

Reconstruction conjointe de l'ordre des gènes de génomés actuels et ancestraux et de leur évolution structurale dans un cadre phylogénétique.

Yoann ANSELMETTI

29 novembre 2017



Plan de la soutenance

Introduction - Contexte historique

Développements algorithmiques

Résultats sur l'assemblage et l'évolution de génomes de moustiques

Conclusion & perspectives

Plan de la soutenance

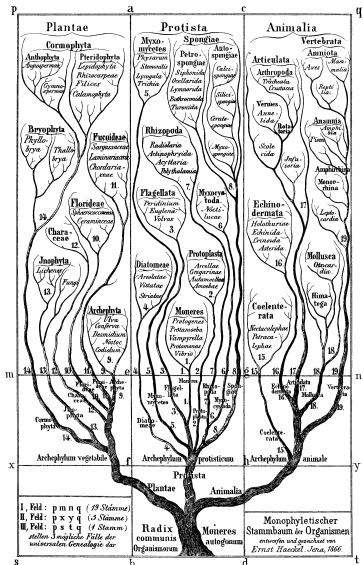
Introduction - Contexte historique

Développements algorithmiques

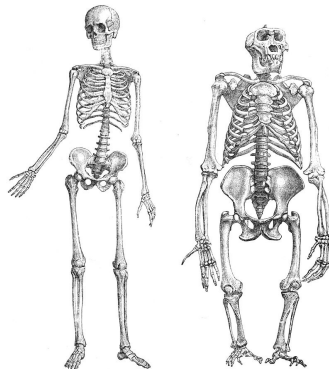
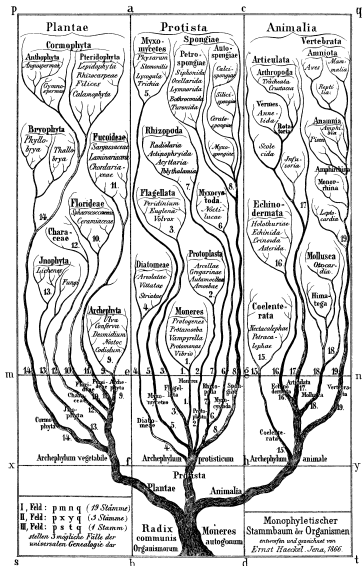
Résultats sur l'assemblage et l'évolution de génomes de moustiques

Conclusion & perspectives

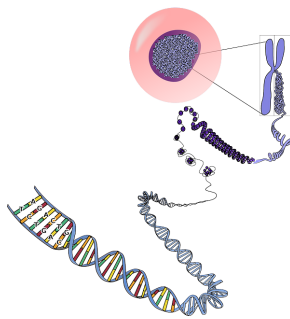
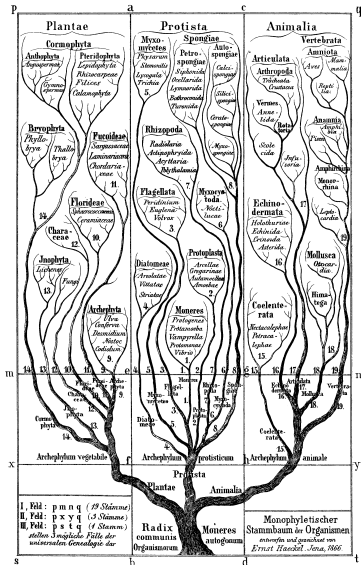
Phylogénie



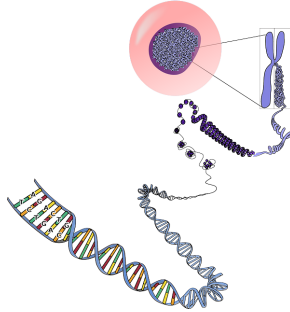
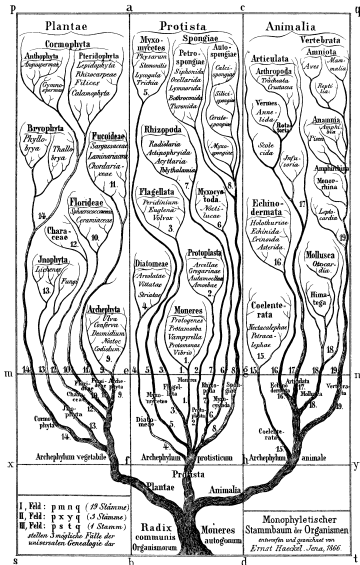
Phylogénie



Phylogenie

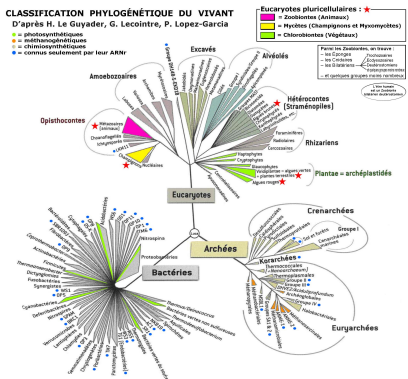


Phylogénie

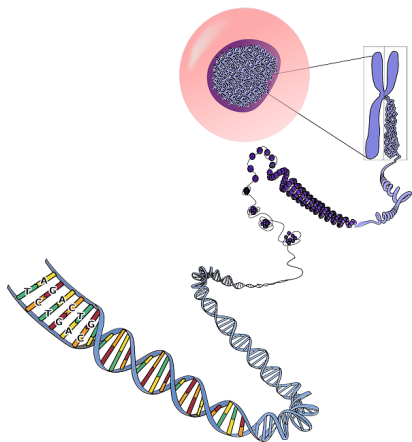


CLASSIFICATION PHYLOGÉNÉTIQUE DU VIVANT

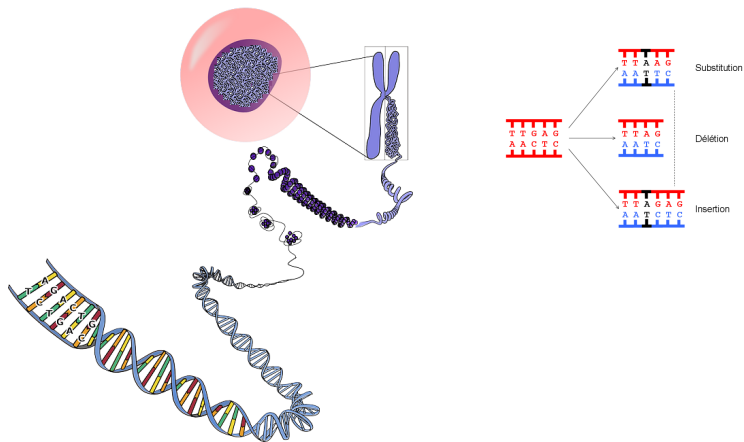
D'après H. Le Guyader, G. Lecointre, P. Lopez-Garcia



Phylogénie moléculaire (mutations de l'ADN)

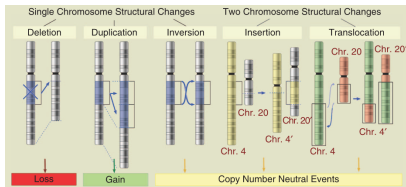
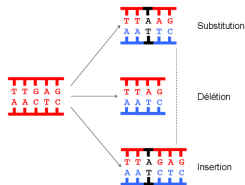
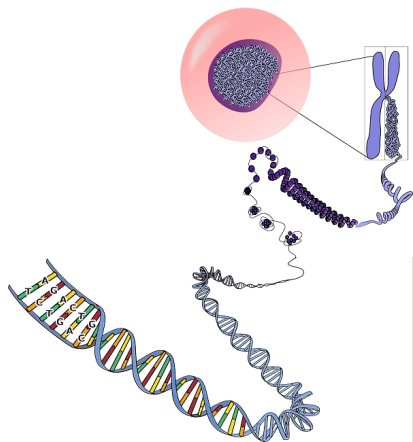


Phylogénie moléculaire (mutations de l'ADN)



► Mutations ponctuelles

Phylogénie moléculaire (mutations de l'ADN)



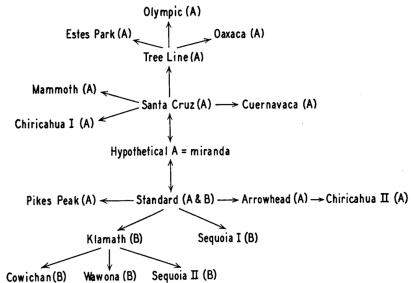
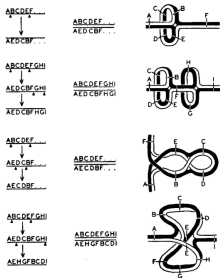
- ▶ Mutations ponctuelles
- ▶ Réarrangements chromosomiques

Phylogénie moléculaire (mutations de l'ADN)

- ▶ Mutations ponctuelles
 - E. Zuckerkandl and L. Pauling, **1965**
- ▶ Réarrangements chromosomiques

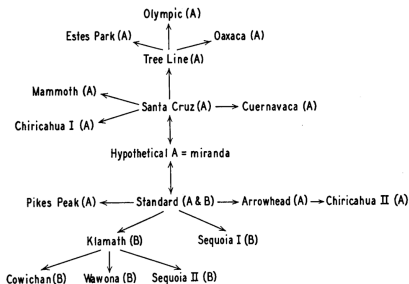
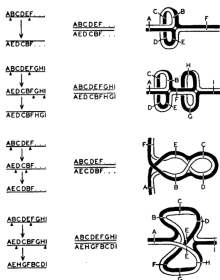
Phylogénie moléculaire (mutations de l'ADN)

- ▶ Mutations ponctuelles
 - E. Zuckerkandl and L. Pauling, **1965**
- ▶ Réarrangements chromosomiques
 - **T. Dobzhansky and A.H. Sturtevant, 1938**



Phylogénie moléculaire (mutations de l'ADN)

- ▶ Mutations ponctuelles
 - E. Zuckerkandl and L. Pauling, **1965**
- ▶ Réarrangements chromosomiques
 - **T. Dobzhansky and A.H. Sturtevant, 1938**



Études de phylogénie moléculaire : mutations ponctuelles » réarrangements

Étude de l'évolution structurale des génomes

Difficultés majeures

1. Complexité algorithmique

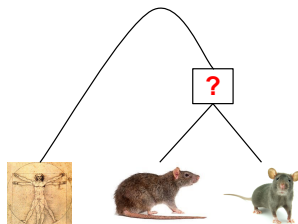
- Difficultés à modéliser l'histoire évolutive des réarrangements génomiques + diversité des événements

Étude de l'évolution structurale des génomes

Difficultés majeures

1. Complexité algorithmique

- Difficultés à modéliser l'histoire évolutive des réarrangements génomiques + diversité des événements



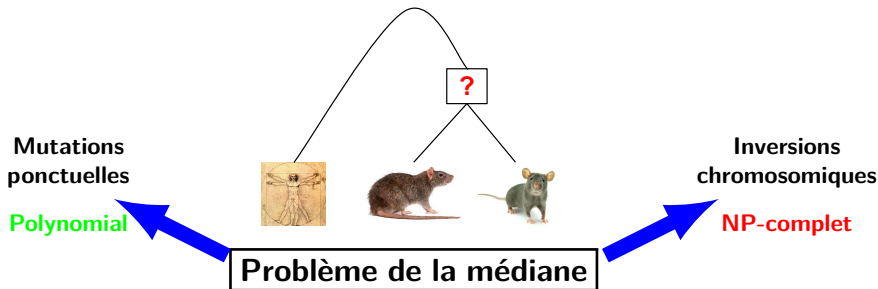
Problème de la médiane

Étude de l'évolution structurale des génomes

Difficultés majeures

1. Complexité algorithmique

- Difficultés à modéliser l'histoire évolutive des réarrangements génomiques + diversité des événements



Étude de l'évolution structurale des génomes

Difficultés majeures

1. Complexité algorithmique
 - Difficultés à modéliser l'histoire évolutive des réarrangements génomiques + diversité des événements
2. Données manquantes : pour beaucoup de génomes la structure demeure incomplète
 - A : séquence et structure complète
 - B : séquence complète mais structure incomplète

A CAGTGCATAGCTCAGTACGCTAGACTGCGTA

B CAGTGCA TAGCTCAGT
AGACT GCGTA ACGCT

Reconstruction de génomes ancestraux

- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Reconstruction de génomes ancestraux

- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Permutations (P)



Reconstruction de génomes ancestraux

- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Permutations (P)

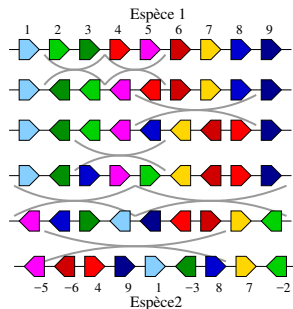


- ▶ Événements :
 - inversion
 - insertion/délétion
 - translocation
 - fusion/fission
 - ...

Reconstruction de génomes ancestraux

- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Permutations (P)

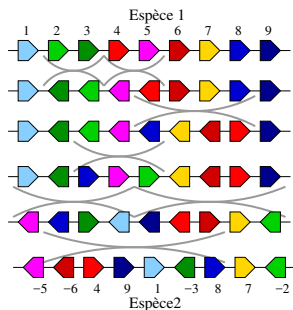


- ▶ Événements :
 - inversion
 - insertion/délétion
 - translocation
 - fusion/fission
 - ...
- ▶ Inférence de scénarios évolutifs

Reconstruction de génomes ancestraux

- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Permutations (P)

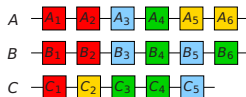


▶ Événements :

- inversion
- insertion/délétion
- translocation
- fusion/fission
- ...

▶ Inférence de scénarios évolutifs

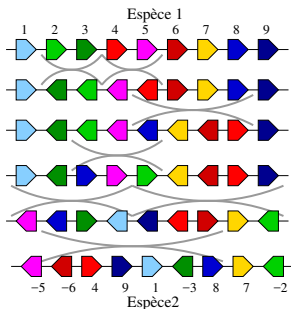
Adjacences/Intervalles (A/I)



Reconstruction de génomes ancestraux

- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Permutations (P)

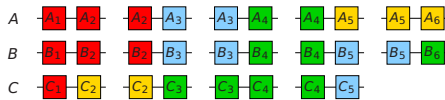
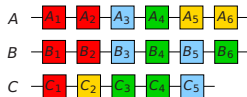


▶ Événements :

- inversion
- insertion/délétion
- translocation
- fusion/fission
- ...

▶ Inférence de scénarios évolutifs

Adjacences/Intervalles (A/I)

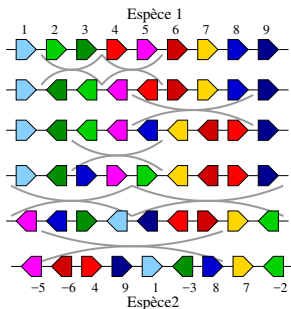


Ensemble d'adjacences

Reconstruction de génomes ancestraux

- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Permutations (P)

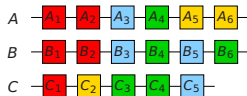


▶ Événements :

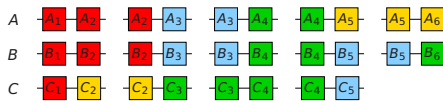
- inversion
- insertion/délétion
- translocation
- fusion/fission
- ...

▶ Inférence de scénarios évolutifs

Adjacences/Intervalles (A/I)



Graphes d'adjacences

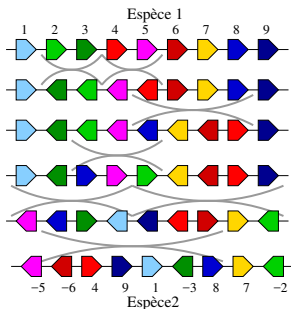


Ensemble d'adjacences

Reconstruction de génomes ancestraux

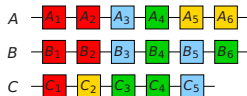
- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Permutations (P)

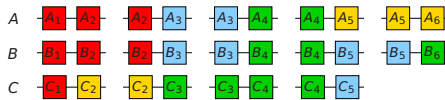


- ▶ Événements :
 - inversion
 - insertion/délétion
 - translocation
 - fusion/fission
 - ...
- ▶ Inférence de scénarios évolutifs

Adjacences/Intervalles (A/I)



Graphe d'adjacences



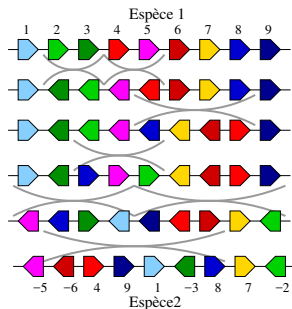
Ensemble d'adjacences

- ▶ Inférence de la **présence/absence** des A/I au niveau ancestral

Reconstruction de génomes ancestraux

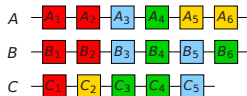
- ▶ 2 classes de méthodes selon comment elles modélisent le génome

Permutations (P)

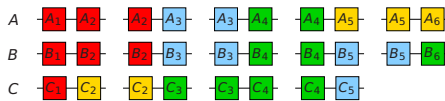


- ▶ Événements :
 - inversion
 - insertion/délétion
 - translocation
 - fusion/fission
 - ...
- ▶ Inférence de scénarios évolutifs

Adjacences/Intervalles (A/I)



Graphe d'adjacences



Ensemble d'adjacences

- ▶ Inférence de la **présence/absence** des A/I au niveau ancestral
- ▶ *Aidée ou pas des histoires de gènes avec pertes/duplications/transferts*

Méthodes de reconstruction de génome ancestraux

Nom	Adjacences Intervalles Permutations	Parcimonie (Pa) Probabiliste (Pr)	Insertions et Pertes	Duplications	Transferts	Explo. solut ^o alter. et/ou support <i>a posteriori</i>
Badger	P	Pr	NON	NON	NON	OUI
GASTS	P	Pa	NON	NON	NON	OUI
PATHGROUPS	P	Pa	NON	NON	NON	NON
MGRA2	P	Pa	OUI	NON	NON	NON
ROCOCO	A/I	Pa	NON	NON	NON	NON
ANGES	A/I	Pa	OUI	NON	NON	OUI
FPSAC	A/I	Pa	OUI	OUI	NON	NON
inferCARs	A	Pa	NON	NON	NON	OUI
DupCAR	A	Pa	OUI	OUI	NON	NON
MLGO	A	Pr	OUI	OUI	NON	NON
DeCo*	A	Pa/Pr	OUI	OUI	OUI	OUI

Anselmetti, Luhman, Bérard, Tannier et Chauve, chapitre "Ancient Genome Reconstruction" du livre "Comparative Genomics : Methods and Protocols" de la série "Methods in Molecular Biology", Springer, À paraître

Méthodes de reconstruction de génome ancestraux

Nom	Adjacences Intervalles Permutations	Parcimonie (Pa) Probabiliste (Pr)	Insertions et Pertes	Duplications	Transferts	Explo. solut ^o alter. et/ou support <i>a posteriori</i>
Badger	P	Pr	NON	NON	NON	OUI
GASTS	P	Pa	NON	NON	NON	OUI
PATHGROUPS	P	Pa	NON	NON	NON	NON
MGRA2	P	Pa	OUI	NON	NON	NON
ROCOCO	A/I	Pa	NON	NON	NON	NON
ANGES	A/I	Pa	OUI	NON	NON	OUI
FPSAC	A/I	Pa	OUI	OUI	NON	NON
inferCARs	A	Pa	NON	NON	NON	OUI
DupCAR	A	Pa	OUI	OUI	NON	NON
MLGO	A	Pr	OUI	OUI	NON	NON
DeCo*	A	Pa/Pr	OUI	OUI	OUI	OUI

Anselmetti, Luhman, Bérard, Tannier et Chauve, chapitre "Ancient Genome Reconstruction" du livre "Comparative Genomics : Methods and Protocols" de la série "Methods in Molecular Biology", Springer, À paraître

Méthodes de reconstruction de génome ancestraux

Nom	Adjacences Intervalles Permutations	Parcimonie (Pa) Probabiliste (Pr)	Insertions et Pertes	Duplications	Transferts	Explo. solut ^o alter. et/ou support <i>a posteriori</i>
Badger	P	Pr	NON	NON	NON	OUI
GASTS	P	Pa	NON	NON	NON	OUI
PATHGROUPS	P	Pa	NON	NON	NON	NON
MGRA2	P	Pa	OUI	NON	NON	NON
ROCOCO	A/I	Pa	NON	NON	NON	NON
ANGES	A/I	Pa	OUI	NON	NON	OUI
FPSAC	A/I	Pa	OUI	OUI	NON	NON
inferCARs	A	Pa	NON	NON	NON	OUI
DupCAR	A	Pa	OUI	OUI	NON	NON
MLGO	A	Pr	OUI	OUI	NON	NON
DeCo*	A	Pa/Pr	OUI	OUI	OUI	OUI

Anselmetti, Luhman, Bérard, Tannier et Chauve, chapitre "Ancient Genome Reconstruction" du livre "Comparative Genomics : Methods and Protocols" de la série "Methods in Molecular Biology", Springer, À paraître

Méthodes de reconstruction de génome ancestraux

Nom	Adjacences Intervalles Permutations	Parcimonie (Pa) Probabiliste (Pr)	Insertions et Pertes	Duplications	Transferts	Explo. solut° alter. et/ou support <i>a posteriori</i>
Badger	P	Pr	NON	NON	NON	OUI
GASTS	P	Pa	NON	NON	NON	OUI
PATHGROUPS	P	Pa	NON	NON	NON	NON
MGRA2	P	Pa	OUI	NON	NON	NON
ROCOCO	A/I	Pa	NON	NON	NON	NON
ANGES	A/I	Pa	OUI	NON	NON	OUI
FPSAC	A/I	Pa	OUI	OUI	NON	NON
inferCARs	A	Pa	NON	NON	NON	OUI
DupCAR	A	Pa	OUI	OUI	NON	NON
MLGO	A	Pr	OUI	OUI	NON	NON
DeCo*	A	Pa/Pr	OUI	OUI	OUI	OUI

Anselmetti, Luhman, Bérard, Tannier et Chauve, chapitre "Ancient Genome Reconstruction" du livre "Comparative Genomics : Methods and Protocols" de la série "Methods in Molecular Biology", Springer, À paraître

Méthodes de reconstruction de génome ancestraux

Nom	Adjacences Intervalles Permutations	Parcimonie (Pa) Probabiliste (Pr)	Insertions et Pertes	Duplications	Transferts	Explo. solut ^o alter. et/ou support <i>a posteriori</i>
Badger	P	Pr	NON	NON	NON	OUI
GASTS	P	Pa	NON	NON	NON	OUI
PATHGROUPS	P	Pa	NON	NON	NON	NON
MGRA2	P	Pa	OUI	NON	NON	NON
ROCOCO	A/I	Pa	NON	NON	NON	NON
ANGES	A/I	Pa	OUI	NON	NON	OUI
FPSAC	A/I	Pa	OUI	OUI	NON	NON
inferCARs	A	Pa	NON	NON	NON	OUI
DupCAR	A	Pa	OUI	OUI	NON	NON
MLGO	A	Pr	OUI	OUI	NON	NON
DeCo*	A	Pa/Pr	OUI	OUI	OUI	OUI


Anselmetti, Luhman, Bérard, Tannier et Chauve, chapitre "Ancient Genome Reconstruction" du livre "Comparative Genomics : Methods and Protocols" de la série "Methods in Molecular Biology", Springer, À paraître

Méthodes de reconstruction de génome ancestraux

Nom	Adjacences Intervalles Permutations	Parcimonie (Pa) Probabiliste (Pr)	Insertions et Pertes	Duplications	Transferts	Explo. solut° alter. et/ou support <i>a posteriori</i>
Badger	P	Pr	NON	NON	NON	OUI
GASTS	P	Pa	NON	NON	NON	OUI
PATHGROUPS	P	Pa	NON	NON	NON	NON
MGRA2	P	Pa	OUI	NON	NON	NON
ROCOCO	A/I	Pa	NON	NON	NON	NON
ANGES	A/I	Pa	OUI	NON	NON	OUI
FPSAC	A/I	Pa	OUI	OUI	NON	NON
inferCARs	A	Pa	NON	NON	NON	OUI
DupCAR	A	Pa	OUI	OUI	NON	NON
MLGO	A	Pr	OUI	OUI	NON	NON
DeCo*	A	Pa/Pr	OUI	OUI	OUI	OUI

Anselmetti, Luhman, Bérard, Tannier et Chauve, chapitre "Ancient Genome Reconstruction" du livre "Comparative Genomics : Methods and Protocols" de la série "Methods in Molecular Biology", Springer, À paraître

Constat de l'assemblage des génomes des bases de données

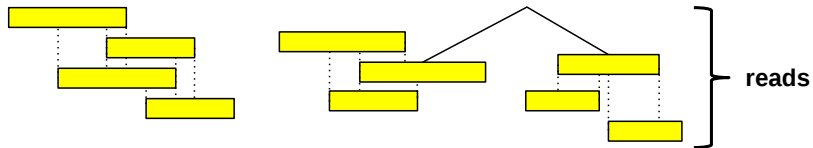
Exemple de la base de données 

Caenorhabditis_elegans 7
Drosophila_melanogaster 11
Monodelphis_domestica 11
Saccharomyces_cerevisiae 17
Gorilla_gorilla 27
Tetraodon_nigroviridis 27
Homo_sapiens 39
Mus_musculus 40
Bos_taurus 43
Pongo_abelii 53
Rattus_norvegicus 69
Taeniopygia_guttata 69
Equus_caballus 102
Chlorocebus_sabaeus 115
Felis_catus 213
Canis_familiaris 227
Pan_troglodytes 245
Ciona_savignyi 332
Papio_anubis 367
Cavia_porcellus 369
Danio_rerio 448
Lepisosteus_oculatus 494
Otolemur_garnettii 523

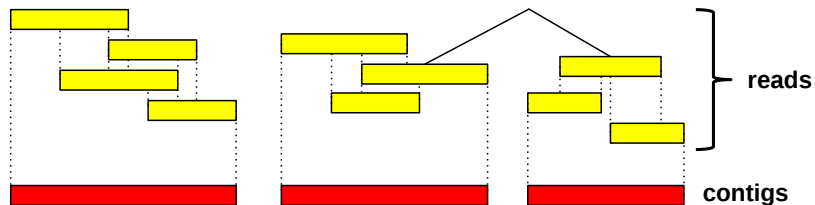
Gasterosteus_aculeatus 550
Loxodonta_africana 583
Ciona_intestinalis 614
Meleagris_gallopavo 631
Ovis_aries 638
Callithrix_jacchus 653
Nomascus_leucogenys 657
Macaca_mulatta 692
Mustela_putorius_furo 731
Ficedula_albicollis 757
Ictidomys_tridecemlineatus 761
Gallus_gallus 766
Oryzias_latipes 807
Oryctolagus_cuniculus 1043
Oreochromis_niloticus 1075
Sus_scrofa 1332
Poecilia_formosa 1504
Takifugu_rubripes 1923
Pelodiscus_sinensis 2162
Xenopus_tropicalis 2232
Anas_platyrhynchos 2338
Anolis_carolinensis 2376
Myotis_lucifugus 2393

Xiphophorus_maculatus 2445
Astyanax_mexicanus 2557
Ailuropoda_melanoleuca 2599
Sarcophilus_harrisii 3505
Latimeria_chalumnae 3948
Dasypus_novemcinctus 3958
Vicugna_pacos 4951
Gadus_morhua 5336
Petromyzon_marinus 5530
Tursiops_truncatus 5624
Pteropus_vampyrus 6328
Microcebus_murinus 6837
Ochotona_princeps 7324
Tupaia_belangeri 8121
Dipodomys_ordii 9720
Sorex_araneus 9740
Ornithorhynchus_anatinus 10158
Procavia_capensis 10465
Choloepus_hoffmanni 10770
Tarsius_syrichta 11528
Erinaceus_europaeus 11536
Echinops_telfairi 12301
Macropus_eugenii 12704

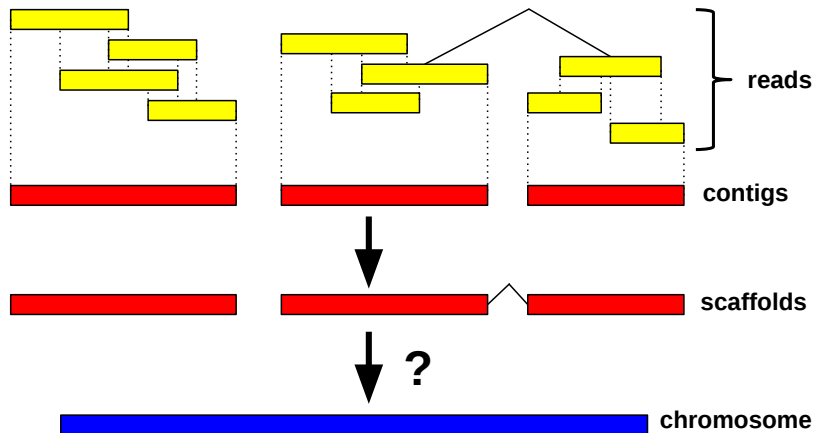
Pipeline général pour l'assemblage de novo de génomes



Pipeline général pour l'assemblage de novo de génomes



Pipeline général pour l'assemblage de novo de génomes



Scaffolding : Reconstruire l'ordre et l'orientation des contigs

Référence unique

Projector2
(2005)

SIS
(2012)

CAR
(2014)

OSLay
(2007)

fillScaffolds
(2010)

ABACAS
(2009)

r2cat
(2009)

Chromosomer
(2016)

Mauve Aligner
(2009)

CLA
(2016)

AlignGraph
(2014)

Multiple référence

MEDUSA
(2015)

Multi-CAR
(2016)

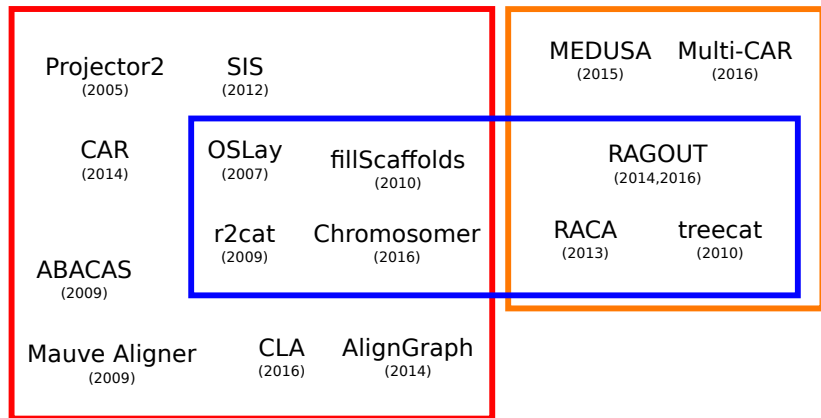
RAGOUT
(2014,2016)

RACA
(2013)

treecat
(2010)

Référence unique

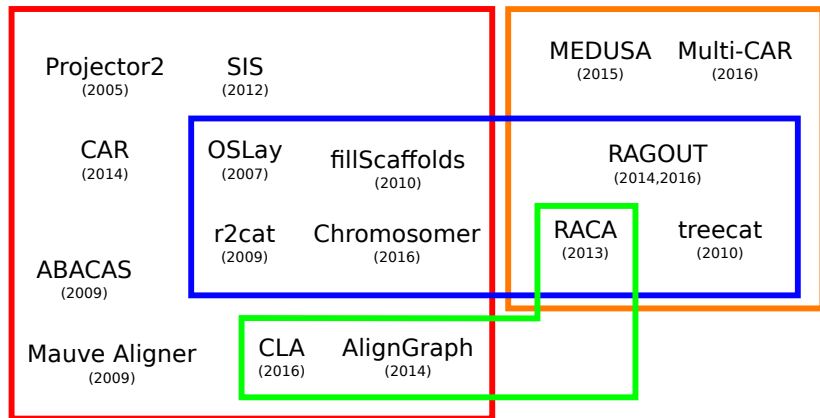
Multiple référence



Génomes de grandes tailles

Référence unique

Multiple référence



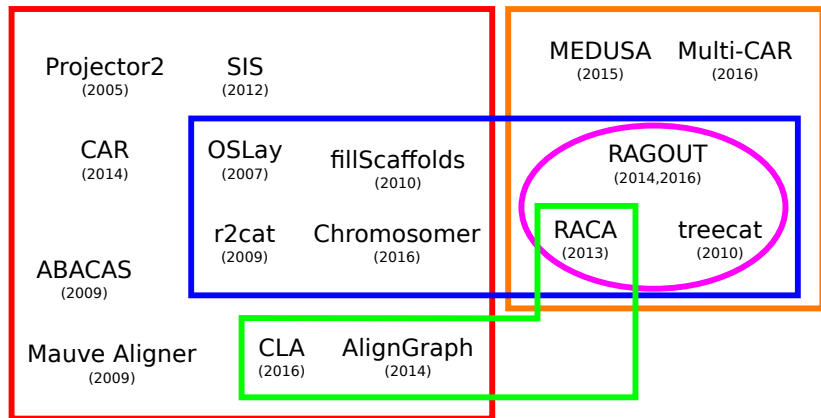
Données de séquençage appariées

Génomes de grandes tailles

Méthodes de *Scaffolding* par génomique comparative

Référence unique

Multiple référence



Données de séquençage appariées

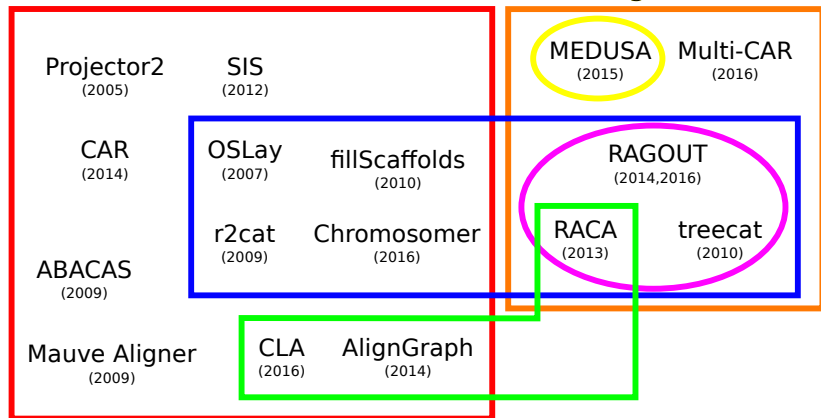
Génomes de grandes tailles

Signal phylogénétique

Méthodes de *Scaffolding* par génomique comparative

Référence unique

Multiple référence Références fragmentées



Données de séquençage appariées

Génomes de grandes tailles

Signal phylogénétique

Méthodes de *Scaffolding* par génomique comparative

Multiple référence

Projector2
(2005)

SIS
(2012)

CAR
(2014)

OSLay
(2007)

fillScaffolds
(2010)

ABACAS
(2009)

r2cat
(2009)

Chromosomer
(2016)

Mauve Aligner
(2009)

CLA
(2016)

AlignGraph
(2014)

MEDUSA
(2015)

Multi-CAR
(2016)

RAGOUT
(2014,2016)

RACA
(2013)

treecat
(2010)

Méthodes de *Scaffolding* par génomique comparative

Multiple référence

Projector2
(2005)

SIS
(2012)

MEDUSA
(2015)

Multi-CAR
(2016)

CAR
(2014)

OSLay
(2007)

fillScaffolds
(2010)

RAGOUT
(2014,2016)

ABACAS
(2009)

r2cat
(2009)

Chromosomer
(2016)

RACA
(2013)

treecat
(2010)

Mauve Aligner
(2009)

CLA
(2016)

AlignGraph
(2014)

Génomes de grandes tailles

Méthodes de *Scaffolding* par génomique comparative

Multiple référence Références fragmentées

Projector2
(2005)

SIS
(2012)

MEDUSA
(2015)

Multi-CAR
(2016)

CAR
(2014)

OSLay
(2007)

fillScaffolds
(2010)

RAGOUT
(2014,2016)

ABACAS
(2009)

r2cat
(2009)

Chromosomer
(2016)

RACA
(2013)

treecat
(2010)

Mauve Aligner
(2009)

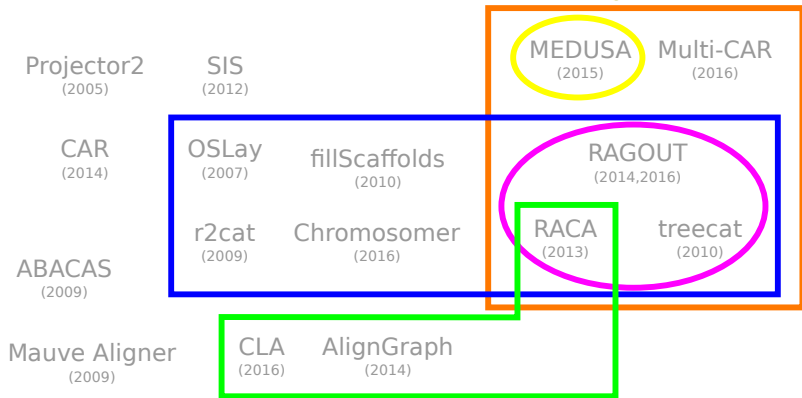
CLA
(2016)

AlignGraph
(2014)

Génomes de grandes tailles

Méthodes de *Scaffolding* par génomique comparative

Multiple référence
Références fragmentées



Données de séquençage appariées

Génomes de grandes tailles

Signal phylogénétique

Méthodes de *Scaffolding* par génomique comparative

Référence et cible

Multiple référence Références fragmentées

Projector2
(2005)

SIS
(2012)

MEDUSA
(2015)

Multi-CAR
(2016)

CAR
(2014)

OSLay
(2007)

fillScaffolds
(2010)

RAGOUT
(2014,2016)

ABACAS
(2009)

r2cat
(2009)

Chromosomer
(2016)

RACA
(2013)

treecat
(2010)

Mauve Aligner
(2009)

CLA
(2016)

AlignGraph
(2014)

Données de séquençage appariées

Génomes de grandes tailles

Signal phylogénétique

2 approches, 1 même problème

Reconstruction de génomes ancestraux & Assemblage de génomes

=

Reconstruire l'ordre de marqueurs le long de génomes

2 approches, 1 même problème

Reconstruction de génomes ancestraux & Assemblage de génomes

=

Reconstruire l'ordre de marqueurs le long de génomes

- ▶ Lien phylogénétique des 2 problèmes :

2 approches, 1 même problème

Reconstruction de génomes ancestraux & Assemblage de génomes

=

Reconstruire l'ordre de marqueurs le long de génomes

- ▶ Lien phylogénétique des 2 problèmes :
 - Génomes actuels dérivent de l'évolution des génomes ancestraux

2 approches, 1 même problème

Reconstruction de génomes ancestraux & Assemblage de génomes

=

Reconstruire l'ordre de marqueurs le long de génomes

- ▶ Lien phylogénétique des 2 problèmes :
 - Génomes actuels dérivent de l'évolution des génomes ancestraux
 - Génomes ancestraux sont reconstruits à partir des génomes actuels

2 approches, 1 même problème

Reconstruction de génomes ancestraux & Assemblage de génomes

=

Reconstruire l'ordre de marqueurs le long de génomes

- ▶ Lien phylogénétique des 2 problèmes :
 - Génomes actuels dérivent de l'évolution des génomes ancestraux
 - Génomes ancestraux sont reconstruits à partir des génomes actuels

Sujet de thèse

Méthode bioinformatique intégrative permettant conjointement de reconstruire l'évolution de la structure des génomes et améliorer l'assemblage des génomes actuels descendants.

Plan de la soutenance

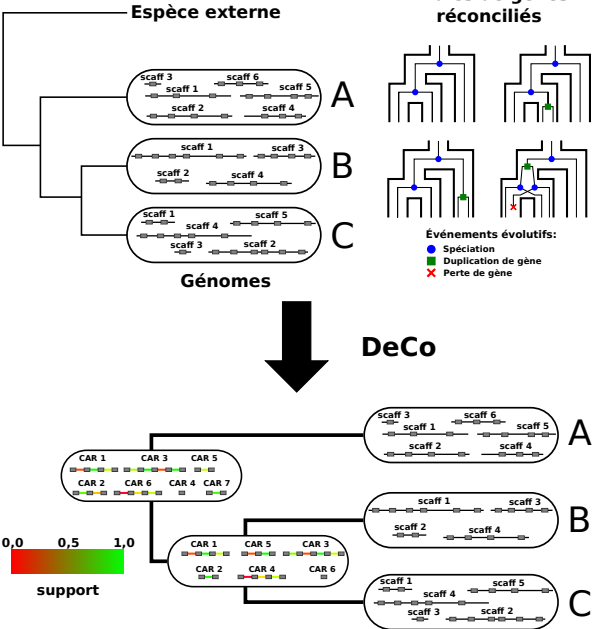
Introduction - Contexte historique

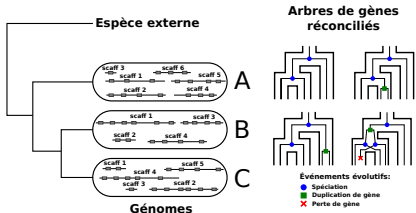
Développements algorithmiques

Résultats sur l'assemblage et l'évolution de génomes de moustiques

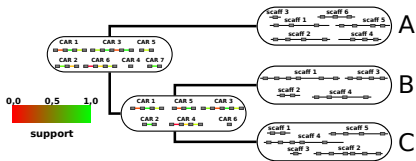
Conclusion & perspectives

Base des développements algorithmiques/méthodologiques



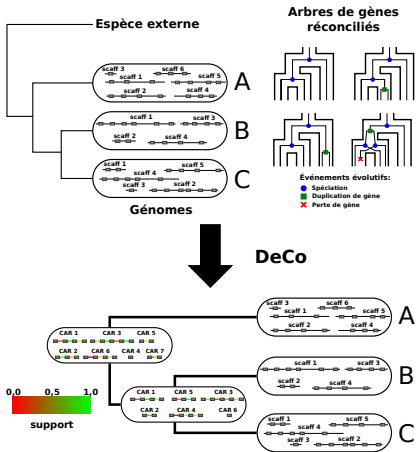


DeCo



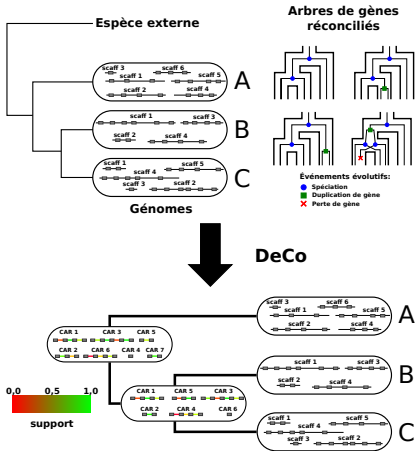
Apports

1. Intégration de la fragmentation des génomes actuels et prédiction de nouvelles adjacents actuelles



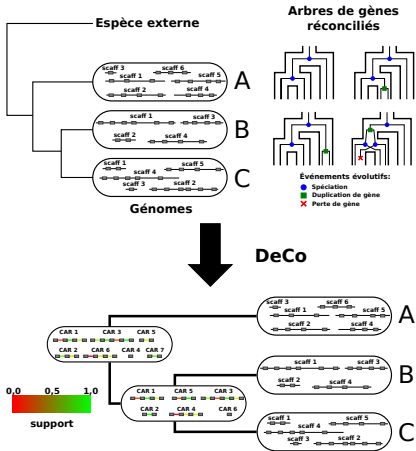
Apports

1. Intégration de la fragmentation des génomes actuels et prédiction de nouvelles adjacences actuelles
2. Intégration d'adjacences avec probabilité *a priori*



Apports

1. Intégration de la fragmentation des génomes actuels et prédiction de nouvelles adjacences actuelles
2. Intégration d'adjacences avec probabilité *a priori*
3. Intégration des extensions dans le logiciel DeCoSTAR



Apports

1. Intégration de la fragmentation des génomes actuels et prédiction de nouvelles adjacences actuelles
2. Intégration d'adjacences avec probabilité *a priori*
3. Intégration des extensions dans le logiciel DECoSTAR
4. Pipelines de pré- et post-traitement des données

Principe de DECO

- ▶ Reconstitue l'histoire évolutive des adjacences de gènes
- ▶ Considère des duplications, pertes (et transferts) de gènes

Principe de DECO

- ▶ Reconstitue l'histoire évolutive des adjacences de gènes
- ▶ Considère des duplications, pertes (et transferts) de gènes
- ▶ Coût d'une histoire évolutive : $\#(\text{créations} + \text{cassures})$ d'adjacences
- ▶ Principe dérivé de la parcimonie : on cherche l'histoire évolutive d'adjacences de coût minimum qui respecte l'histoire des gènes

Principe de DECO

- ▶ Reconstitue l'histoire évolutive des adjacences de gènes
- ▶ Considère des duplications, pertes (et transferts) de gènes
- ▶ Coût d'une histoire évolutive : $\#(\text{créations} + \text{cassures})$ d'adjacences
- ▶ Principe dérivé de la parcimonie : on cherche l'histoire évolutive d'adjacences de coût minimum qui respecte l'histoire des gènes
- ▶ Hypothèse : adjacences évoluent indépendamment

Principe de DECO

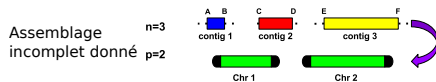
- ▶ Reconstitue l'histoire évolutive des adjacences de gènes
- ▶ Considère des duplications, pertes (et transferts) de gènes
- ▶ Coût d'une histoire évolutive : $\#(\text{créations} + \text{cassures})$ d'adjacences
- ▶ Principe dérivé de la parcimonie : on cherche l'histoire évolutive d'adjacences de coût minimum qui respecte l'histoire des gènes
- ▶ Hypothèse : adjacences évoluent indépendamment
- ▶ **Ne considère pas la fragmentation des génomes**
Pour deux gènes g_1 et g_2 d'un génome actuel.

$$\text{coût}(g_1 \sim g_2) = \begin{cases} 0 & \text{si } g_1 \text{ est adjacent à } g_2 \text{ dans les données} \\ \infty & \text{sinon} \end{cases}$$

Où $\text{coût}(g_1 \sim g_2) = \text{coût d'avoir } g_1 \text{ adjacent à } g_2 \text{ dans l'histoire évolutive}$

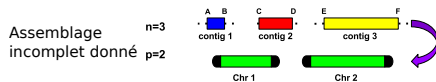
Considération de la fragmentation des génomes actuels

- ▶ $n = \#$ contigs dans l'assemblage
- ▶ $p = \#$ chromosomes
- ▶ $n > p$



Considération de la fragmentation des génomes actuels

- ▶ $n = \#$ contigs dans l'assemblage
- ▶ $p = \#$ chromosomes
- ▶ $n > p$

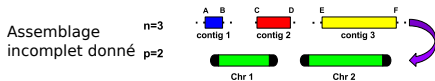


Considération de la fragmentation des génomes actuels

- ▶ $n = \# \text{contigs}$ dans l'assemblage
- ▶ $p = \# \text{chromosomes}$
- ▶ $n > p$

Question

Les gènes en extrémités de contigs dans les génomes actuels sont-ils en réalité adjacents ou non adjacents ?

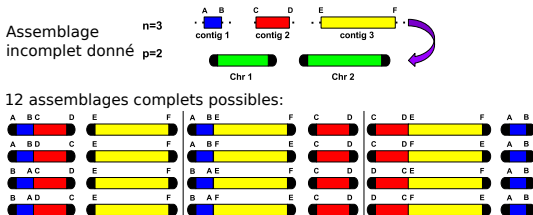


Considération de la fragmentation des génomes actuels

- ▶ $n = \# \text{contigs dans l'assemblage}$
- ▶ $p = \# \text{chromosomes}$
- ▶ $n > p$

Question

Les gènes en extrémités de contigs dans les génomes actuels sont-ils en réalité adjacents ou non adjacents ?

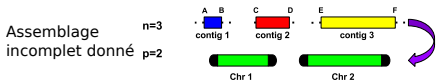


Considération de la fragmentation des génomes actuels

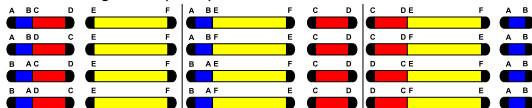
- ▶ $n = \# \text{contigs dans l'assemblage}$
- ▶ $p = \# \text{chromosomes}$
- ▶ $n > p$

Question

Les gènes en extrémités de contigs dans les génomes actuels sont-ils en réalité adjacents ou non adjacents ?



12 assemblages complets possibles:



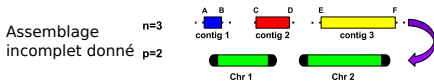
$$P(B \sim C) = \frac{1}{12}$$

Considération de la fragmentation des génomes actuels

- ▶ $n = \# \text{contigs dans l'assemblage}$
- ▶ $p = \# \text{chromosomes}$
- ▶ $n > p$

Question

Les gènes en extrémités de contigs dans les génomes actuels sont-ils en réalité adjacents ou non adjacents ?



12 assemblages complets possibles:



$$P(B \sim C) = \frac{1}{12}$$

Calcul de la probabilité : $P(g_1 \sim g_2)$

$$P(g_1 \sim g_2) = \frac{\# \text{ d'assemblage complet où } g_1 \sim g_2}{\# \text{ total d'assemblages complets}}$$

Intégration de la fragmentation des génomes actuels

Rappels de DECO

Pour deux gènes g_1 et g_2 d'un génome actuel.

$$\text{coût}(g_1 \sim g_2) = \begin{cases} 0 & \text{si } g_1 \text{ est adjacent à } g_2 \text{ dans les données} \\ \infty & \text{sinon} \end{cases}$$

Intégration de la fragmentation des génomes actuels

Rappels de DECO

Pour deux gènes g_1 et g_2 d'un génome actuel.

$$\text{coût}(g_1 \sim g_2) = \begin{cases} 0 & \text{si } g_1 \text{ est adjacent à } g_2 \text{ dans les données} \\ \infty & \text{sinon} \end{cases}$$

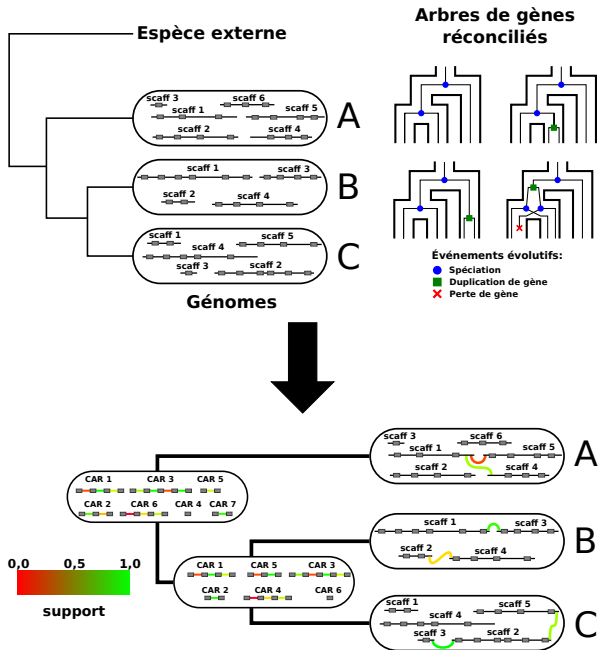
Transformation $P(g_1 \sim g_2) \rightarrow \text{coût}(g_1 \sim g_2)$

- ▶ $P(g_1 \sim g_2) = 1 \rightarrow \text{coût}(g_1 \sim g_2) = 0$
- ▶ $P(g_1 \sim g_2) = 0 \rightarrow \text{coût}(g_1 \sim g_2) = \infty$

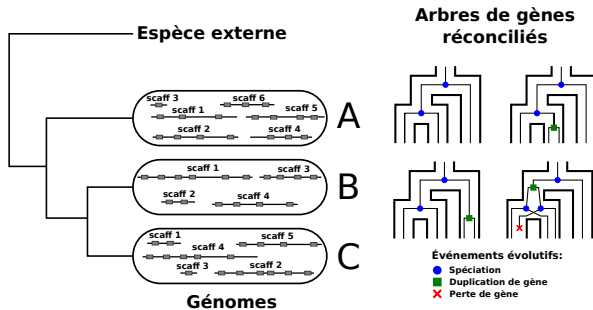
⇒ Utilisation de la fonction \log : gradient de coûts entre 0 et ∞

$$\text{coût}(g_1 \sim g_2) = -\log_b(P(g_1 \sim g_2))$$

Intégration de la fragmentation des génomes actuels



Intégration de la fragmentation des génomes actuels

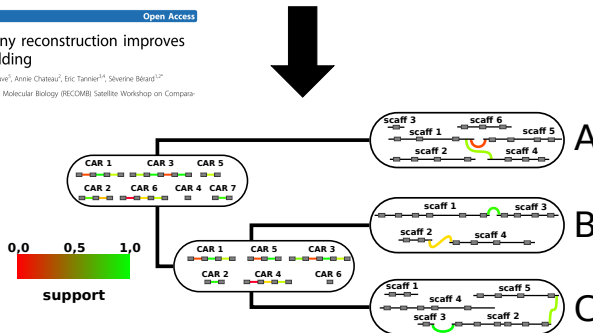


RESEARCH Open Access

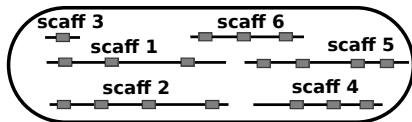
Ancestral gene synteny reconstruction improves extant species scaffolding

Yoann Anselmetti^{1,3}, Vincent Berry², Cedric Chauve², Annie Chateau², Eric Tannier^{1,4}, Sylvaine Bédard^{1,2}

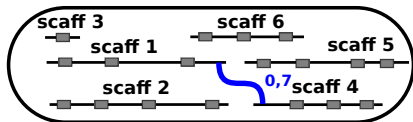
From 13th Annual Research in Computational Molecular Biology (RECOMB) Satellite Workshop on Comparative Genomics
Frankfurt, Germany, 4-7 October 2015



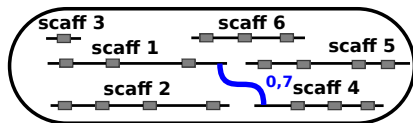
Ajout d'adjacences avec probabilité *a priori*



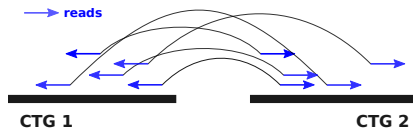
Ajout d'adjacences avec probabilité *a priori*



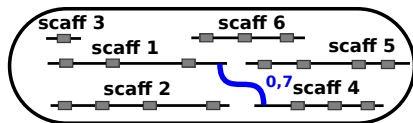
Ajout d'adjacences avec probabilité *a priori*



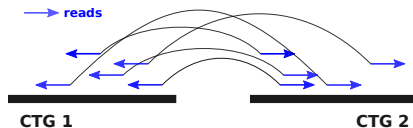
Adjacences de scaffolding (reads appariés)



Ajout d'adjacences avec probabilité *a priori*



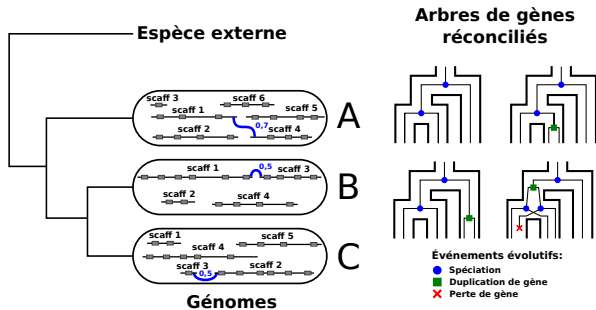
Adjacences de scaffolding (reads appariés)



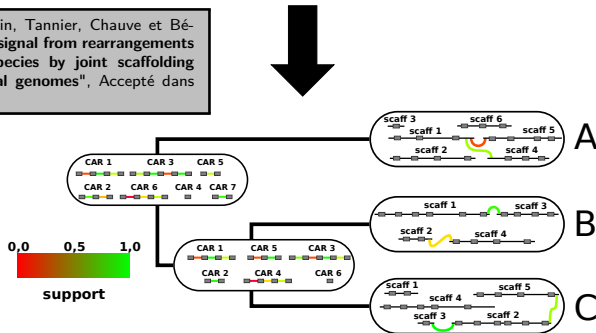
3 types d'adjacences

- ▶ Adjacences observées
- ▶ Adjacences scaffolding / avec probabilité *a priori*
- ▶ Adjacences non observées

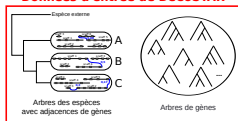
Intégration d'adjacences avec probabilité *a priori*



Anselmetti, Duchemin, Tannier, Chauve et Bérard, "Phylogenetic signal from rearrangements in 18 *Anopheles* species by joint scaffolding extant and ancestral genomes", *Accepted dans BMC Genomics*



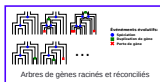
Données d'entrée de DeCoSTAR



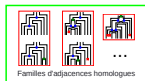
DeCoSTAR: Reconstructing the Ancestral Organization of Genes or Genomes Using Reconciled Phylogenies

Wandille Duchemin^{1,2,*}, Yoann Anselmetti^{2,3}, Murray Patterson^{2,4}, Yann Ponty^{5,6}, Séverine Bérard^{3,7}, Cedric Chauve⁸, Celine Scornavacca¹, Vincent Daubin², and Eric Tannier^{1,2}

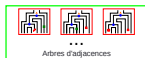
1/ Réconciliation avec ecceTERA



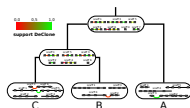
2/ Création des familles d'adjacences homologues



3/ Inférence des arbres d'adjacences



4/ Reconstruction de l'ordre des gènes

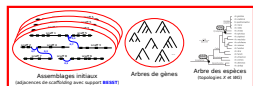


Résultats de DeCoSTAR

Algorithmes intégrés

- ▶ ECCE_{TERA} (réconciliation)
- ▶ DECo (algorithme initial)
- ▶ DECoLT (transfert de gène)
- ▶ DEClone (support *a posteriori*)
- ▶ ART-DECo (fragmentation des génomes)
- ▶ ADSEQ (adjacences avec probabilité *a priori*)

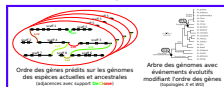
Pré-traitement et linéarisation des prédictions



Données initiales de DeCoSTAR



II/ Exécution de DeCoSTAR
(avec les algorithmes DeClone et ADseq+DeClone)

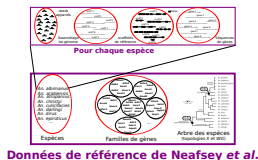


Prédictions de DeCoSTAR

DeCoSTAR

- ▶ Volume généré : 100Mo-1Go
- ▶ Temps de calcul : 1h-1j/run

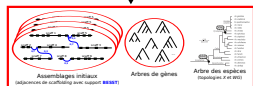
Pré-traitement et linéarisation des prédictions



I/ Pipeline de génération des données de DeCoSTAR

Pré-traitement

- ▶ Pour générer les données d'entrée :
 - adjacences (et support)
 - arbres de gènes
- ▶ cas des adj. de *scaffolding*
 - Volume généré : plusieurs To
 - Temps de calcul : 2 semaines/espèce

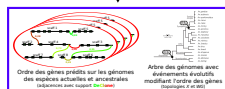


Données initiales de DeCoSTAR

II/ Exécution de DeCoSTAR (avec les algorithmes DeClone et ADseq+DeClone)

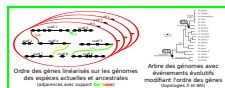
DeCoSTAR

- ▶ Volume généré : 100Mo-1Go
- ▶ Temps de calcul : 1h-1j/run



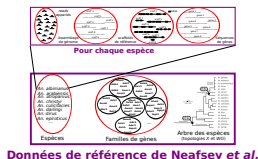
Prédictions de DeCoSTAR

III/ Linéarisation des prédictions de DeCoSTAR

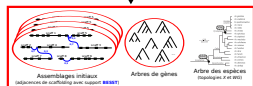


Prédictions de DeCoSTAR linéarisées

Pré-traitement et linéarisation des prédictions

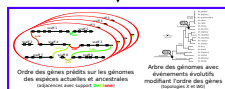


I/ Pipeline de génération des données de DeCoSTAR



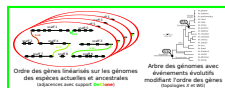
Données initiales de DeCoSTAR

II/ Exécution de DeCoSTAR (avec les algorithmes DeClone et ADseq+DeClone)



Prédictions de DeCoSTAR

III/ Linéarisation des prédictions de DeCoSTAR



Prédictions de DeCoSTAR linéarisées

Pré-traitement

- ▶ Pour générer les données d'entrée :
 - adjacences (et support)
 - arbres de gènes
- ▶ cas des adj. de *scaffolding*
 - Volume généré : plusieurs To
 - Temps de calcul : 2 semaines/espèce

DeCoSTAR

- ▶ Volume généré : 100Mo-1Go
- ▶ Temps de calcul : 1h-1j/run

Linéarisation

- ▶ Pour résoudre les conflits synténiques
- ▶ Temps de calcul : 1-2j/run

Plan de la soutenance

Introduction - Contexte historique

Développements algorithmiques

Résultats sur l'assemblage et l'évolution de génomes de moustiques

Conclusion & perspectives

Paludisme

Une maladie qui devient de plus en plus résistante aux traitements

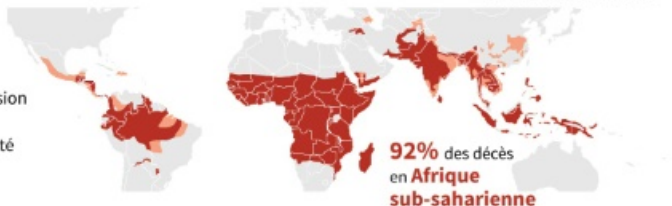
3,2 milliards de personnes exposées
214 millions de cas en 2015

429 000 morts en 2015, dont
303 000 enfants

En Afrique
Premier cas de résistance observé chez un malade chinois

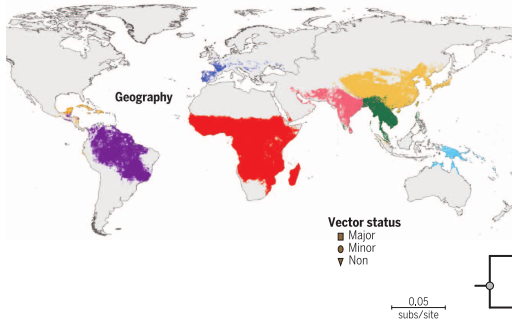
En Asie du Sud-Est
Une nouvelle souche résistante aux traitements identifiée au Cambodge, en Thaïlande et au Laos

- Zones de transmission
- Zones à risque limité



Jeu de données de 18 *Anopheles*

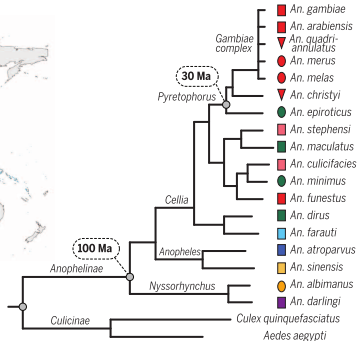
- ▶ 18 espèces d'*Anopheles*
- ▶ Nombre de *scaffolds* dans les génomes de référence :
 - Moyenne : 1 899,22 ($\pm 2 020,99$)
 - Médiane : 507
 - Min : 6 (*Anopheles gambiae*)
 - Max : 8 600 (*Anopheles maculatus*)
- ▶ 14 940 arbres de gènes
- ▶ 183 680 gènes



MOSQUITO GENOMICS

Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes

Daniel E. Neafsey,^{††} Robert M. Waterhouse,^{*} *et al.*



Jeu de données de 18 *Anopheles*

- ▶ 18 espèces d'*Anopheles*
- ▶ Nombre de *scaffolds* dans les génomes de référence :
 - Moyenne : 1 899,22 ($\pm 2\ 020,99$)
 - Médiane : 507
 - Min : 6 (*Anopheles gambiae*)
 - Max : 8 600 (*Anopheles maculatus*)
- ▶ 14 940 arbres de gènes
- ▶ 183 680 gènes
- ▶ 2 topologies pour la phylogénie des espèces
 - topologie du chromosome X (considérée comme la topologie des espèces)
 - topologie du génome entier

MOSQUITO GENOMICS

Extensive introgression in a malaria vector species complex revealed by phylogenomics

Michael C. Fontaine,^{1,2,3,†} James B. Pease,^{3,†} Aaron Steele,⁴ Robert M. Waterhouse,^{5,6,7,8} Daniel E. Neafsey,⁹ Igor V. Sharakhov,^{9,10} Xiaofang Jang,¹⁰ Andrew B. Hall,¹⁰ Flaminia Catteruccia,^{11,12} Evdokia Kakani,^{11,12} Sara N. Mitchell,¹¹ Yi-Chieh Wu,⁵ Hilary A. Smith,^{1,2} R. Rebecca Love,^{1,2} Mara K. Lawniczak,^{13,†} Michel A. Slotman,¹⁴ Scott J. Emrich,^{2,4} Matthew W. Hahn,^{3,15} Nora J. Besansky^{1,2,5}

Jeu de données de 18 *Anopheles*

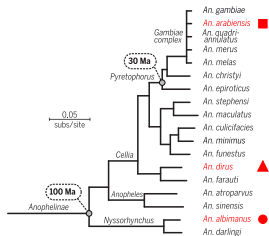
- ▶ 18 espèces d'*Anopheles*
- ▶ Nombre de *scaffolds* dans les génomes de référence :
 - Moyenne : 1 899,22 ($\pm 2 020,99$)
 - Médiane : 507
 - Min : 6 (*Anopheles gambiae*)
 - Max : 8 600 (*Anopheles maculatus*)
- ▶ 14 940 arbres de gènes
- ▶ 183 680 gènes
- ▶ 2 topologies pour la phylogénie des espèces
 - topologie du chromosome X (considérée comme la topologie des espèces)
 - topologie du génome entier
- ▶ *Collaboration avec les membres du consortium Anopheles*
 - v2.0 des assemblages des génomes d'*Anopheles*

Validation de DECOSTAR par baisse de la qualité d'assemblage

- ▶ Génération d'un assemblage plus conservatif/fragmenté
- ▶ Annotation nouvel assemblage avec gènes de l'assemblage de référence

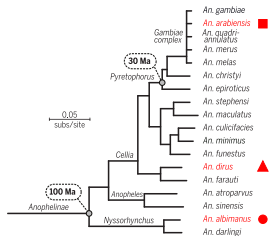
Validation de DECoSTAR par baisse de la qualité d'assemblage

- ▶ Génération d'un assemblage plus conservatif/fragmenté
- ▶ Annotation nouvel assemblage avec gènes de l'assemblage de référence



Validation de DECoSTAR par baisse de la qualité d'assemblage

- ▶ Génération d'un assemblage plus conservatif/fragmenté
- ▶ Annotation nouvel assemblage avec gènes de l'assemblage de référence

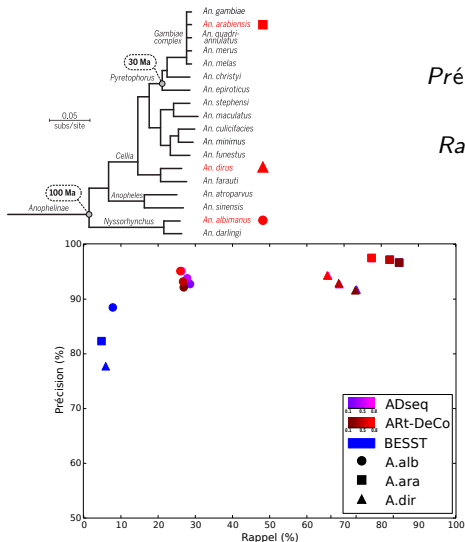


$$\text{Précision} = \frac{\# \text{adj inférées vraies}}{\# \text{adj inférées}}$$

$$\text{Rappel} = \frac{\# \text{adj inférées vraies}}{\# \text{adj occultées}}$$

Validation de DECoSTAR par baisse de la qualité d'assemblage

- ▶ Génération d'un assemblage plus conservatif/fragmenté
- ▶ Annotation nouvel assemblage avec gènes de l'assemblage de référence



$$\text{Précision} = \frac{\# \text{adj inférées vraies}}{\# \text{adj inférées}}$$

$$\text{Rappel} = \frac{\# \text{adj inférées vraies}}{\# \text{adj occultées}}$$

ADseq: Signal **phylogénie** + **séquence**

ART-DeCo: Signal **phylogénie**

BESST: Signal **séquence**

Résoudre l'assemblage du génome d'*Anopheles funestus*

Anopheles funestus

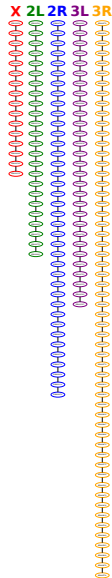
- ▶ Assemblage de référence : 534 *scaffolds*
- ▶ Assemblage DECoSTAR : 203 *scaffolds*
 - DECoSTAR : 331 nouvelles adjacences

Résoudre l'assemblage du génome d'*Anopheles funestus*

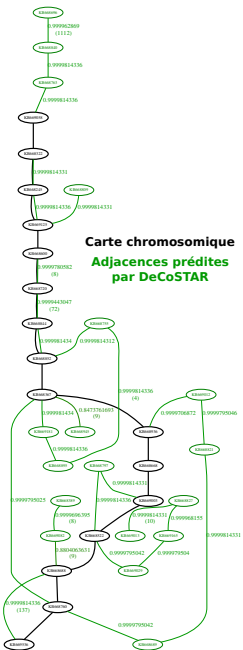
Anopheles funestus

- ▶ Assemblage de référence : 534 *scaffolds*
- ▶ Assemblage DECoSTAR : 203 *scaffolds*
 - DECoSTAR : 331 nouvelles adjacences

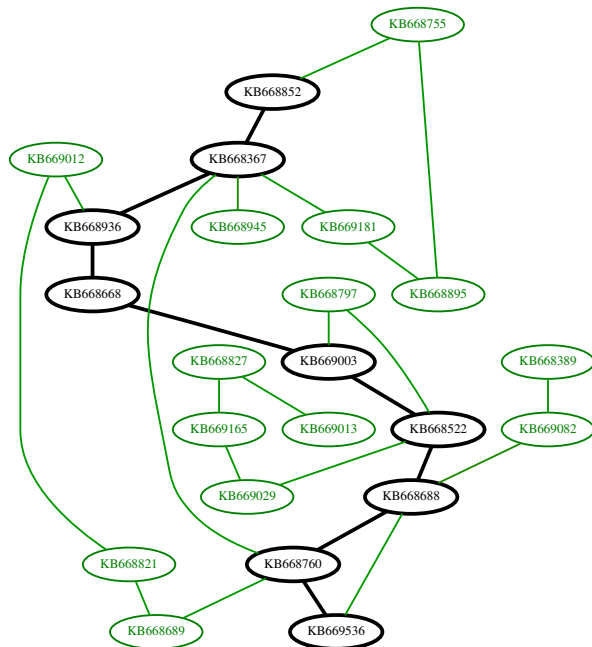
Carte chrom. d'*An. funestus*
(équipe d'I. Sharakhov)



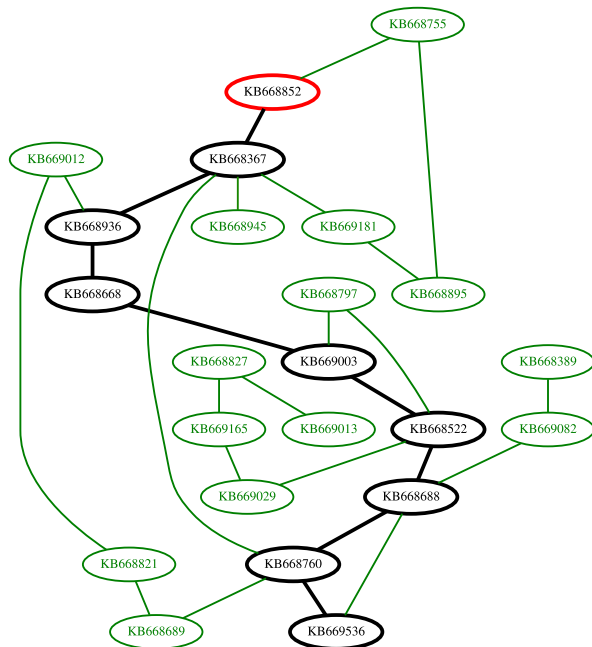
Graphe d'adjacences du chromosome X d'*Anopheles funestus*



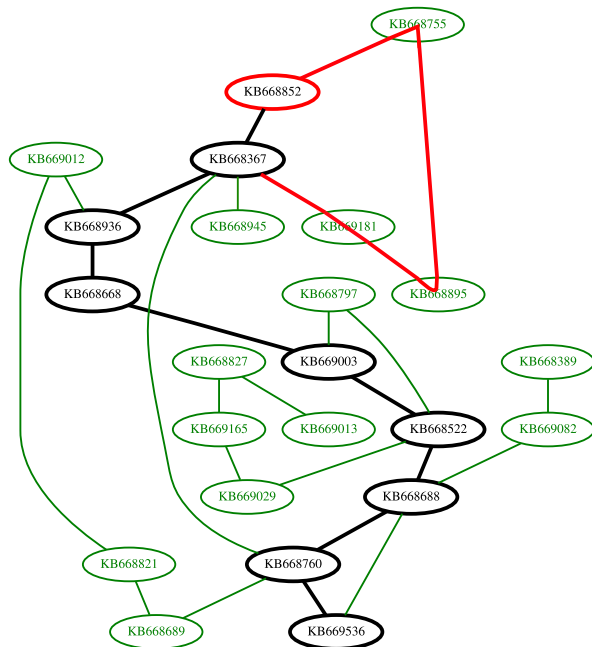
Graphe d'adjacences du chromosome X d'*Anopheles funestus*



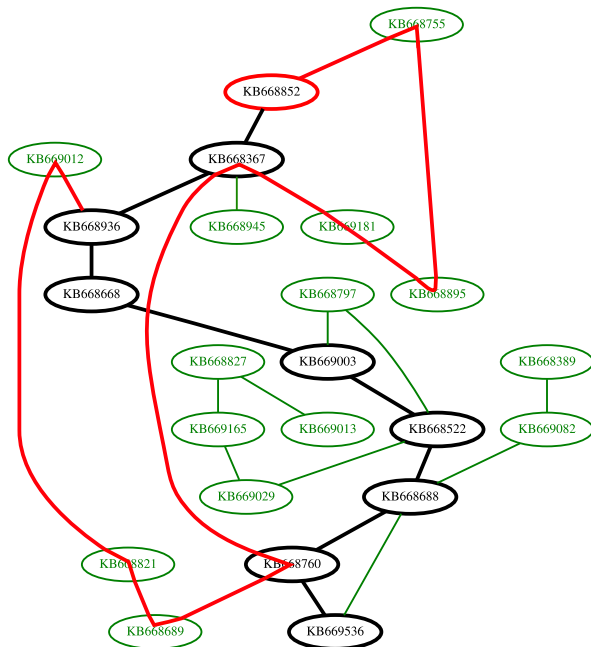
Grphe d'adjacences du chromosome X d'*Anopheles funestus*



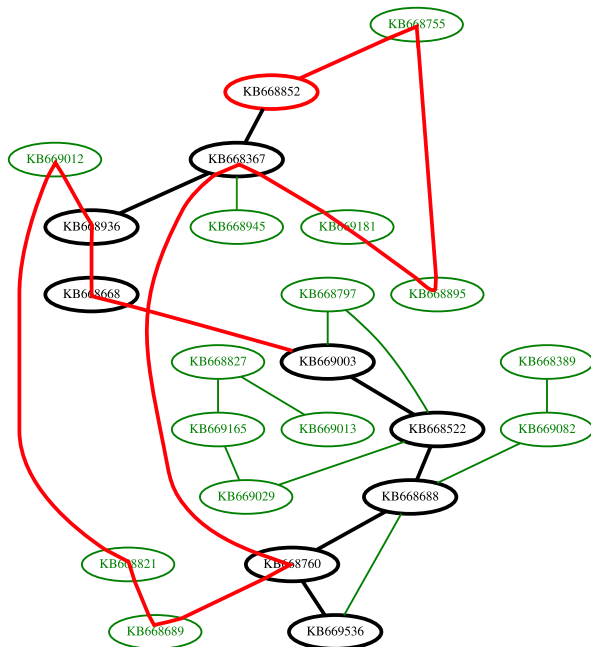
Graphe d'adjacences du chromosome X d'*Anopheles funestus*



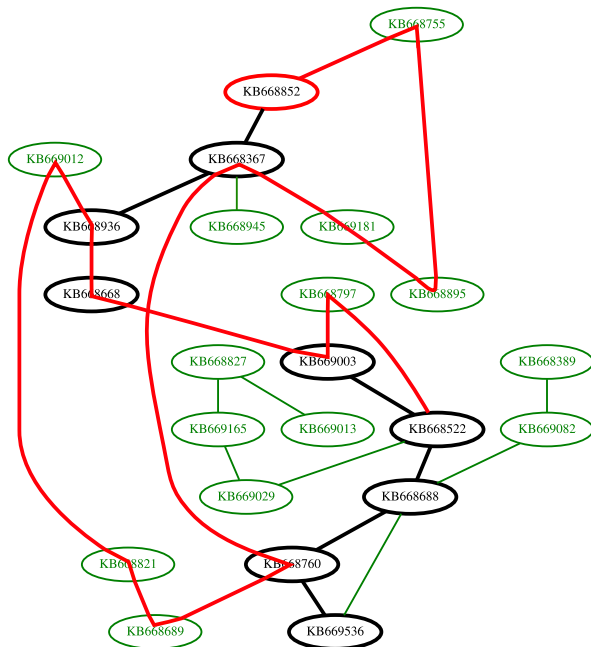
Graphe d'adjacences du chromosome X d'*Anopheles funestus*



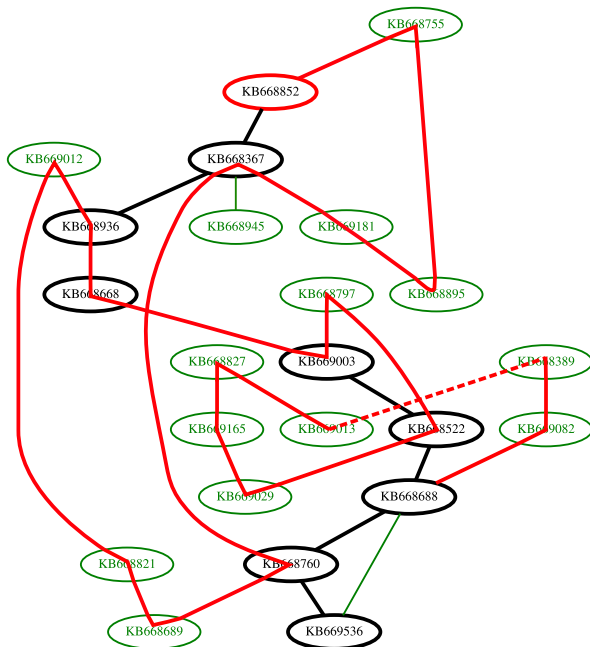
Graphe d'adjacences du chromosome X d'*Anopheles funestus*



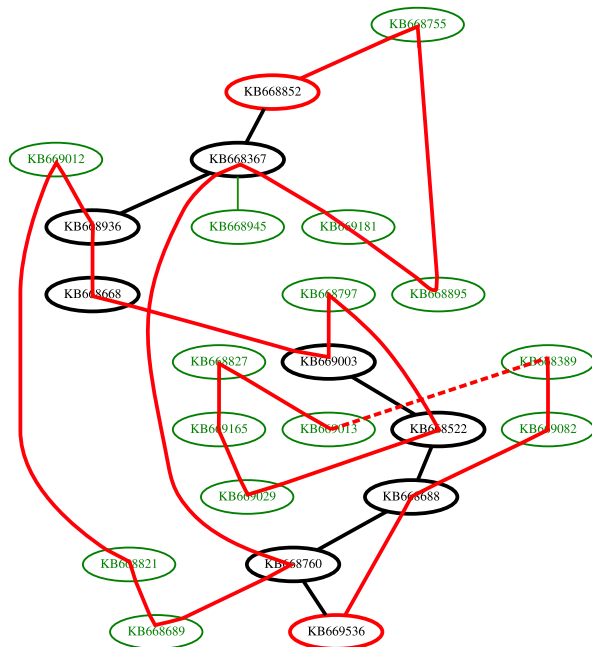
Grphe d'adjacences du chromosome X d'*Anopheles funestus*



Graphe d'adjacences du chromosome X d'*Anopheles funestus*

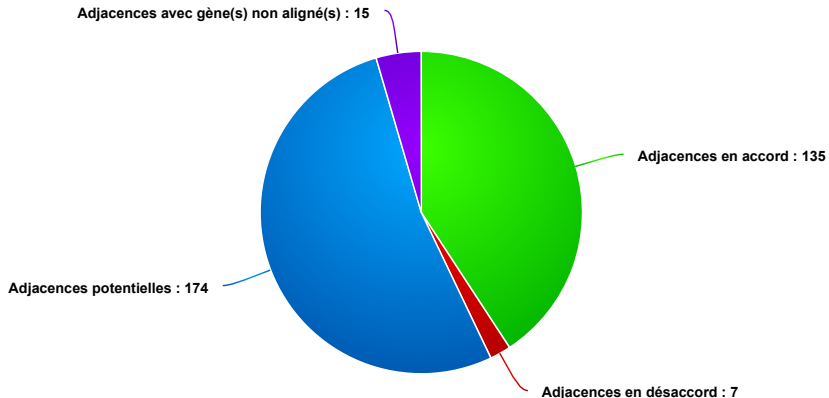


Graphe d'adjacences du chromosome X d'*Anopheles funestus*

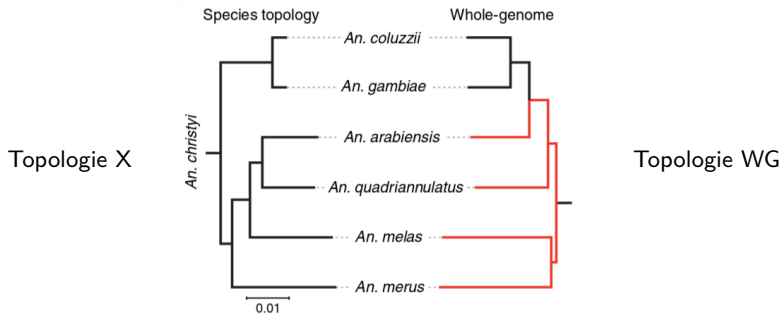


Analyse de 331 nouvelles adjacences sur des scaffolds PacBio

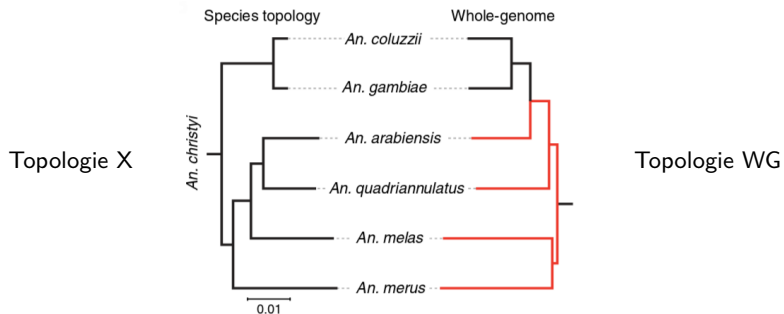
- ▶ *scaffolds* PacBio produits par le consortium *Anopheles*



Évaluation des 2 topologies (Fontaine et al., 2015)



Évaluation des 2 topologies (Fontaine et al., 2015)



- ▶ Choix de la topologie du X (hypothèse d'introgession) :
 - Comparaison noeuds ancestraux des arbres de gènes inférés sur le chr. X vs autosomes

Notre évaluation des 2 topologies par l'étude des réarrangements

Exécution de DECOSTAR avec les 2 topologies X et WG

+

Analyse des réarrangements

Notre évaluation des 2 topologies par l'étude des réarrangements

Hypothèses :

- ▶ Parcimonie : Événements de réarrangements sont rares

Topologie WG

Topologie X

Notre évaluation des 2 topologies par l'étude des réarrangements

Hypothèses :

- ▶ Parcimonie : Événements de réarrangements sont rares

Topologie WG

- ▶ #duplications : 6.159
- ▶ #(créations+cassures) : 3.337
- ▶ #réarrangements/adj : $9.15.10^{-3}$

Topologie X

- ▶ #duplications : 6.461
- ▶ #(créations+cassures) : 3.515
- ▶ #réarrangements/adj : $9.68.10^{-3}$

Notre évaluation des 2 topologies par l'étude des réarrangements

Hypothèses :

- ▶ Parcimonie : Événements de réarrangements sont rares
- ▶ La topologie correcte permet un meilleur assemblage des génomes

Topologie WG

- ▶ #duplications : 6.159
- ▶ #(créations+cassures) : 3.337
- ▶ #réarrangements/adj : $9.15.10^{-3}$

Topologie X

- ▶ #duplications : 6.461
- ▶ #(créations+cassures) : 3.515
- ▶ #réarrangements/adj : $9.68.10^{-3}$

Notre évaluation des 2 topologies par l'étude des réarrangements

Hypothèses :

- ▶ Parcimonie : Événements de réarrangements sont rares
- ▶ La topologie correcte permet un meilleur assemblage des génomes

Topologie WG

- ▶ #duplications : 6.159
- ▶ #(créations+cassures) : 3.337
- ▶ #réarrangements/adj : $9.15.10^{-3}$
- ▶ μ (#scaffolds actuels) : 749
- ▶ μ (#scaffolds ancestraux) : 1.670
- ▶ Σ (conflits synténiques) : 6.319

Topologie X

- ▶ #duplications : 6.461
- ▶ #(créations+cassures) : 3.515
- ▶ #réarrangements/adj : $9.68.10^{-3}$
- ▶ μ (#scaffolds actuels) : 751
- ▶ μ (#scaffolds ancestraux) : 1.771
- ▶ Σ (conflits synténiques) : 7.655

Notre évaluation des 2 topologies par l'étude des réarrangements

Hypothèses :

- ▶ Parcimonie : Événements de réarrangements sont rares
- ▶ La topologie correcte permet un meilleur assemblage des génomes

Topologie WG

- ▶ #duplications : 6.159
- ▶ #(créations+cassures) : 3.337
- ▶ #réarrangements/adj : $9.15.10^{-3}$
- ▶ μ (#scaffolds actuels) : 749
- ▶ μ (#scaffolds ancestraux) : 1.670
- ▶ Σ (conflits synténiques) : 6.319

Topologie X

- ▶ #duplications : 6.461
- ▶ #(créations+cassures) : 3.515
- ▶ #réarrangements/adj : $9.68.10^{-3}$
- ▶ μ (#scaffolds actuels) : 751
- ▶ μ (#scaffolds ancestraux) : 1.771
- ▶ Σ (conflits synténiques) : 7.655

Topologie des espèces du point de vue des réarrangements :
topologie WG

Plan de la soutenance

Introduction - Contexte historique

Développements algorithmiques

Résultats sur l'assemblage et l'évolution de génomes de moustiques

Conclusion & perspectives

Conclusion

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique
- ▶ Thématique en plein développement :
 - GOS-ASM (2016)
 - RACA (2013) + DESCHRAMBLER (2017)

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique
- ▶ Thématique en plein développement :
 - GOS-ASM (2016)
 - RACA (2013) + DESCHRAMBLER (2017)
- ▶ DECoSTAR :
 - Signal phylogénétique

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique
- ▶ Thématique en plein développement :
 - GOS-ASM (2016)
 - RACA (2013) + DESCHRAMBLER (2017)
- ▶ DECoSTAR :
 - Signal phylogénétique
 - Multiple référence

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique
- ▶ Thématique en plein développement :
 - GOS-ASM (2016)
 - RACA (2013) + DESCHRAMBLER (2017)
- ▶ DECoSTAR :
 - Signal phylogénétique
 - Multiple référence
 - Chaque génome est une cible et une référence
 - Références peuvent être fragmentées

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique
- ▶ Thématique en plein développement :
 - GOS-ASM (2016)
 - RACA (2013) + DESCHRAMBLER (2017)
- ▶ DECoSTAR :
 - Signal phylogénétique
 - Multiple référence
 - Chaque génome est une cible et une référence
 - Références peuvent être fragmentées
 - Génomes de grandes tailles

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique
- ▶ Thématique en plein développement :
 - GOS-ASM (2016)
 - RACA (2013) + DESCHRAMBLER (2017)
- ▶ DECoSTAR :
 - Signal phylogénétique
 - Multiple référence
 - Chaque génome est une cible et une référence
 - Références peuvent être fragmentées
 - Génomes de grandes tailles
 - insertions/délétions, duplications et transferts de gènes

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique
- ▶ Thématique en plein développement :
 - GOS-ASM (2016)
 - RACA (2013) + DESCHRAMBLER (2017)
- ▶ DECoSTAR :
 - Signal phylogénétique
 - Multiple référence
 - Chaque génome est une cible et une référence
 - Références peuvent être fragmentées
 - Génomes de grandes tailles
 - insertions/délétions, duplications et transferts de gènes
 - Adjacences avec probabilité *a priori* (*reads* appariés)

Association de 2 problèmes dans une même approche

- ▶ Méthode de reconstruction conjointe de l'ordre des gènes de génomes actuels et ancestraux dans un cadre phylogénétique
- ▶ Thématique en plein développement :
 - GOS-ASM (2016)
 - RACA (2013) + DESCHRAMBLER (2017)
- ▶ DECoSTAR :
 - Signal phylogénétique
 - Multiple référence
 - Chaque génome est une cible et une référence
 - Références peuvent être fragmentées
 - Génomes de grandes tailles
 - insertions/délétions, duplications et transferts de gènes
 - Adjacences avec probabilité *a priori* (*reads* appariés)
 - support *a posteriori* aux adjacences inférées

Perspectives

Étude de l'assemblage et de l'évolution structurale des génomes d'oiseaux

- ▶ Collaboration avec B. Nabholz et C. Scornavacca (ISE-M)
- ▶ 27 génomes d'oiseaux → proposition d'assemblages améliorés
- ▶ Comparaisons génomes ancestraux des oiseaux reconstruits par différentes méthodes + papier Romanov et al., 2014 (FISH+MGRA)

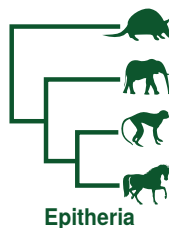
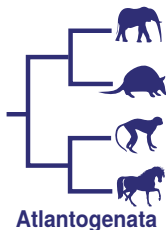
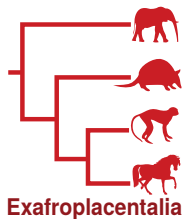
Perspectives

Étude de l'assemblage et de l'évolution structurale des génomes d'oiseaux

- ▶ Collaboration avec B. Nabholz et C. Scornavacca (ISE-M)
- ▶ 27 génomes d'oiseaux → proposition d'assemblages améliorés
- ▶ Comparaisons génomes ancestraux des oiseaux reconstruits par différentes méthodes + papier Romanov et al., 2014 (FISH+MGRA)

Racine de la phylogénie des mammifères

- ▶ 3 scénarios débattus : Exafroplacentalia vs Atlantogenata vs Epitheria
- ▶ Tester les 3 topologies du point de vue des réarrangements chromosomiques



Perspectives

Étude de l'assemblage et de l'évolution structurale des génomes d'oiseaux

- ▶ Collaboration avec B. Nabholz et C. Scornavacca (ISE-M)
- ▶ 27 génomes d'oiseaux → proposition d'assemblages améliorés
- ▶ Comparaisons génomes ancestraux des oiseaux reconstruits par différentes méthodes + papier Romanov et al., 2014 (FISH+MGRA)

Racine de la phylogénie des mammifères

- ▶ 3 scénarios débattus : Exafroplacentelia vs Atlantogenata vs Epitheria
- ▶ Tester les 3 topologies du point de vue des réarrangements chromosomiques

Application pour l'assemblage de génomes anciens

- ▶ Séquençage de génomes d'espèces disparues (ADN ancien)
- ▶ ADN ancien pour valider des génomes ancestraux (ex : Y. pestis, Duchemin et al., 2015)
- ▶ Méthodes basées sur la reconstruction de génomes ancestraux pour assembler des génomes anciens :
 - FPSAC
 - PHYSCA
 - DECoSTAR ?

Membres du jury

- ▶ Christophe DESSIMOZ
- ▶ Gilles FISCHER (rapporteur)
- ▶ Dominique MOUCHIROUD
- ▶ Aïda OUANGRAOUA (rapporteuse)

Membres du jury

- ▶ Christophe DESSIMOZ
- ▶ Gilles FISCHER (rapporteur)
- ▶ Dominique MOUCHIROUD
- ▶ Aïda OUANGRAOUA (rapporteuse)

SUPERvisors

- ▶ Sèverine BÉRARD (co-directrice)
- ▶ Éric TANNIER (co-directeur)
- ▶ Cedric CHAUVE
- ▶ Annie CHATEAU

Membres du jury

- ▶ Christophe DESSIMOZ
- ▶ Gilles FISCHER (rapporteur)
- ▶ Dominique MOUCHIROUD
- ▶ Aïda OUANGRAOUA (rapporteuse)

*SUPER*visors

- ▶ Sèverine BÉRARD (co-directrice)
- ▶ Éric TANNIER (co-directeur)
- ▶ Cedric CHAUVE
- ▶ Annie CHATEAU

ISE-M



- ▶ équipe Phylogénie et Évolution Moléculaire
- ▶ cluster ISE-M/MBB
- ▶ doctorants/post-docs

LBBE



- ▶ équipe Bioinfo, Phylogénie et Génétique Évolutive

Membres du jury

- ▶ Christophe DESSIMOZ
- ▶ Gilles FISCHER (rapporteur)
- ▶ Dominique MOUCHIROUD
- ▶ Aïda OUANGRAOUA (rapporteuse)

*SUPER*visors

- ▶ Sèverine BÉRARD (co-directrice)
- ▶ Éric TANNIER (co-directeur)
- ▶ Cedric CHAUVE
- ▶ Annie CHATEAU

ISE-M



- ▶ équipe Phylogénie et Évolution Moléculaire
- ▶ cluster ISE-M/MBB
- ▶ doctorants/post-docs

LBBE



- ▶ équipe Bioinfo, Phylogénie et Génomique Évolutive

Financements



- ▶ ANR Ancestrome
- ▶ Campus France-Mitacs (collab. C. CHAUVE)

Membres du jury

- ▶ Christophe DESSIMOZ
- ▶ Gilles FISCHER (rapporteur)
- ▶ Dominique MOUCHIROUD
- ▶ Aïda OUANGRAOUA (rapporteuse)

*SUPER*visors

- ▶ Sèverine BÉRARD (co-directrice)
- ▶ Éric TANNIER (co-directeur)
- ▶ Cedric CHAUVE
- ▶ Annie CHATEAU

Famille

ISE-M



- ▶ équipe Phylogénie et Évolution Moléculaire
- ▶ cluster ISE-M/MBB
- ▶ doctorants/post-docs

LBBE



- ▶ équipe Bioinfo, Phylogénie et Génomique Évolutive

Financements



- ▶ ANR Ancestrome
- ▶ Campus France-Mitacs (collab. C. CHAUVE)

Membres du jury

- ▶ Christophe DESSIMOZ
- ▶ Gilles FISCHER (rapporteur)
- ▶ Dominique MOUCHIROUD
- ▶ Aïda OUANGRAOUA (rapporteuse)

*SUPER*visors

- ▶ Sèverine BÉRARD (co-directrice)
- ▶ Éric TANNIER (co-directeur)
- ▶ Cedric CHAUVE
- ▶ Annie CHATEAU

Famille

Belle-famille

ISE-M



- ▶ équipe Phylogénie et Évolution Moléculaire
- ▶ cluster ISE-M/MBB
- ▶ doctorants/post-docs

LBBE



- ▶ équipe Bioinfo, Phylogénie et Génétique Évolutive

Financements



- ▶ ANR Ancestrome
- ▶ Campus France-Mitacs (collab. C. CHAUVE)

Membres du jury

- ▶ Christophe DESSIMOZ
- ▶ Gilles FISCHER (rapporteur)
- ▶ Dominique MOUCHIROUD
- ▶ Aïda OUANGRAOUA (rapporteuse)

*SUPER*visors

- ▶ Sèverine BÉRARD (co-directrice)
- ▶ Éric TANNIER (co-directeur)
- ▶ Cedric CHAUVE
- ▶ Annie CHATEAU

Famille

Belle-famille

ISE-M



- ▶ équipe Phylogénie et Évolution Moléculaire
- ▶ cluster ISE-M/MBB
- ▶ doctorants/post-docs

Amis

LBBE



- ▶ équipe Bioinfo, Phylogénie et Génomique Évolutive

Financements



- ▶ ANR Ancestrome
- ▶ Campus France-Mitacs (collab. C. CHAUVE)

Membres du jury

- ▶ Christophe DESSIMOZ
- ▶ Gilles FISCHER (rapporteur)
- ▶ Dominique MOUCHIROUD
- ▶ Aïda OUANGRAOUA (rapporteuse)

SUPERvisors

- ▶ Sèverine BÉRARD (co-directrice)
- ▶ Éric TANNIER (co-directeur)
- ▶ Cedric CHAUVE
- ▶ Annie CHATEAU

Famille

Belle-famille

ISE-M



- ▶ équipe Phylogénie et Évolution Moléculaire
- ▶ cluster ISE-M/MBB
- ▶ doctorants/post-docs

Amis

Aude

LBBE



- ▶ équipe Bioinfo, Phylogénie et Génomique Évolutive

Financements



- ▶ ANR Ancestrome
- ▶ Campus France-Mitacs (collab. C. CHAUVE)