Alignement, la spécificité d'abord !

Mikaël Salson

Équipe Bonsai, LIFL (Université Lille 1 et CNRS) – Inria

Les séquenceurs à haut débit

Les séquenceurs à haut débit

Roche 454

Illumina Hi-Seq

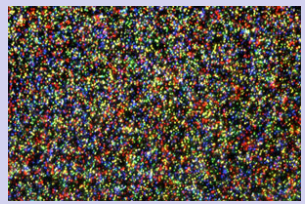
Ion Torrent

Solid

Les séquenceurs à haut débit

Roche 454

Illumina Hi-Seq



NIH

Ion Torrent

Solid

Les séquenceurs à haut débit

Roche 454

Haut débit
($\simeq$ 1 Gpb)

Ion Torrent

Illumina Hi-Seq

Très haut débit
($>$ 100 Gpb)

Solid

Les séquenceurs à haut débit

Roche 454

Moins de
une journée

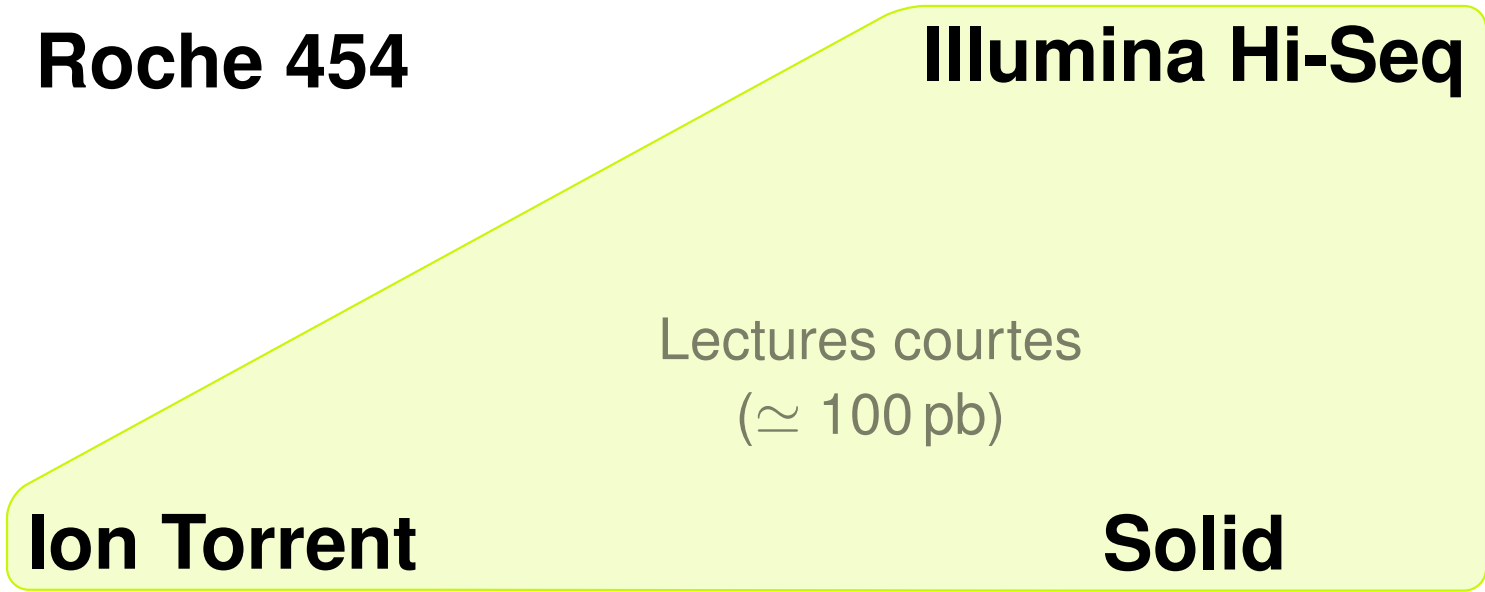
Ion Torrent

Illumina Hi-Seq

Plusieurs jours

Solid

Les séquenceurs à haut débit



Les séquenceurs à haut débit

Roche 454

Homopolymères

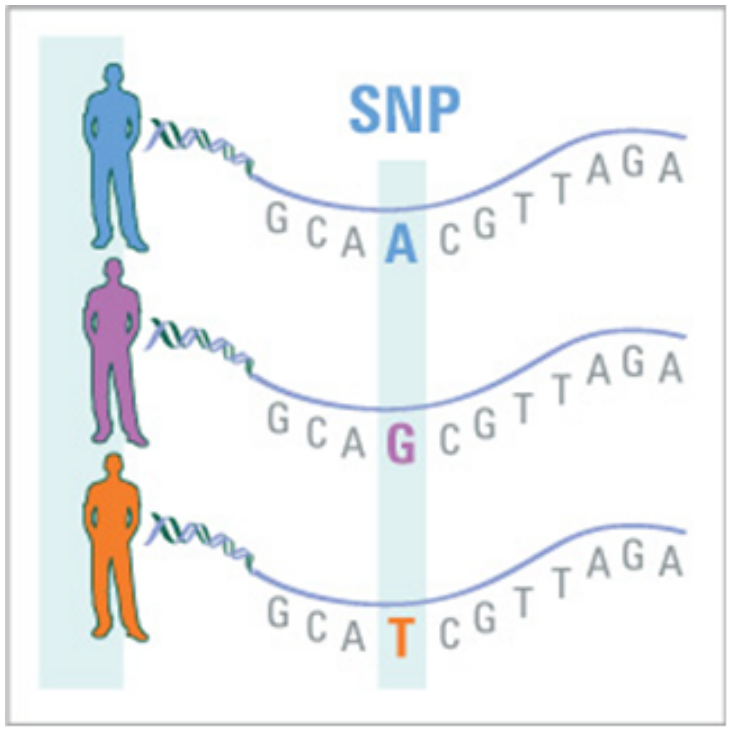
Ion Torrent

Illumina Hi-Seq

Substitutions

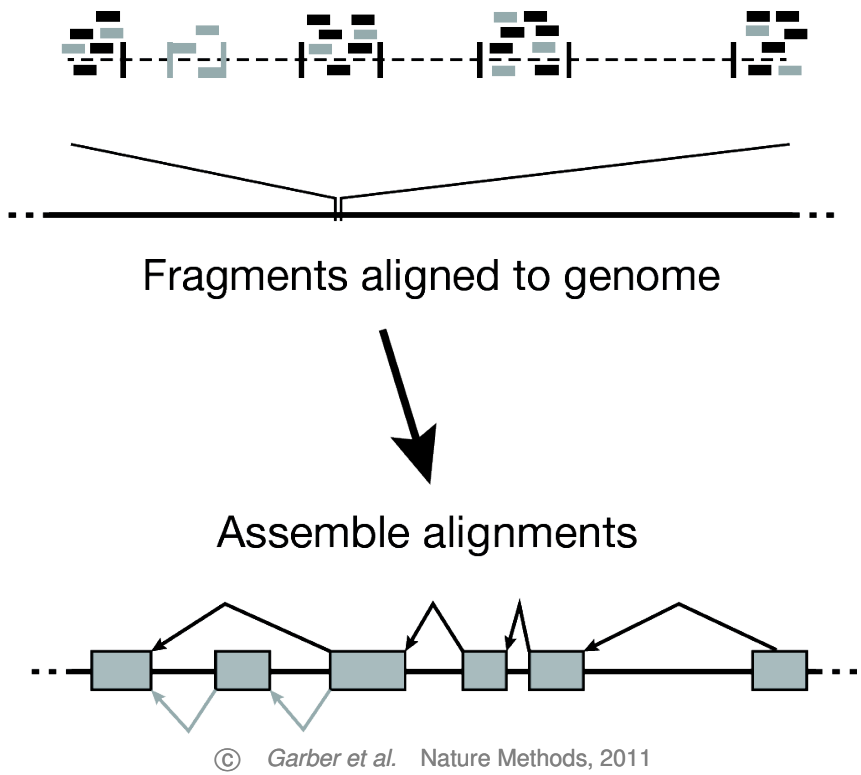
Solid

Le séquençage pour quoi faire ?

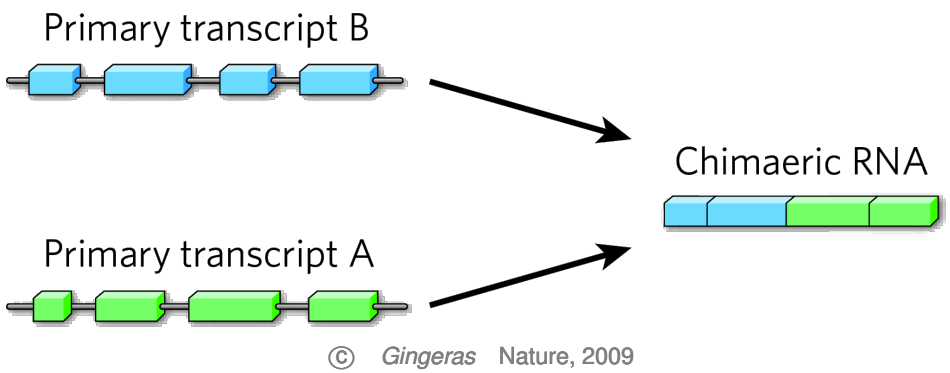


© *Lauren Solomon* Broad institute

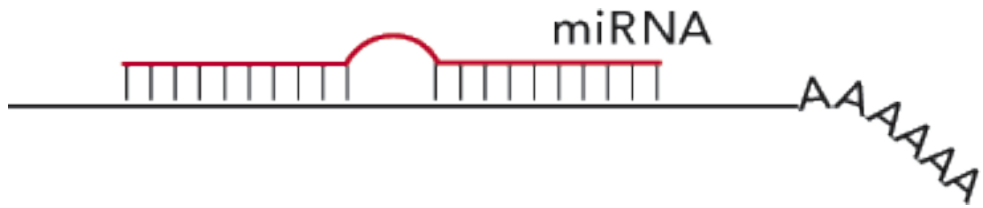
Le séquençage pour quoi faire ?



Le séquençage pour quoi faire ?

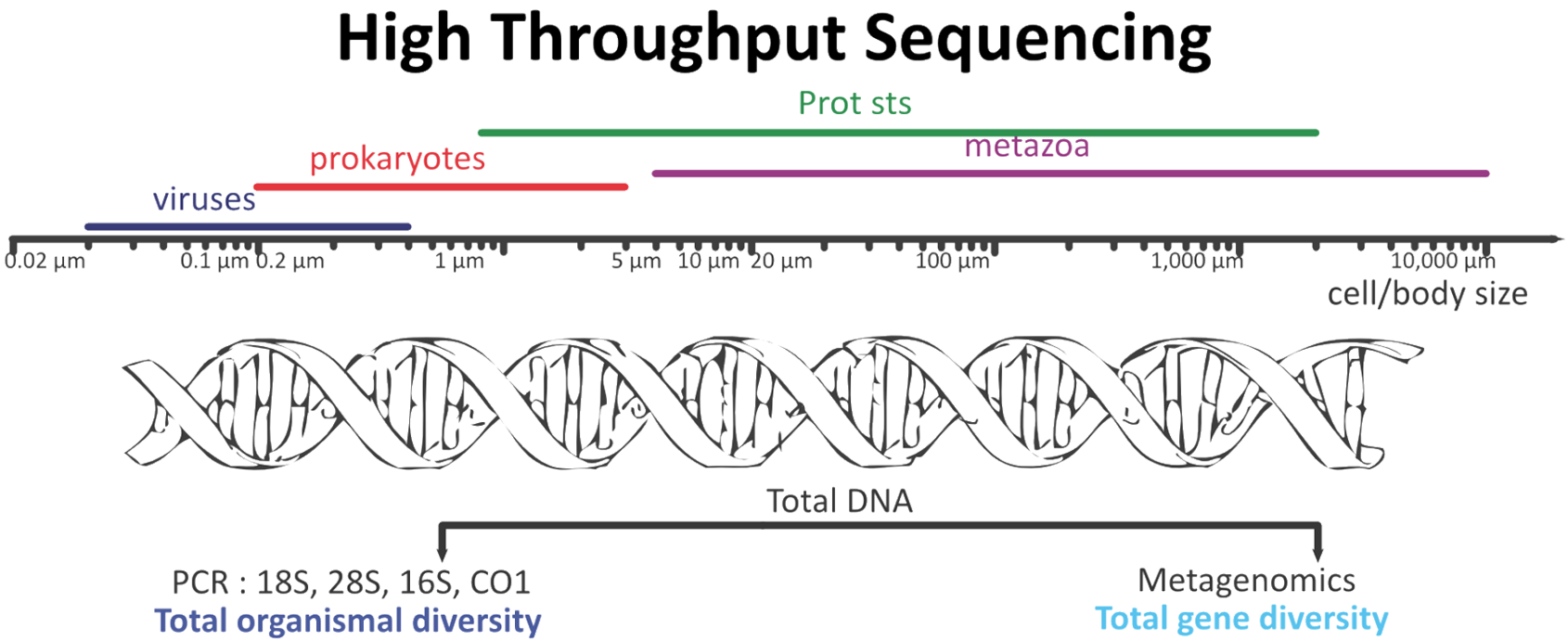


Le séquençage pour quoi faire ?

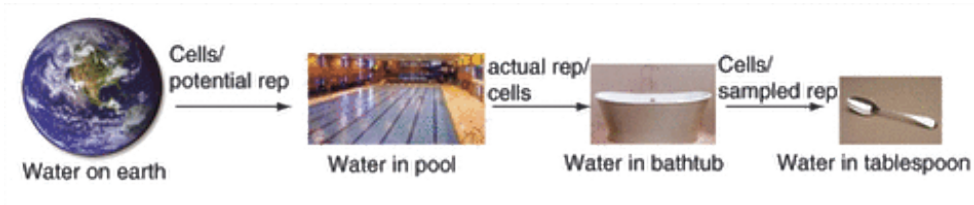


© *Carin Cain, Victor Ambros* Lasker foundation

Le séquençage pour quoi faire ?

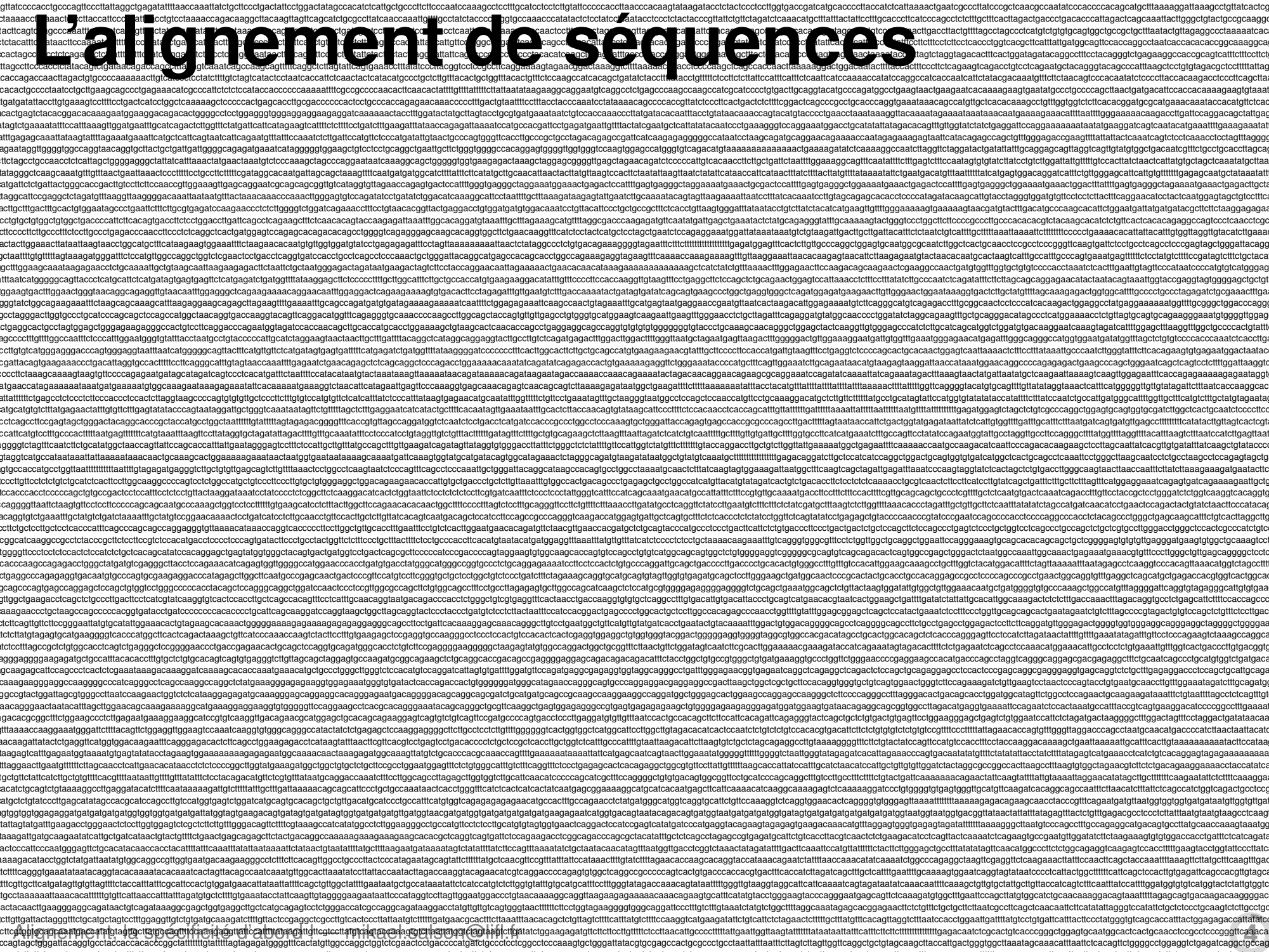


Le séquençage pour quoi faire ?

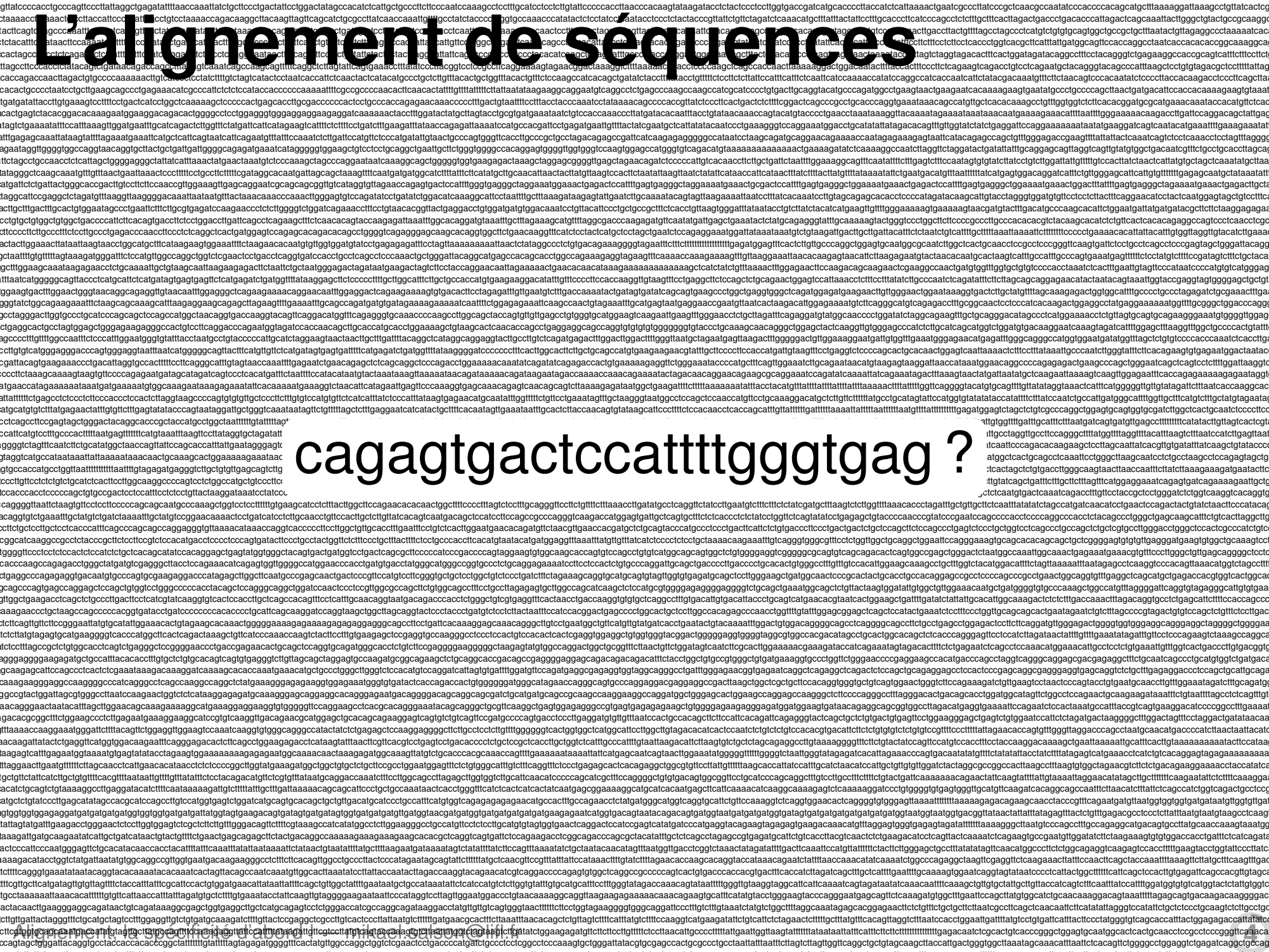


© *Benichou et al* Immunology, 2012

L’alignement de séquences


L'alignement de séquences

L'alignement de séquences



L'alignement de séquences

cagagtgactccattttgggtgag ?

Recherche exacte

Knuth-Morris-Pratt

Boyer-Moore

Arbre des suffixes

Recherche exacte

Knuth-Morris-Pratt

Boyer-Moore

Années 1970

Arbre des suffixes

L'alignement de séquences

tgggcacctctatcttccg ?

L'alignement de séquences

tgggcacctctatcttccg ?

Recherche avec erreur(s)

Programmation dynamique

Shift-or

Recherche avec erreur(s)

Programmation dynamique

Shift-or

Années 1960–1980

Un problème compliqué ?

Knuth-Morris-Pratt

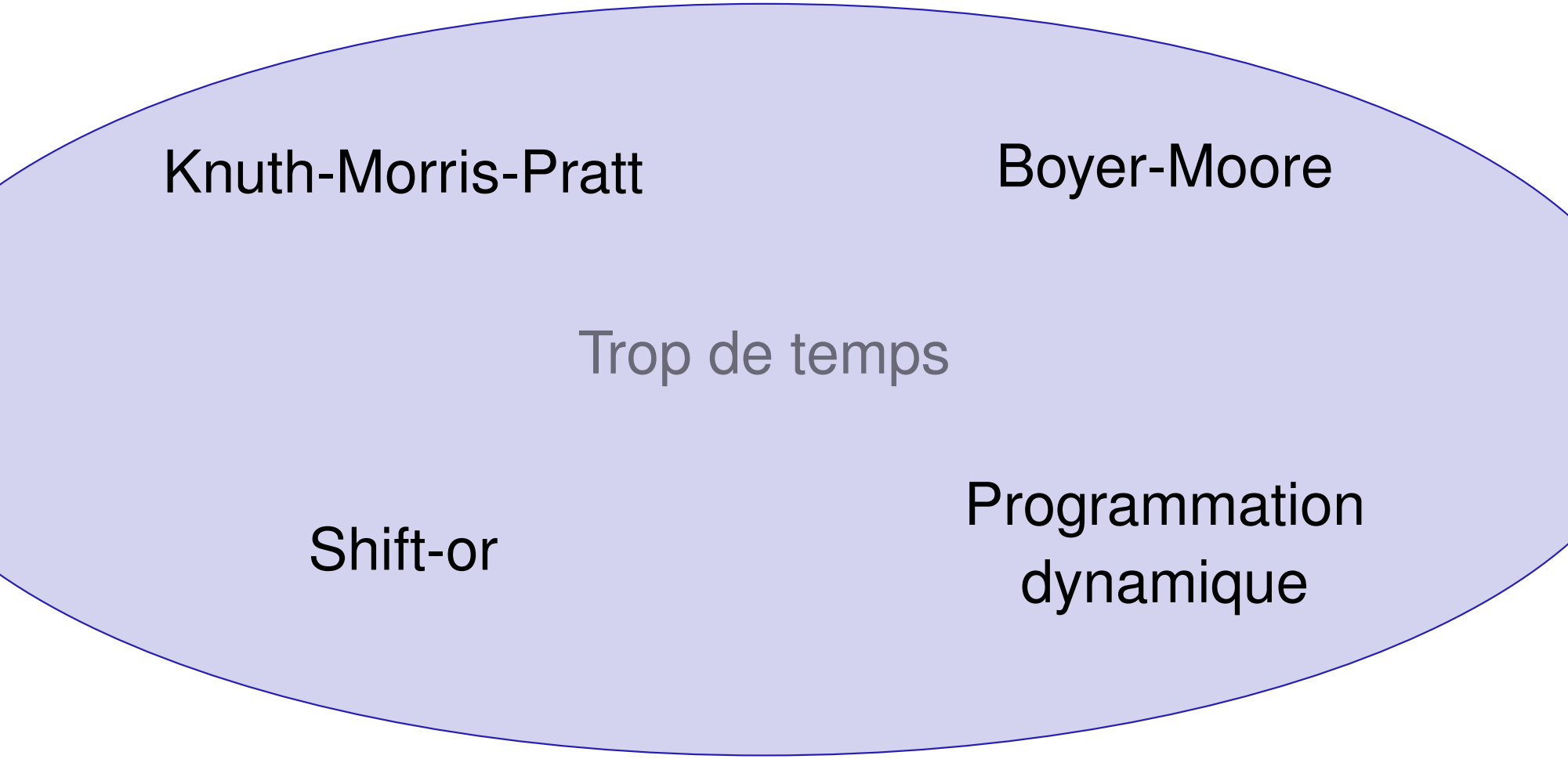
Boyer-Moore

Shift-or

Programmation dynamique

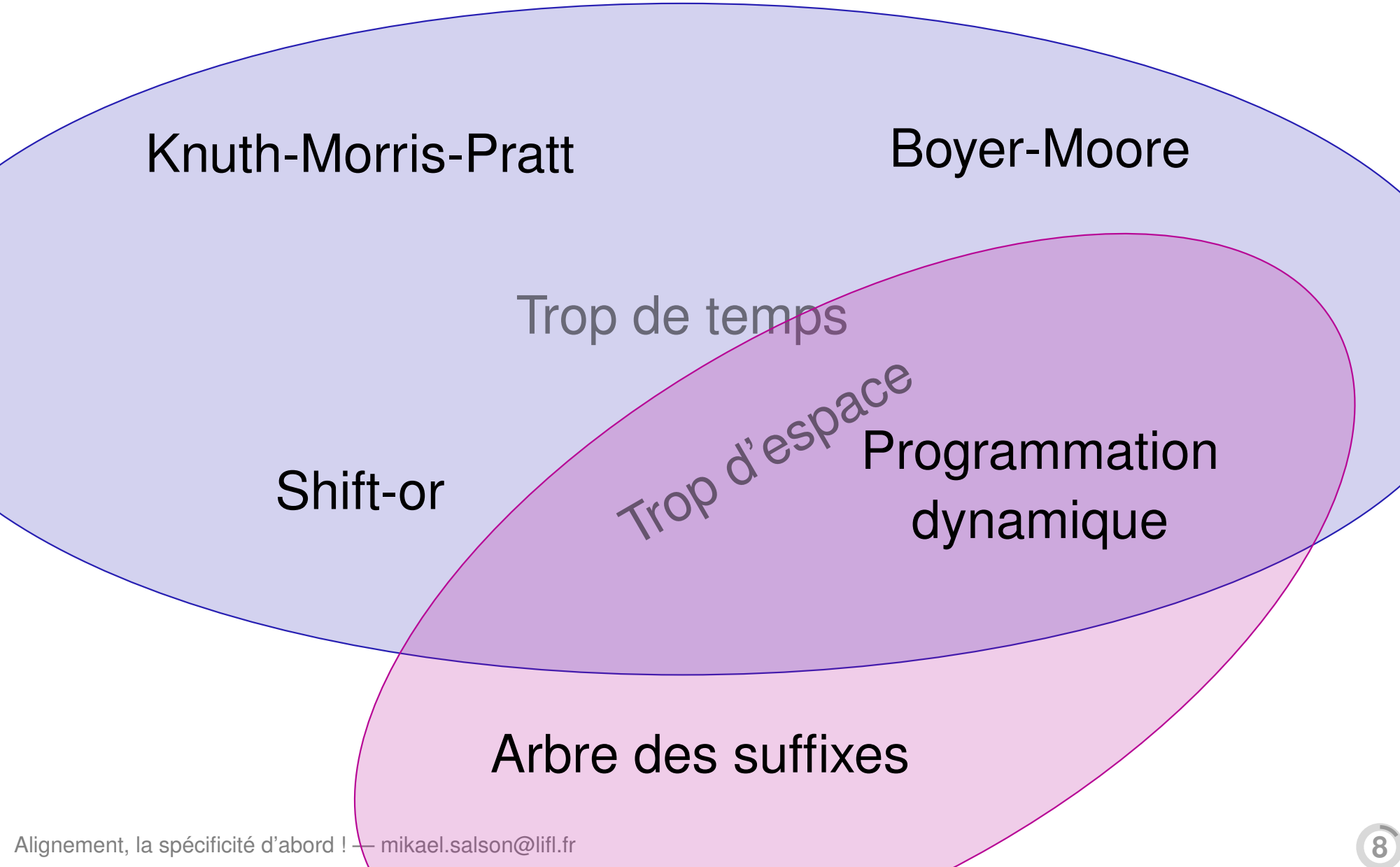
Arbre des suffixes

Un problème compliqué ?



Arbre des suffixes

Un problème compliqué ?



Un problème compliqué ?

Beaucoup de données

Spécificités du protocole et du séquençage

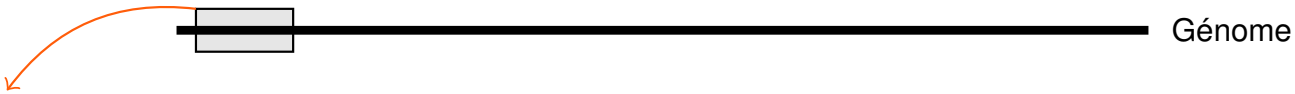
Graines

Graines

Génome

Graines

Graines



Graines

Graines

Graines

Graines

Graines

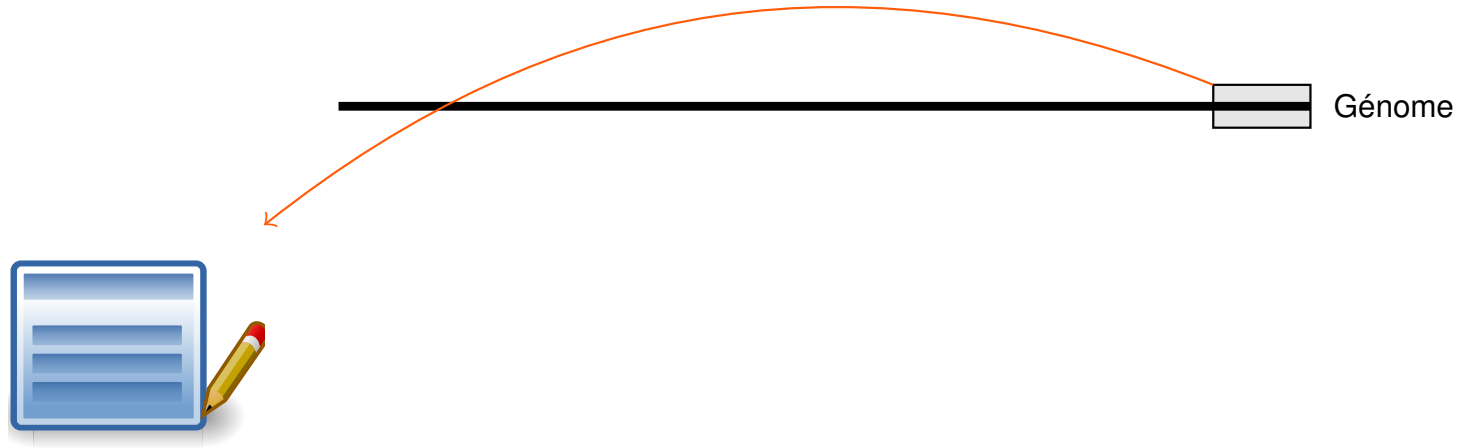
Graines

Graines

Graines

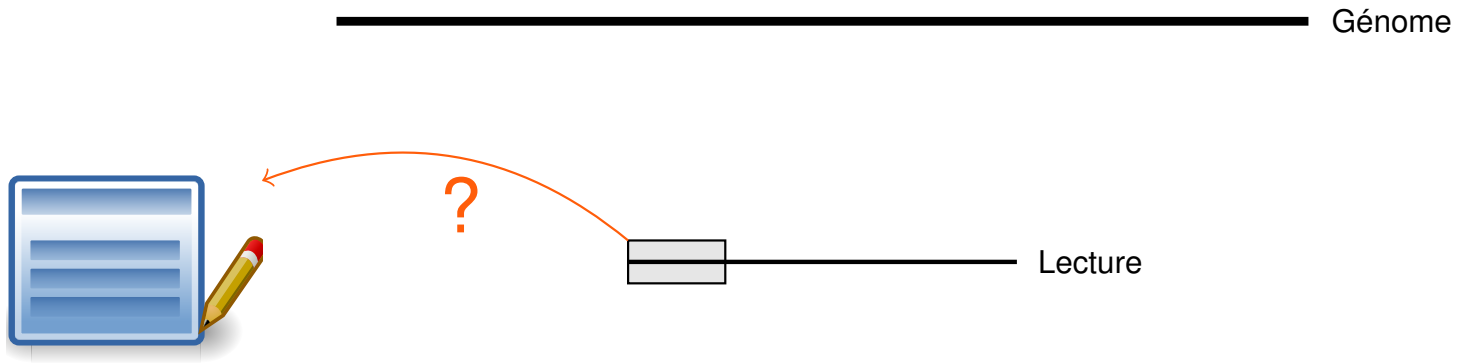
Graines

Graines

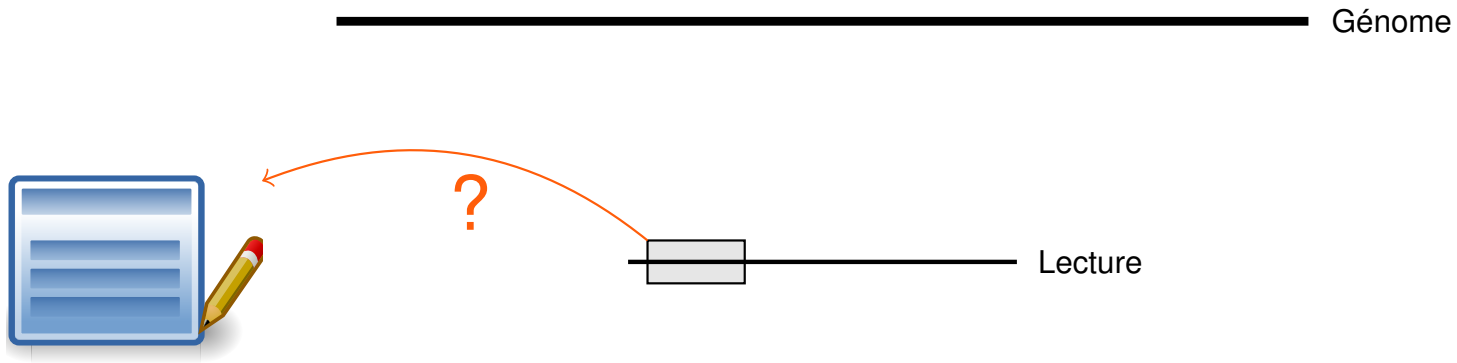


Graines

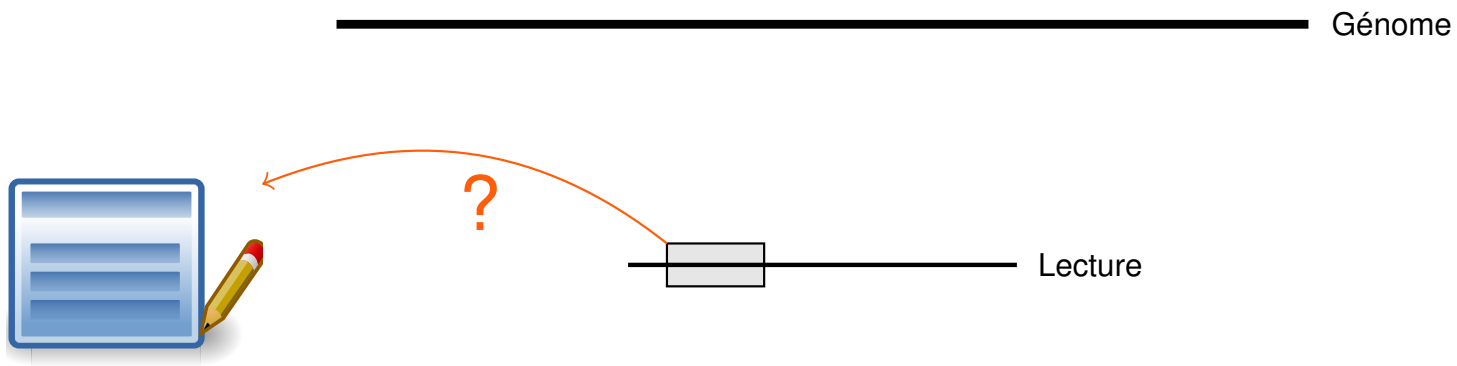
Graines



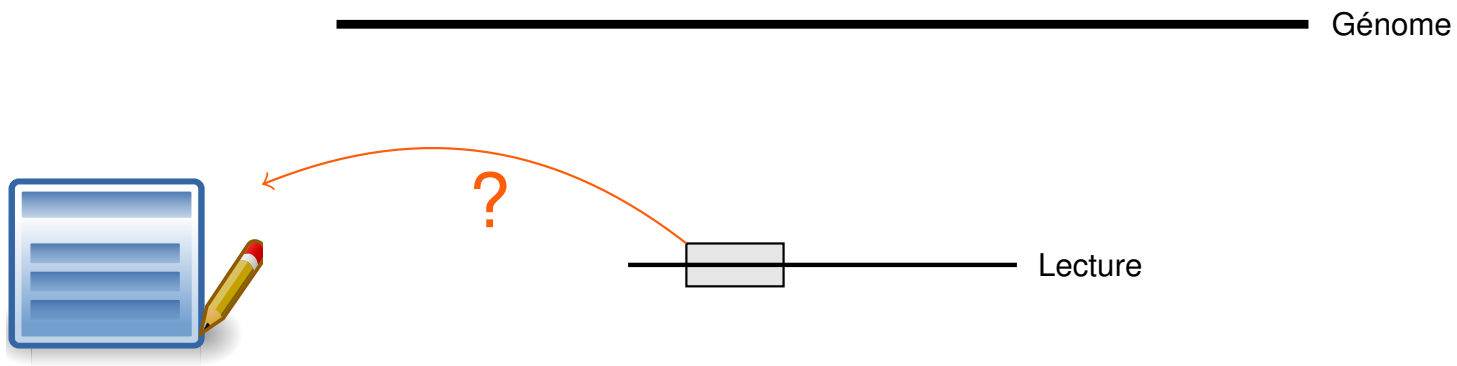
Graines



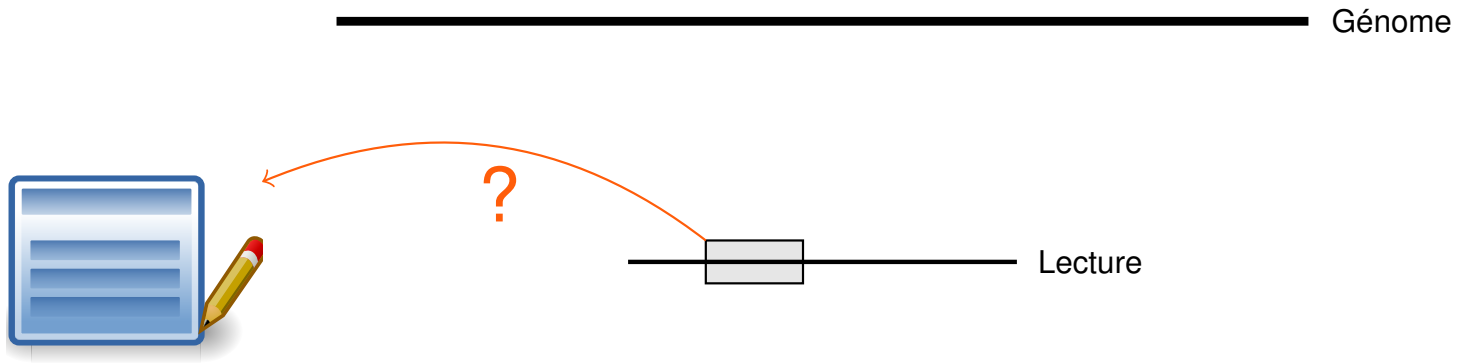
Graines



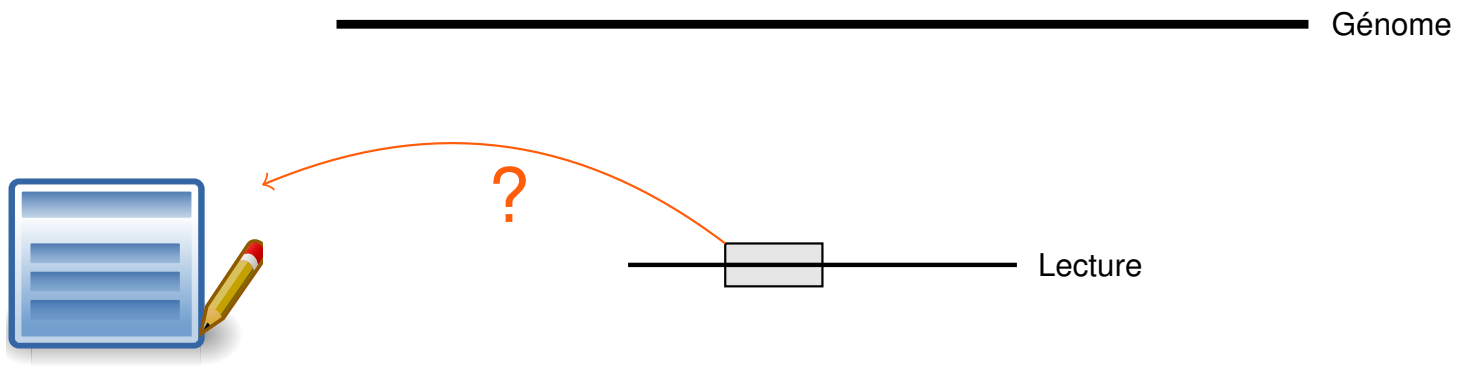
Graines



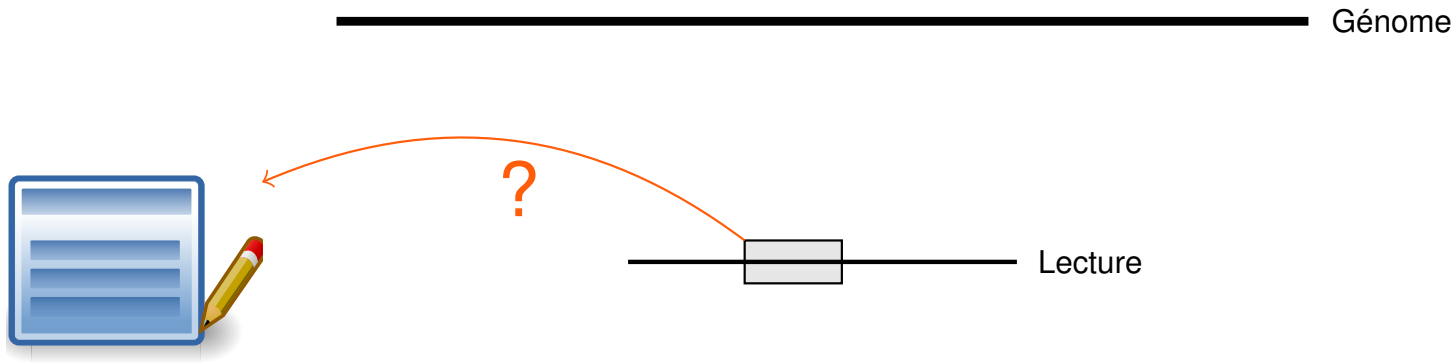
Graines



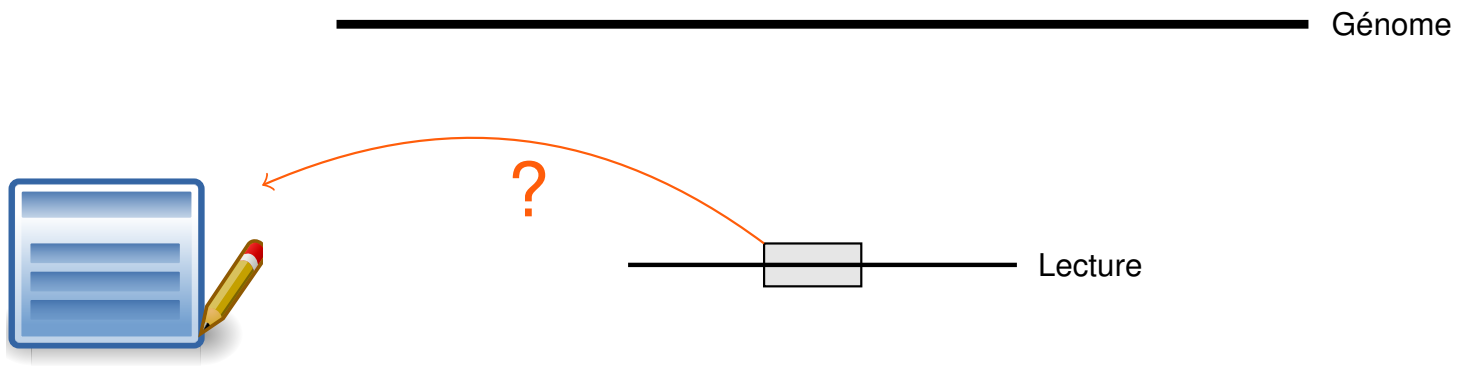
Graines



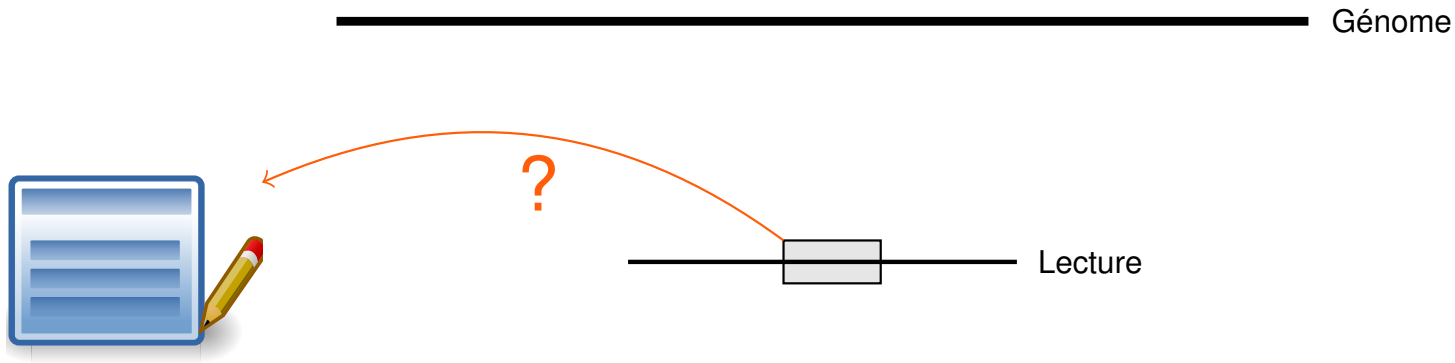
Graines



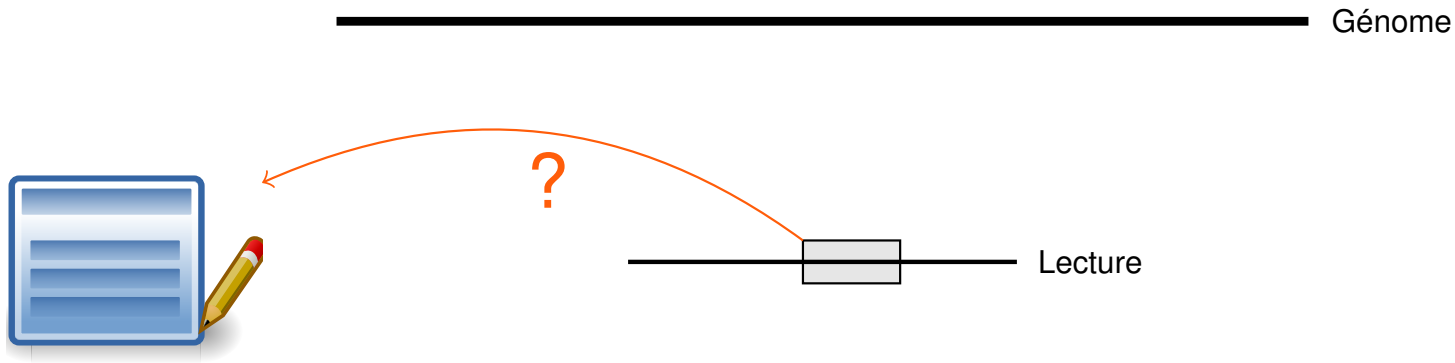
Graines



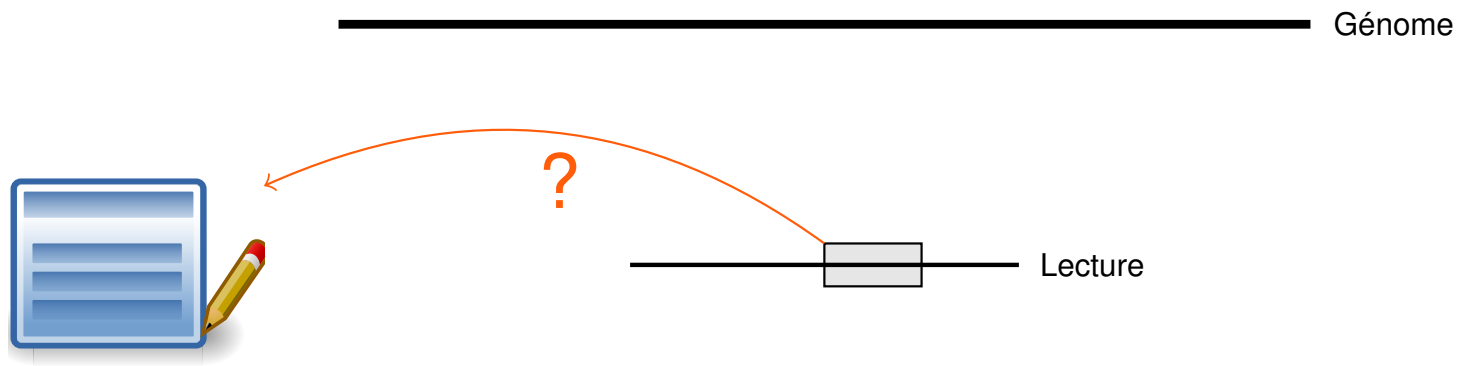
Graines



Graines

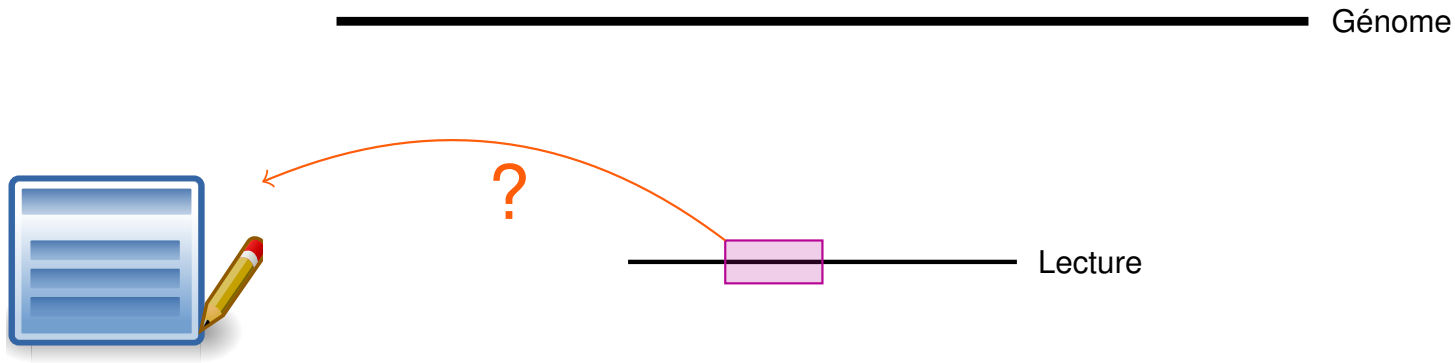


Graines

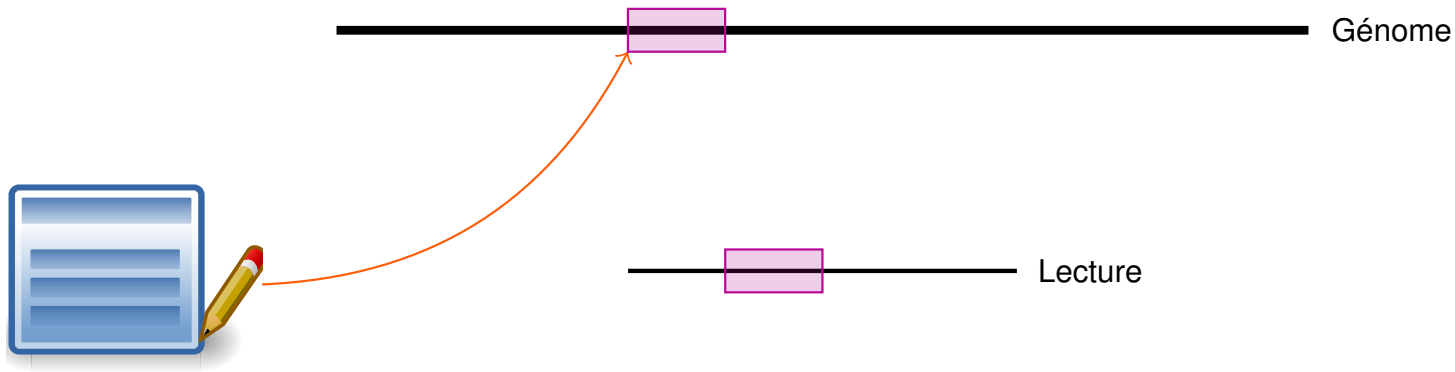


Graines

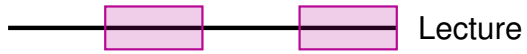
Graines



Graines



Graines

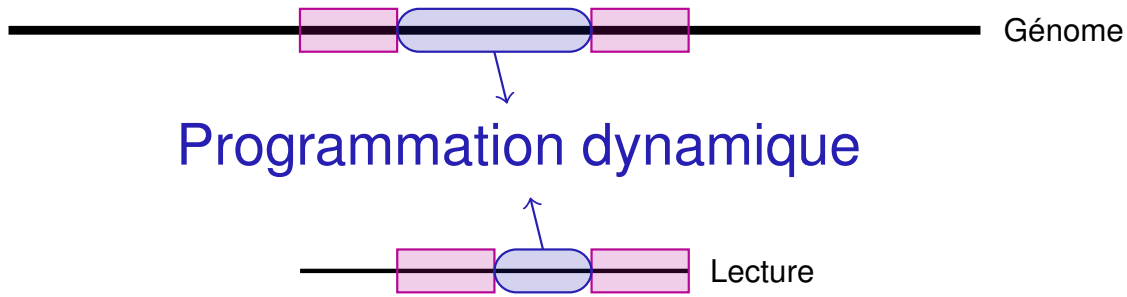


Graines

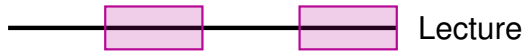
Génome

Lecture

Graines



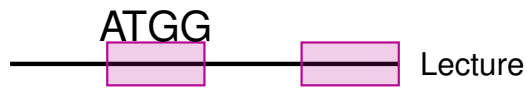
Graines



Graines contiguës

####

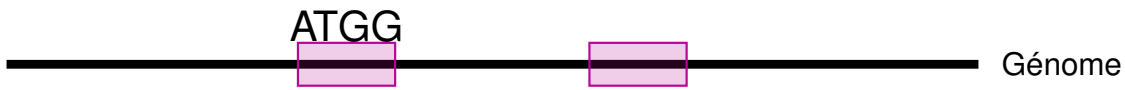
Graines

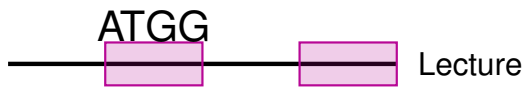


Graines contiguës

####

Graines

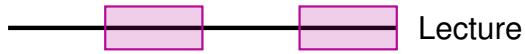




Graines contiguës

####

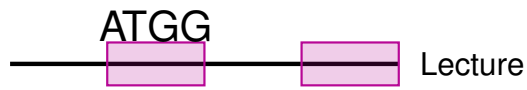
Graines



Graines espacées

##–#

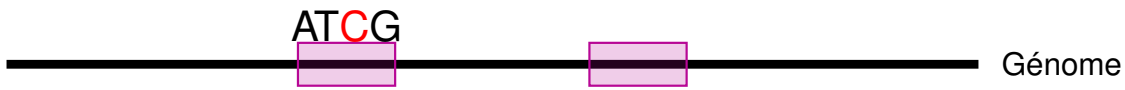
Graines

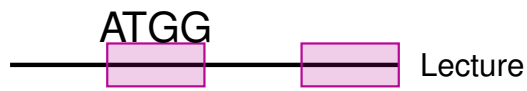


Graines espacées

##–#

Graines

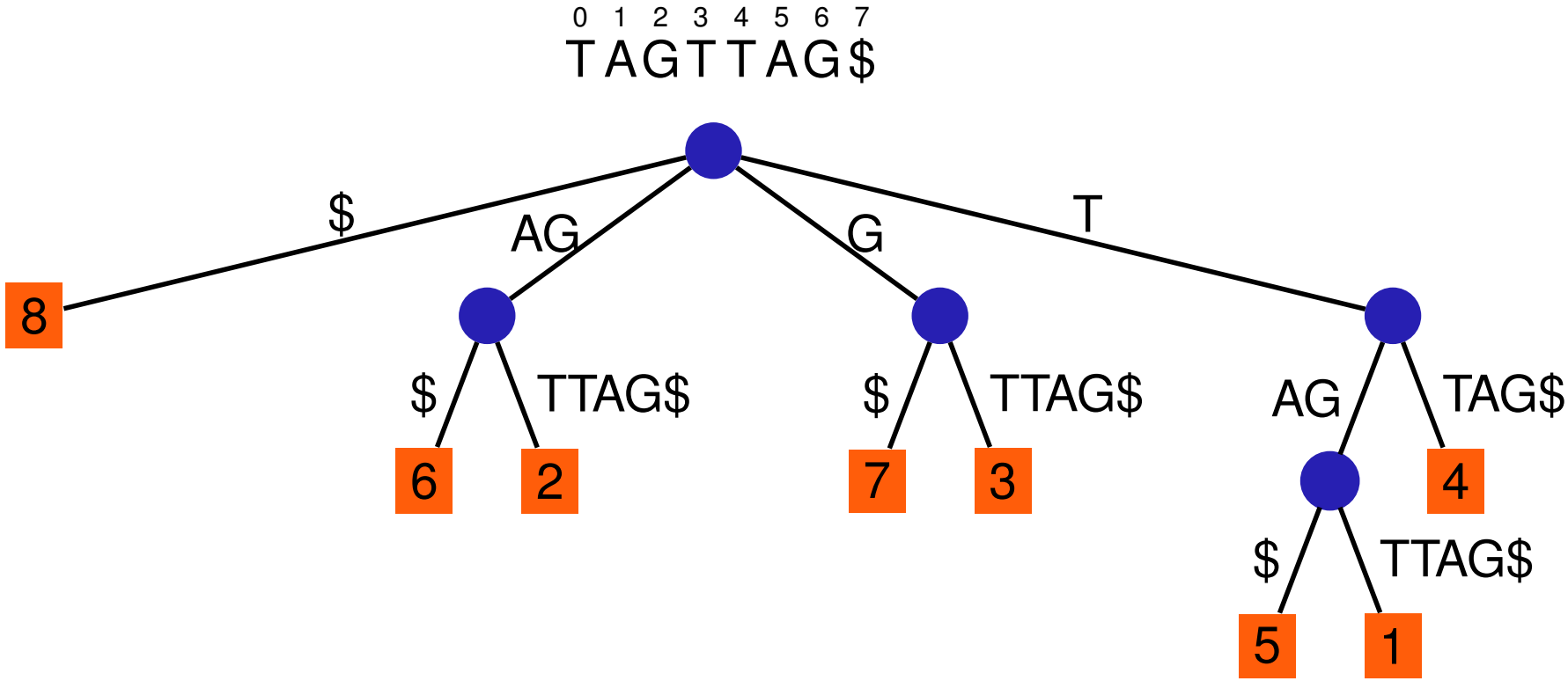




Graines espacées

##–#

Indexation de tout le texte



Les aligneurs

© *Fonseca et al* Bioinformatics, 2012

`http://wwwdev.ebi.ac.uk/fg/hts_mappers/`

Les aligneurs les plus utilisés

Bowtie

Blat

MAQ

BWA

TopHat

SOAP

SOAP2

Mummer3

BWA-SW

mrFAST

SHRiMP

Les aligneurs ayant le plus de citations par an, d'après Fonseca *et al*, Bioinformatics 2012.

Les aligneurs les plus utilisés

Bowtie

Blat

MAQ

BWA

TopHat

SOAP

SOAP2

Mummer3

BWA-SW

mrFAST

SHRiMP

Les aligneurs ayant le plus de citations par an, d'après Fonseca *et al*, Bioinformatics 2012.

Les aligneurs les plus utilisés

Bowtie

Blat

MAQ

BWA

TopHat

SOAP

SOAP2

Mummer3

BWA-SW

mrFAST

SHRiMP

BWT

Les aligneurs ayant le plus de citations par an, d'après Fonseca *et al*, Bioinformatics 2012.

Les aligneurs les plus utilisés

Bowtie

Blat

MAQ

BWA

TopHat

SOAP

SOAP2

Mummer3

BWA-SW

mrFAST

SHRiMP

BWT

Arbre des suffixes

Les aligneurs ayant le plus de citations par an, d'après Fonseca *et al*, Bioinformatics 2012.

Les aligneurs les plus utilisés

Bowtie

Blat

MAQ

BWA

TopHat

SOAP

SOAP2

Mummer3

BWA-SW

mrFAST

SHRiMP

BWT

Arbre des suffixes

Graines contiguës

Les aligneurs ayant le plus de citations par an, d'après Fonseca *et al*, Bioinformatics 2012.

Les aligneurs les plus utilisés

Bowtie

Blat

MAQ

BWA

TopHat

SOAP

SOAP2

Mummer3

BWA-SW

mrFAST

SHRiMP

BWT

Arbre des suffixes

Graines contiguës

Graines espacées

Les aligneurs ayant le plus de citations par an, d'après Fonseca *et al*, Bioinformatics 2012.

Les aligneurs les plus utilisés

Bowtie

Blat

MAQ

BWA

TopHat

SOAP

SOAP2

Mummer3

BWA-SW

mrFAST

SHRiMP

BWT

Arbre des suffixes

Graines contiguës

Graines espacées

Les aligneurs ayant le plus de citations par an, d'après Fonseca *et al*, Bioinformatics 2012.

Qu’est-ce qu’un bon alignement ?

Quelle utilisation ?

Qu’est-ce qu’un bon alignement ?

Quelle utilisation ?

Quel protocole ?

Qu’est-ce qu’un bon alignement ?

Quelle utilisation ?

Quel protocole ?

Quel séquenceur ?

Qu’est-ce qu’un bon alignement ?

Quelle utilisation ?

Quel protocole ?

Quel séquenceur ?

Quelles machines ?

Qu’est-ce qu’un bon alignement ?

Quelle utilisation ?

Quel protocole ?

Tout est relatif !

Quel séquenceur ?

Quelles machines ?

Qu’est-ce qu’un bon alignement ?

Pourcentage de lectures correctement alignées ?

Qu’est-ce qu’un bon alignement ?

Pourcentage de lectures correctement alignées ?

Longueur ?

Qu’est-ce qu’un bon alignement ?

Pourcentage de lectures correctement alignées ?

Longueur ?

Génome ?

Qu’est-ce qu’un bon alignement ?

Pourcentage de lectures correctement alignées ?

Longueur ?

Génome ?

Quantité ?

Qu’est-ce qu’un bon alignement ?

Pourcentage de lectures correctement alignées ?

Longueur ?

Génome ?

Quantité ?

Paires ?

Qu’est-ce qu’un bon alignement ?

Pourcentage de lectures correctement alignées ?

Longueur ?

Génome ?

Quantité ?

Paires ?

Erreur ?

Qu’est-ce qu’un bon alignement ?

Pourcentage de lectures correctement alignées ?

Longueur ?

Génome ?

Quantité ?

Paires ?

Erreur ?

Alignement correct ?

La qualité, une donnée fiable ?

$$Q = -10 \log_{10} p$$

La qualité, une donnée fiable ?

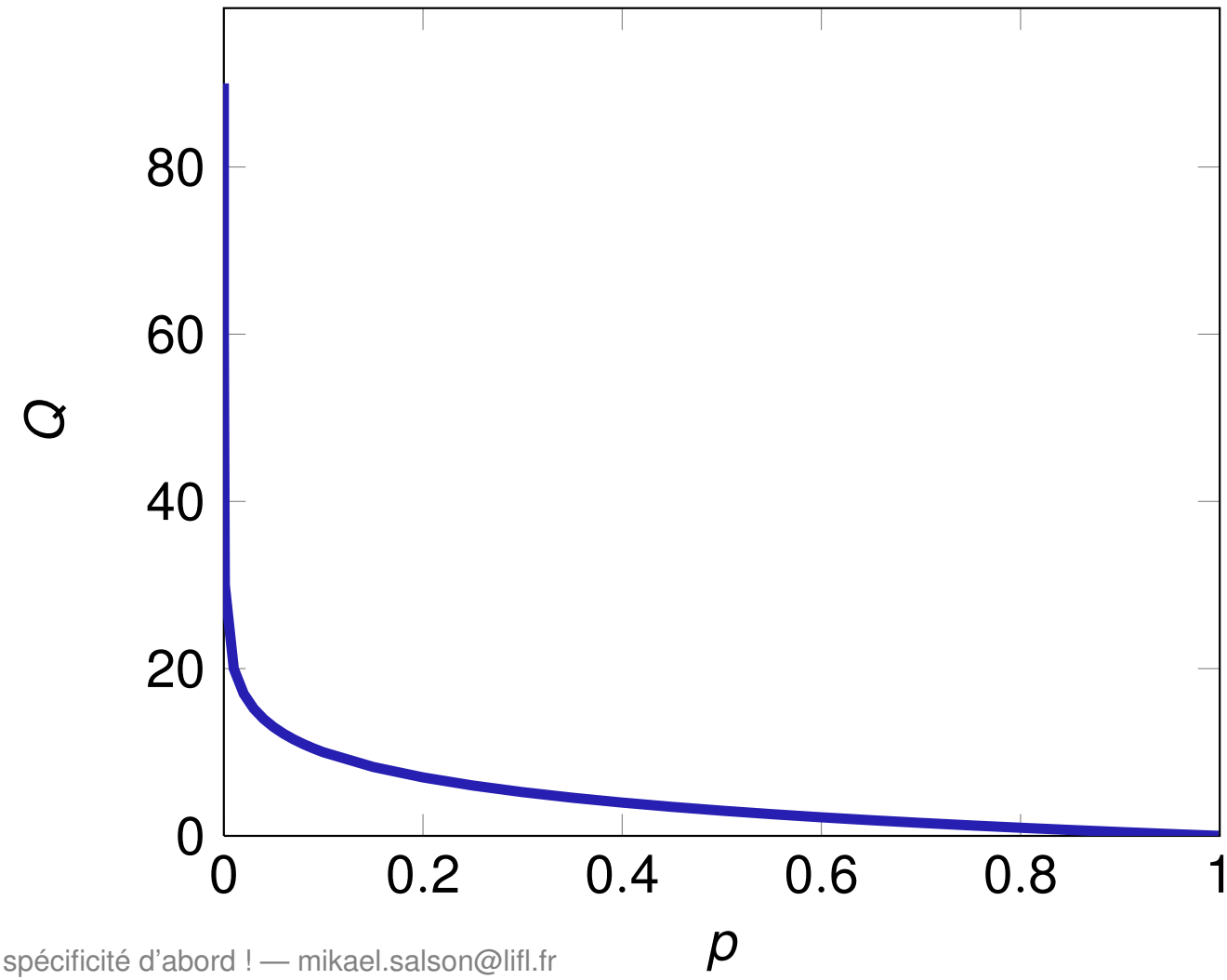
$$Q = -10 \log_{10} p$$

probabilité d'avoir une erreur

La qualité, une donnée fiable ?

$$Q = -10 \log_{10} p$$

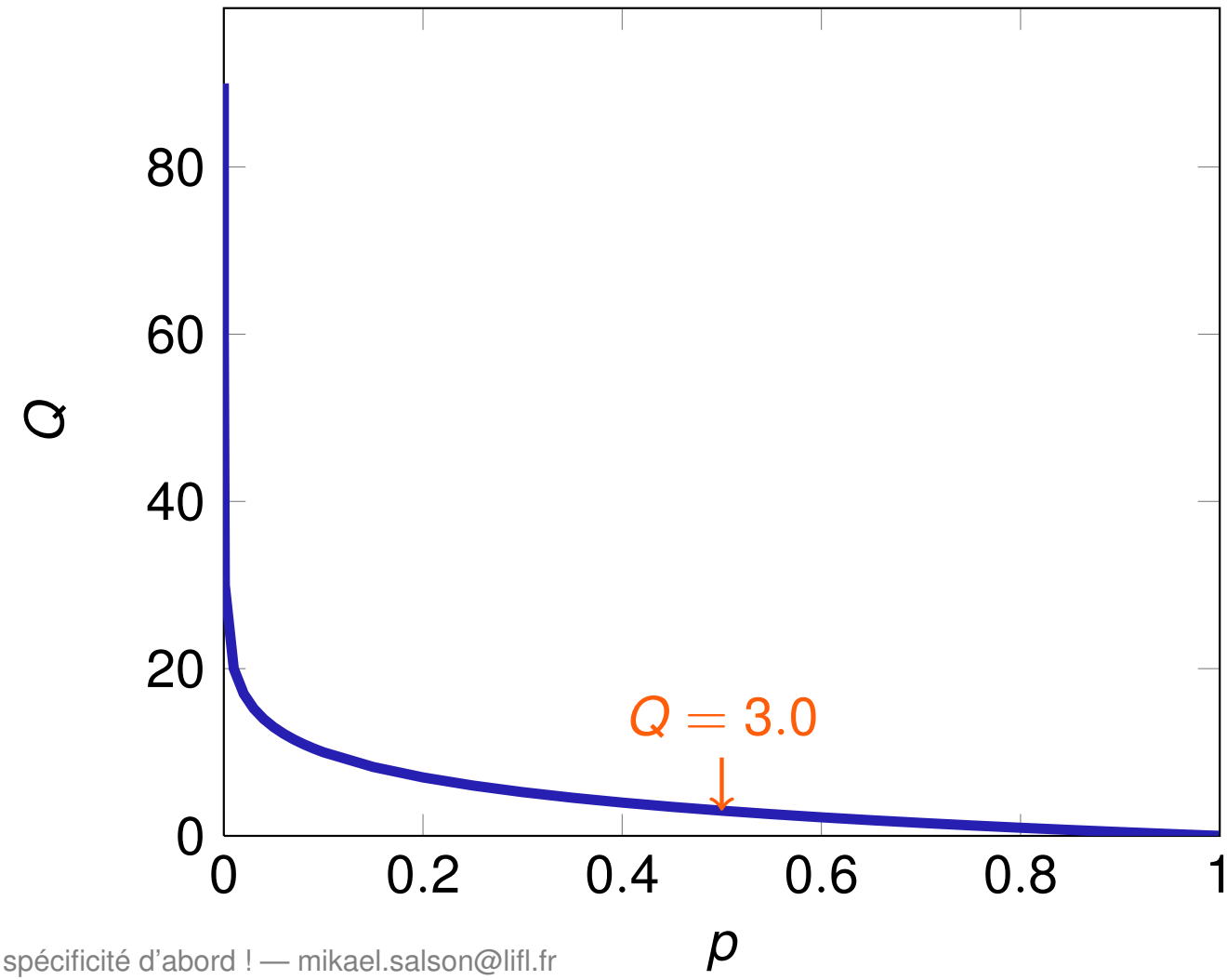
probabilité d'avoir une erreur



La qualité, une donnée fiable ?

$$Q = -10 \log_{10} p$$

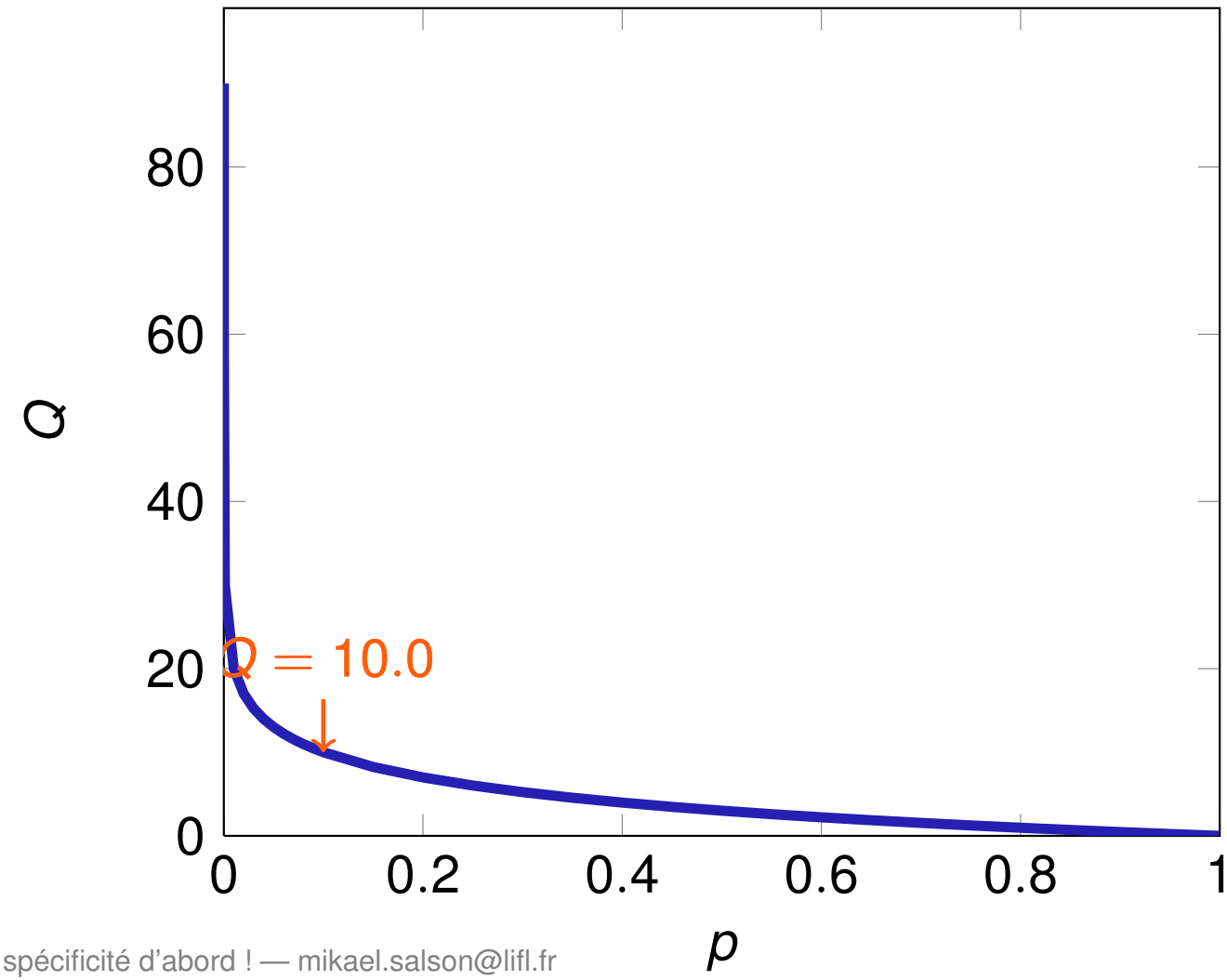
probabilité d'avoir une erreur



La qualité, une donnée fiable ?

$$Q = -10 \log_{10} p$$

probabilité d'avoir une erreur



La qualité, une donnée fiable !

La qualité, une donnée fiable !

« *Our results indicate that significant gains in Solexa read mapping performance can be achieved by* [...] *appropriately using the base-call quality scores* »

Smith *et al*, BMC Bioinformatics, 2008

La qualité, une donnée fiable !

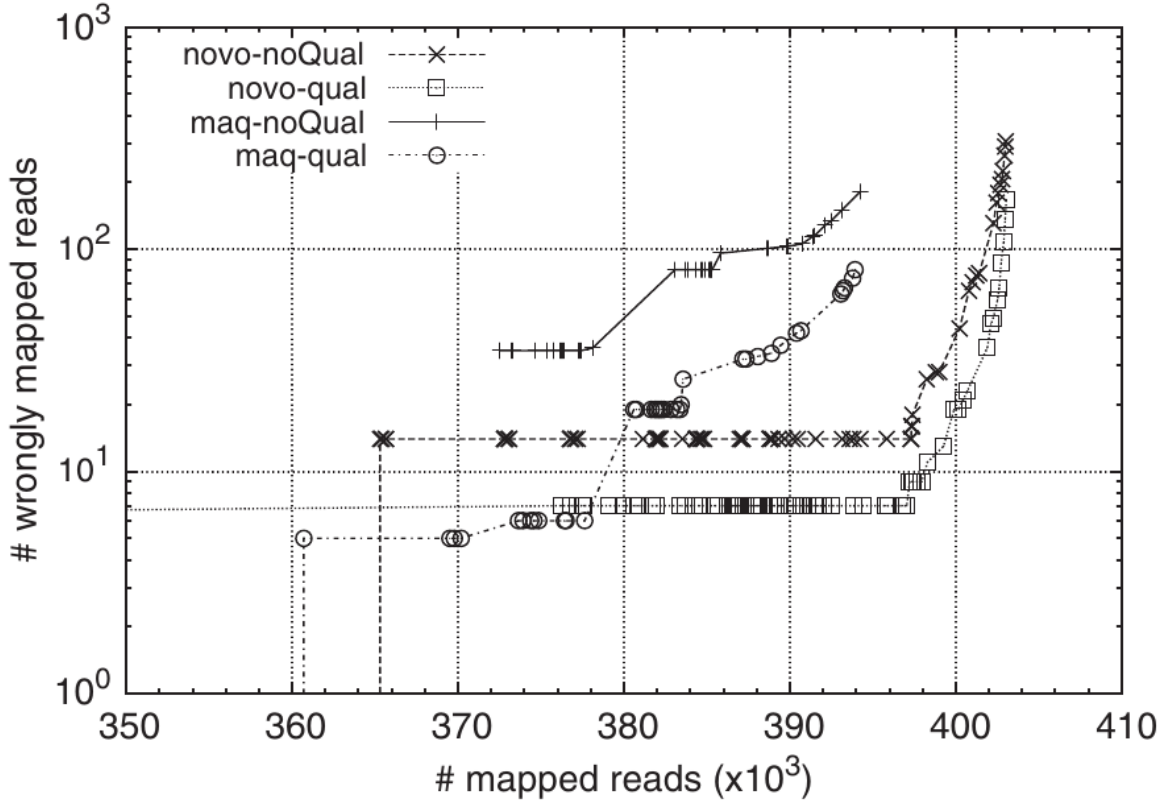
« *Our results indicate that significant gains in Solexa read mapping performance can be achieved by* [...] *appropriately using the base-call quality scores* »

Smith *et al*, BMC Bioinformatics, 2008

Alignement avec
X substitutions

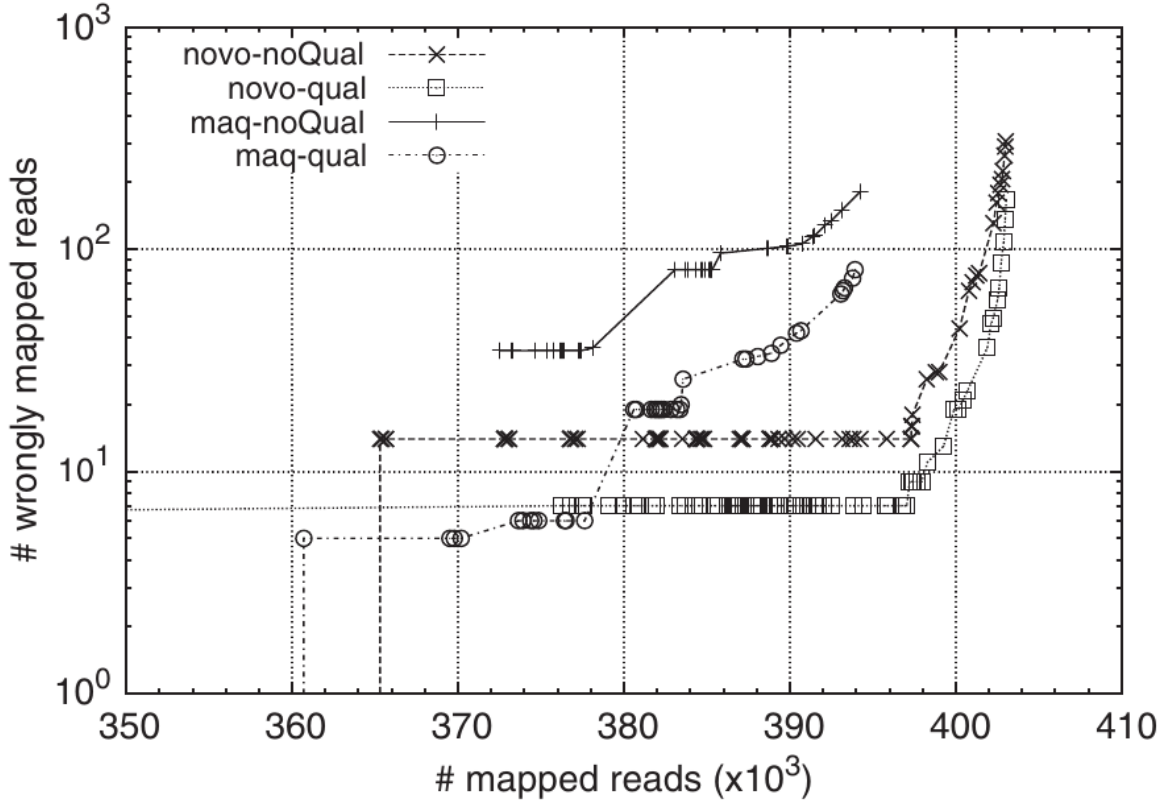
Alignement sans
substitution ou
permise si $Q \leq 8$

La qualité, une donnée fiable !



© *Li et Homer* Briefings in Bioinformatics, 2010

La qualité, une donnée fiable !



© *Li et Homer* Briefings in Bioinformatics, 2010

« *Figure 3 shows that using base quality score halves alignment errors when the quality score is accurate* »

Li et Homer, Briefings in Bioinformatics, 2010

La qualité, une donnée fiable ! ?

La qualité, une donnée fiable ! ?

chr. 21
humain

DRX000307

Illumina GA II
76 pb
20 M lectures

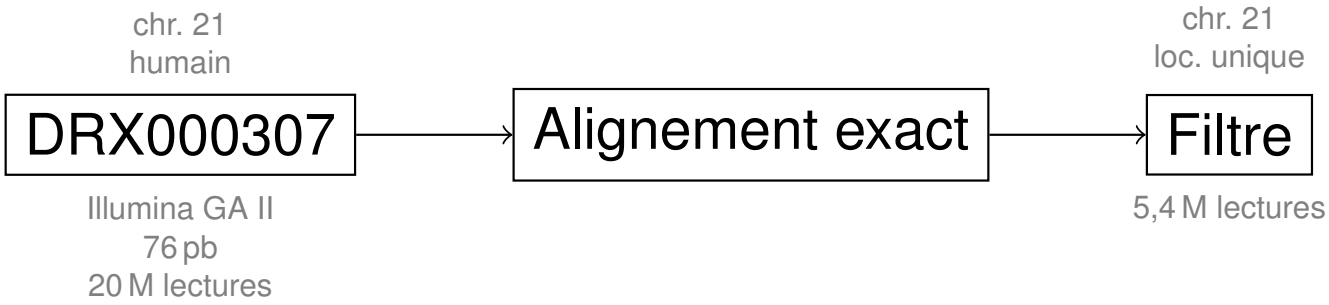
La qualité, une donnée fiable ! ?

chr. 21
humain

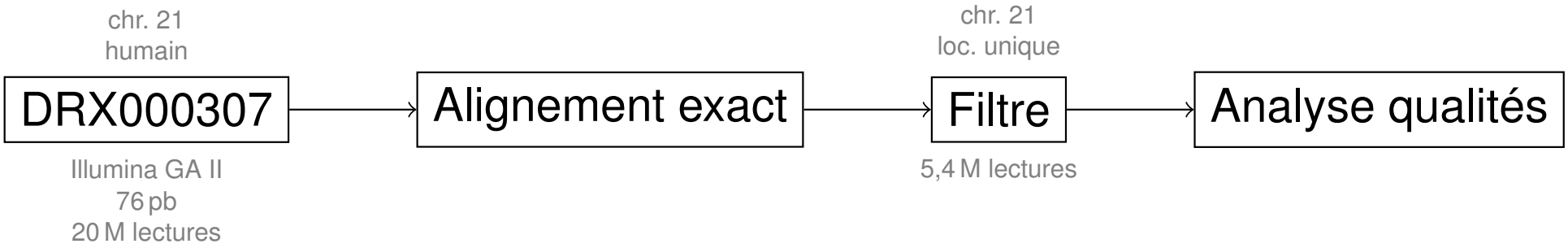
DRX000307 → Alignement exact

Illumina GA II
76 pb
20 M lectures

La qualité, une donnée fiable ! ?



La qualité, une donnée fiable ! ?



La qualité, une donnée fiable ! ?

chr. 21
humain

DRX000307

Illumina GA II
76 pb
20 M lectures

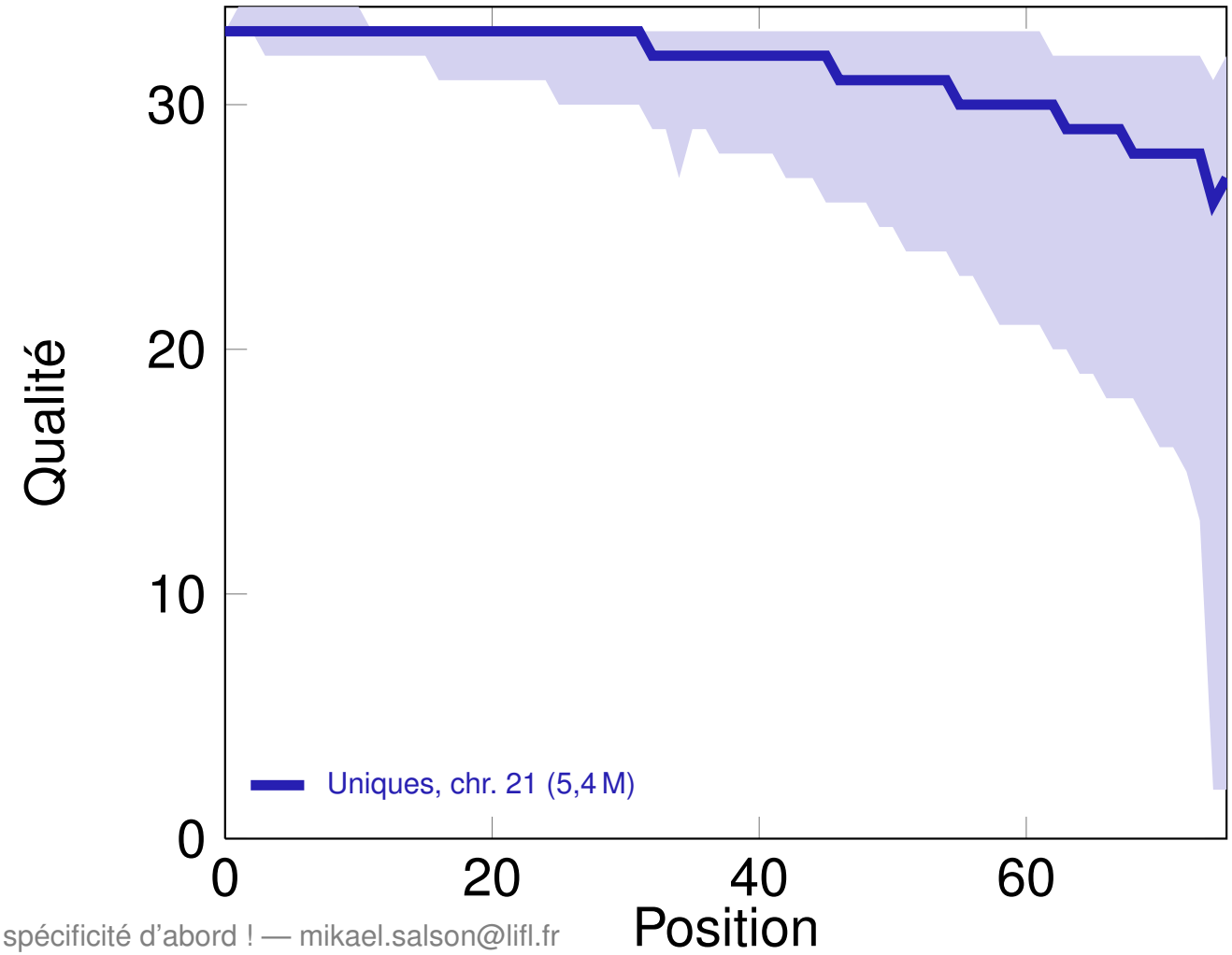
→ Alignement exact →

chr. 21
loc. unique

Filtre

5,4 M lectures

→ Analyse qualités



La qualité, une donnée fiable ! ?

chr. 21
humain

DRX000307

Illumina GA II
76 pb
20 M lectures

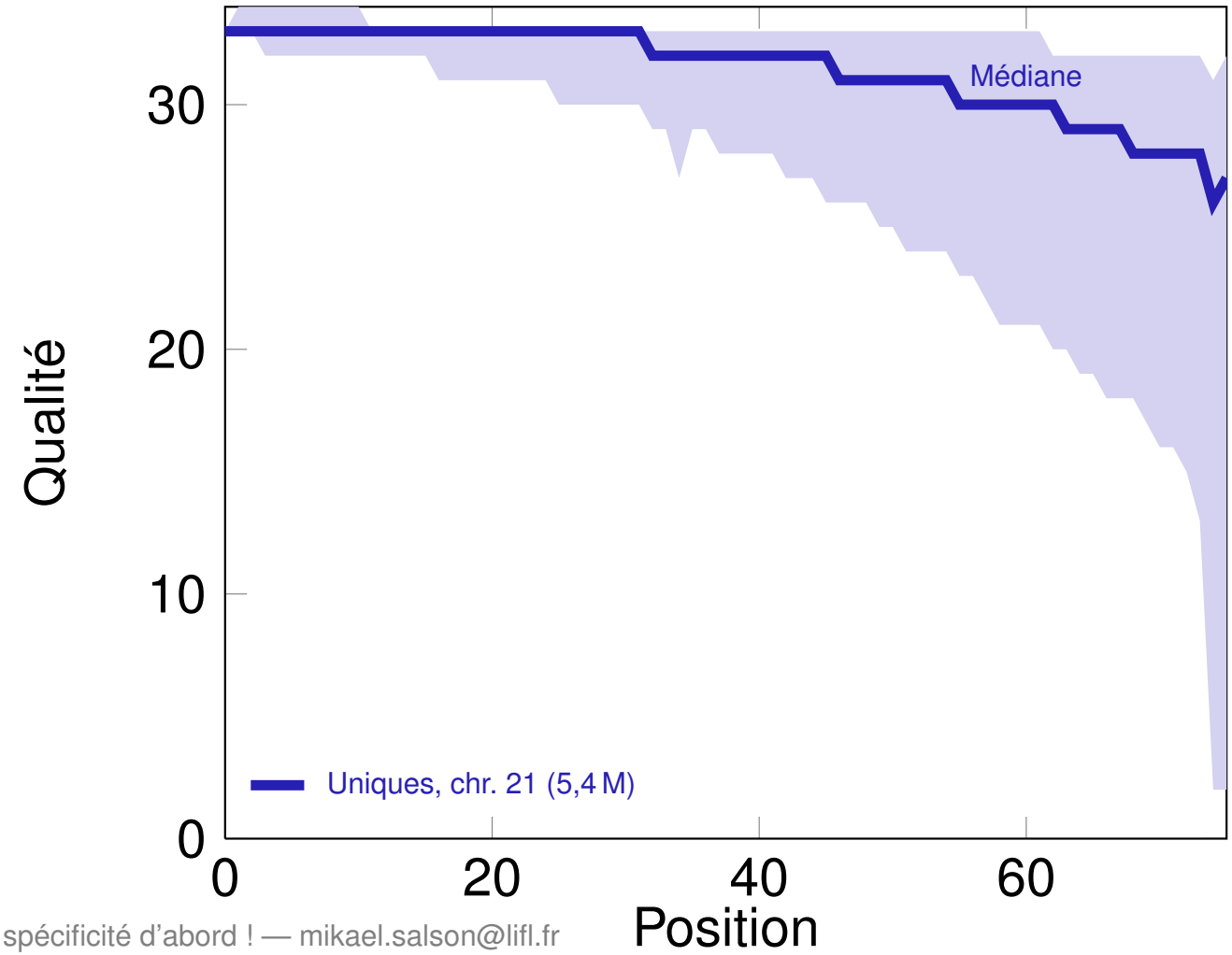
Alignement exact

chr. 21
loc. unique

Filtre

5,4 M lectures

Analyse qualités



La qualité, une donnée fiable ! ?

chr. 21
humain

DRX000307

Illumina GA II
76 pb
20 M lectures

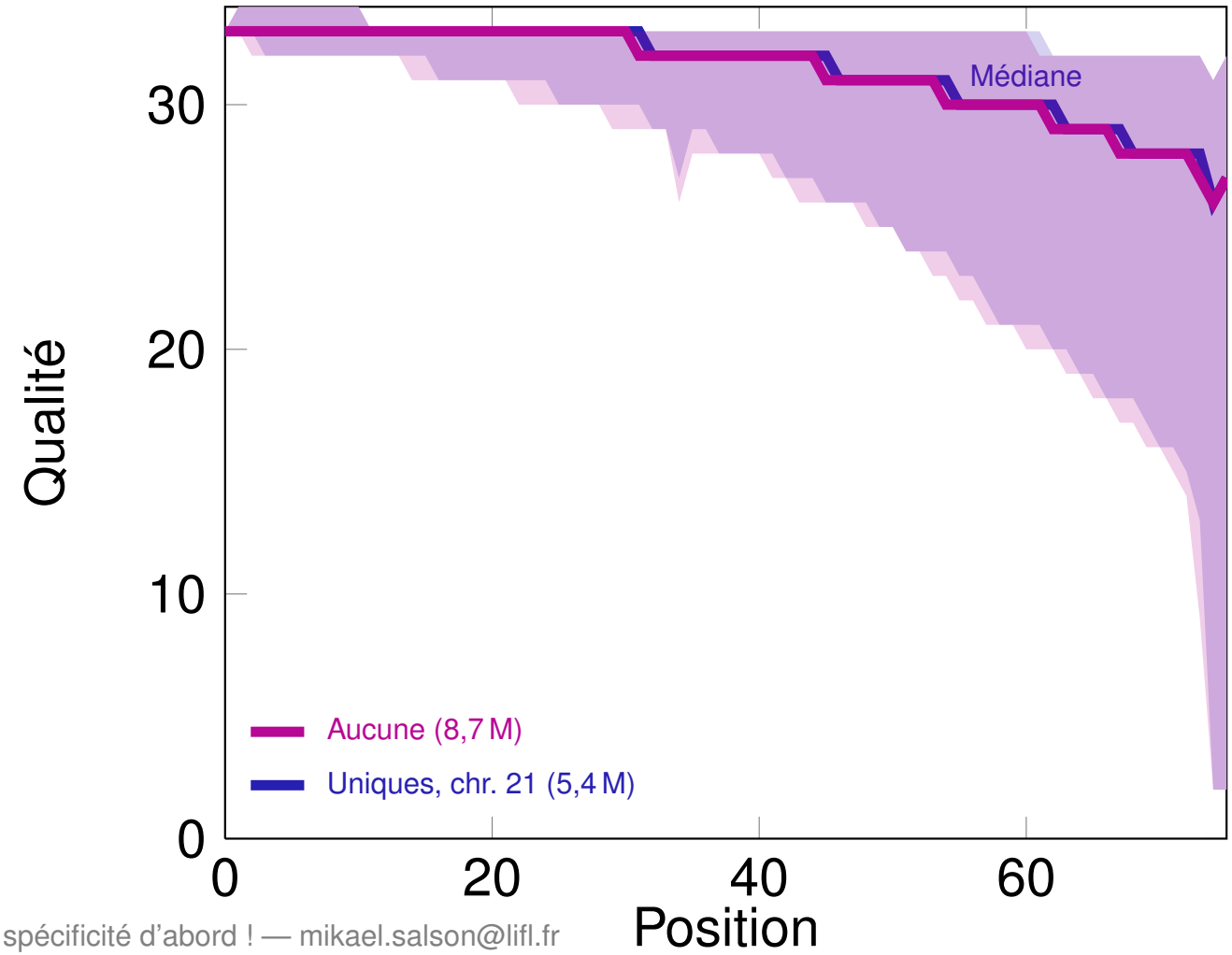
→ Alignement exact →

chr. 21
loc. unique

Filtre

5,4 M lectures

→ Analyse qualités



Qualité, une donnée fiable ! ? ?

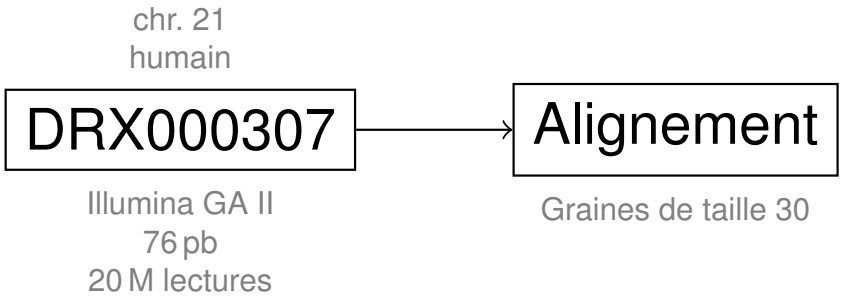
Qualité, une donnée fiable ! ? ?

chr. 21
humain

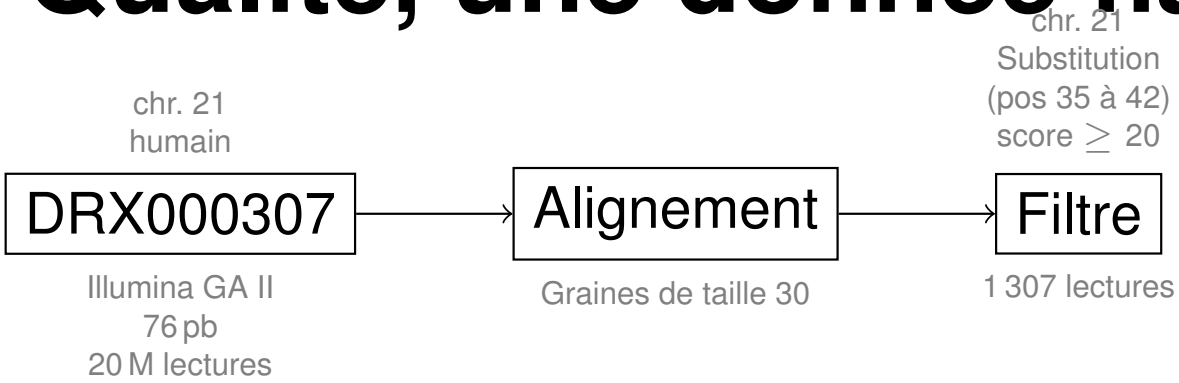
DRX000307

Illumina GA II
76 pb
20 M lectures

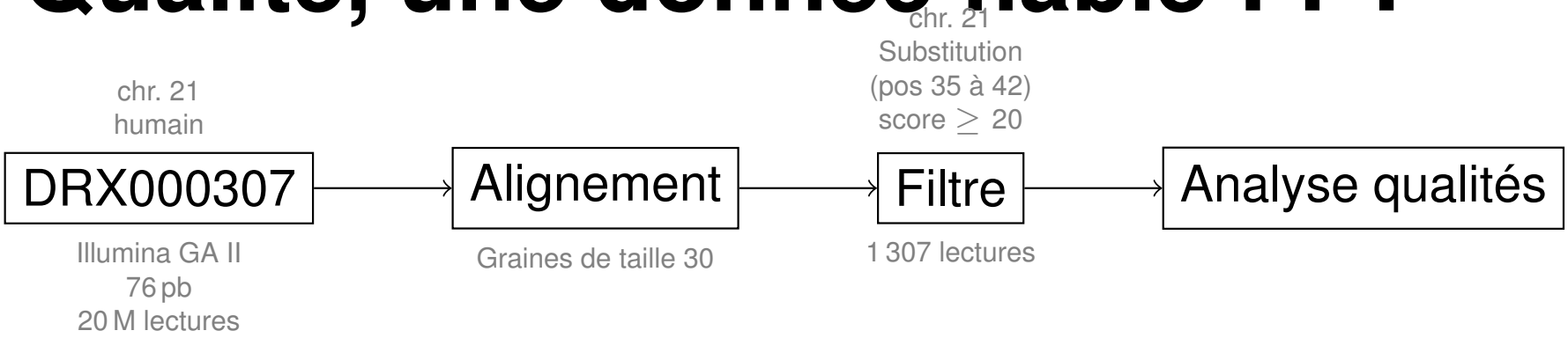
Qualité, une donnée fiable ! ? ?



Qualité, une donnée fiable ! ? ?



Qualité, une donnée fiable ! ? ?



Qualité, une donnée fiable ! ? ?

chr. 21
humain

DRX000307

Illumina GA II
76 pb
20 M lectures

→ Alignement

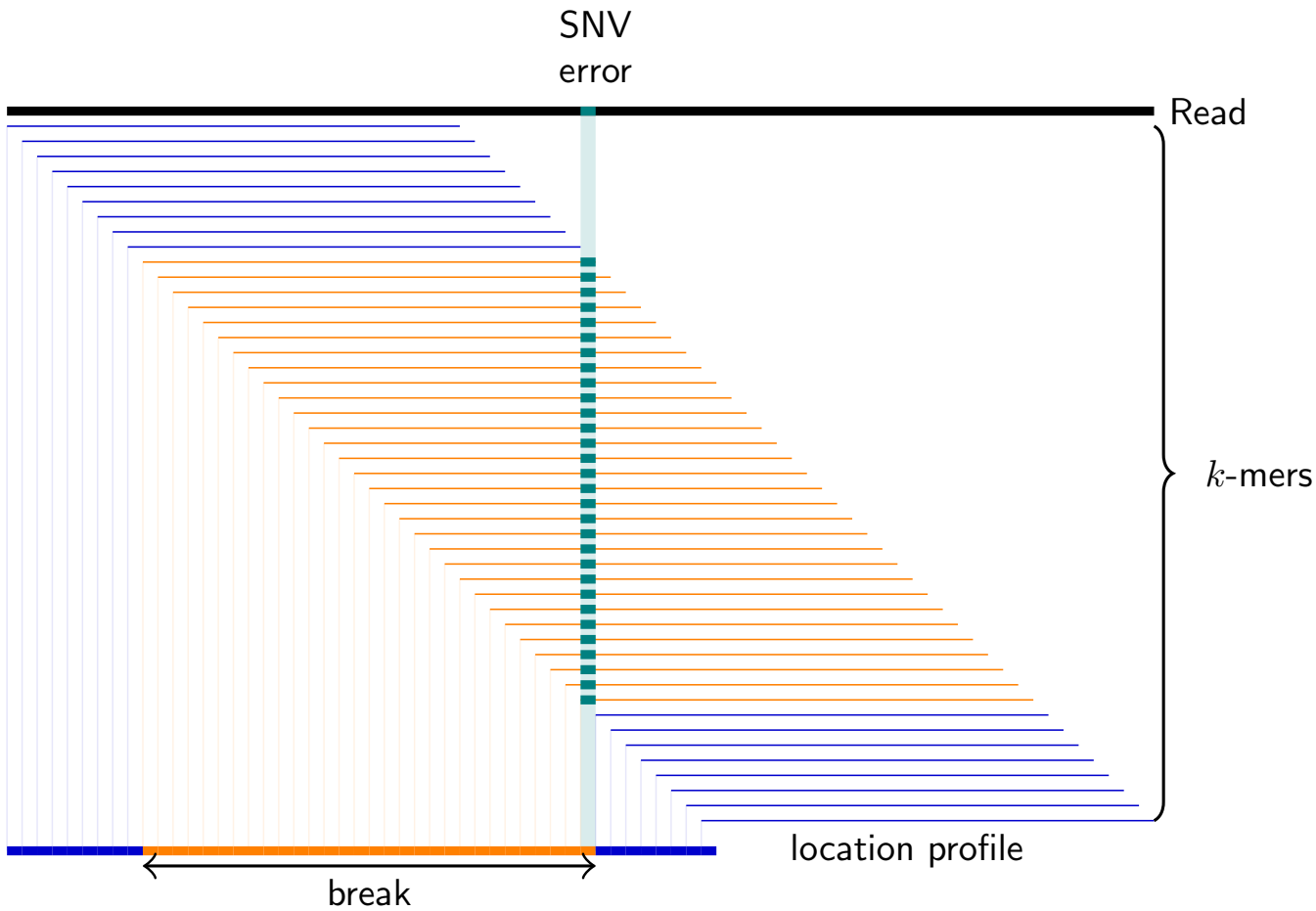
Graines de taille 30

→ Filtre

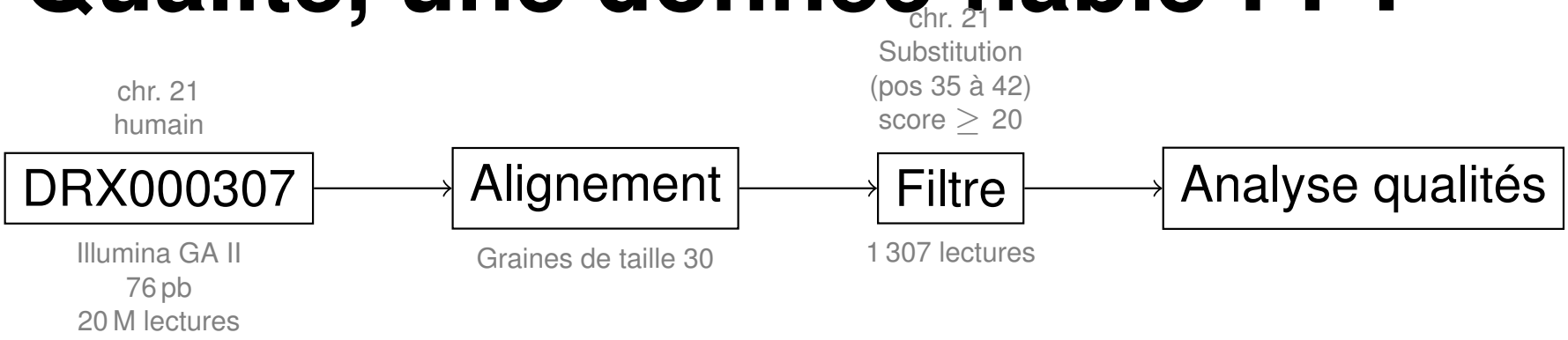
chr. 21
Substitution
(pos 35 à 42)
score $\geq$ 20

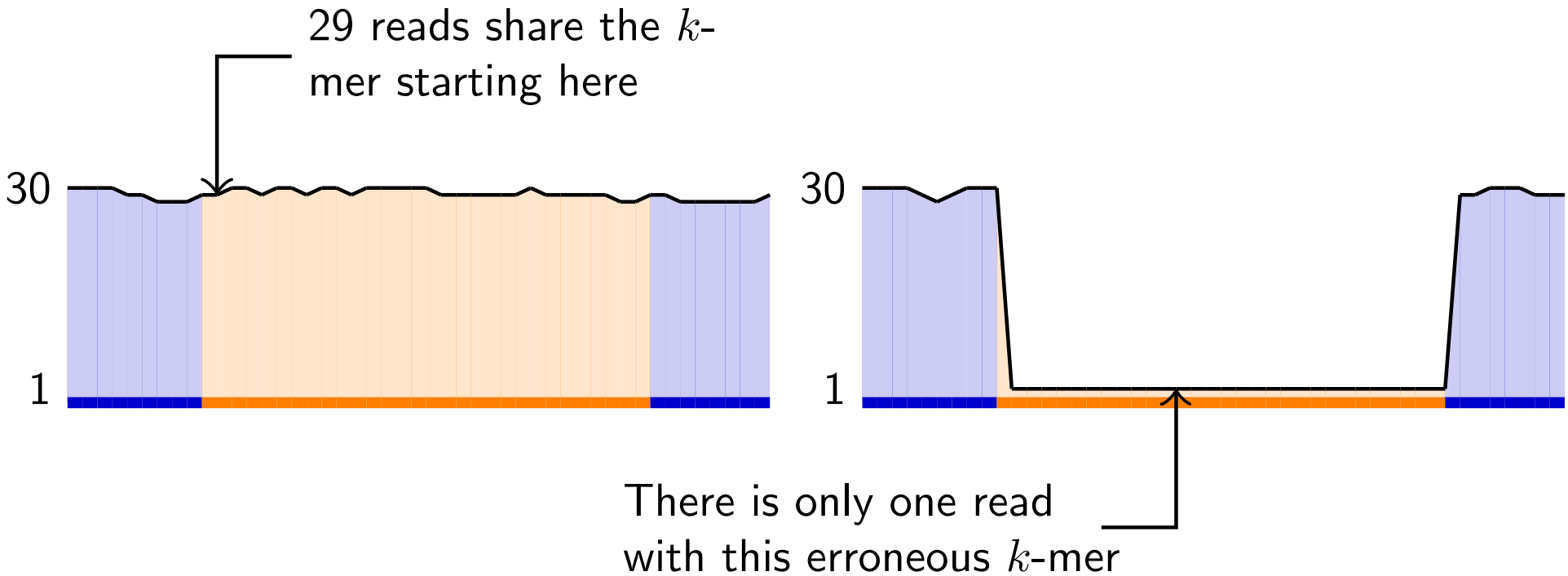
1 307 lectures

→ Analyse qualités



Qualité, une donnée fiable ! ? ?





Qualité, une donnée fiable ! ? ?

chr. 21
humain

DRX000307

Illumina GA II
76 pb
20 M lectures

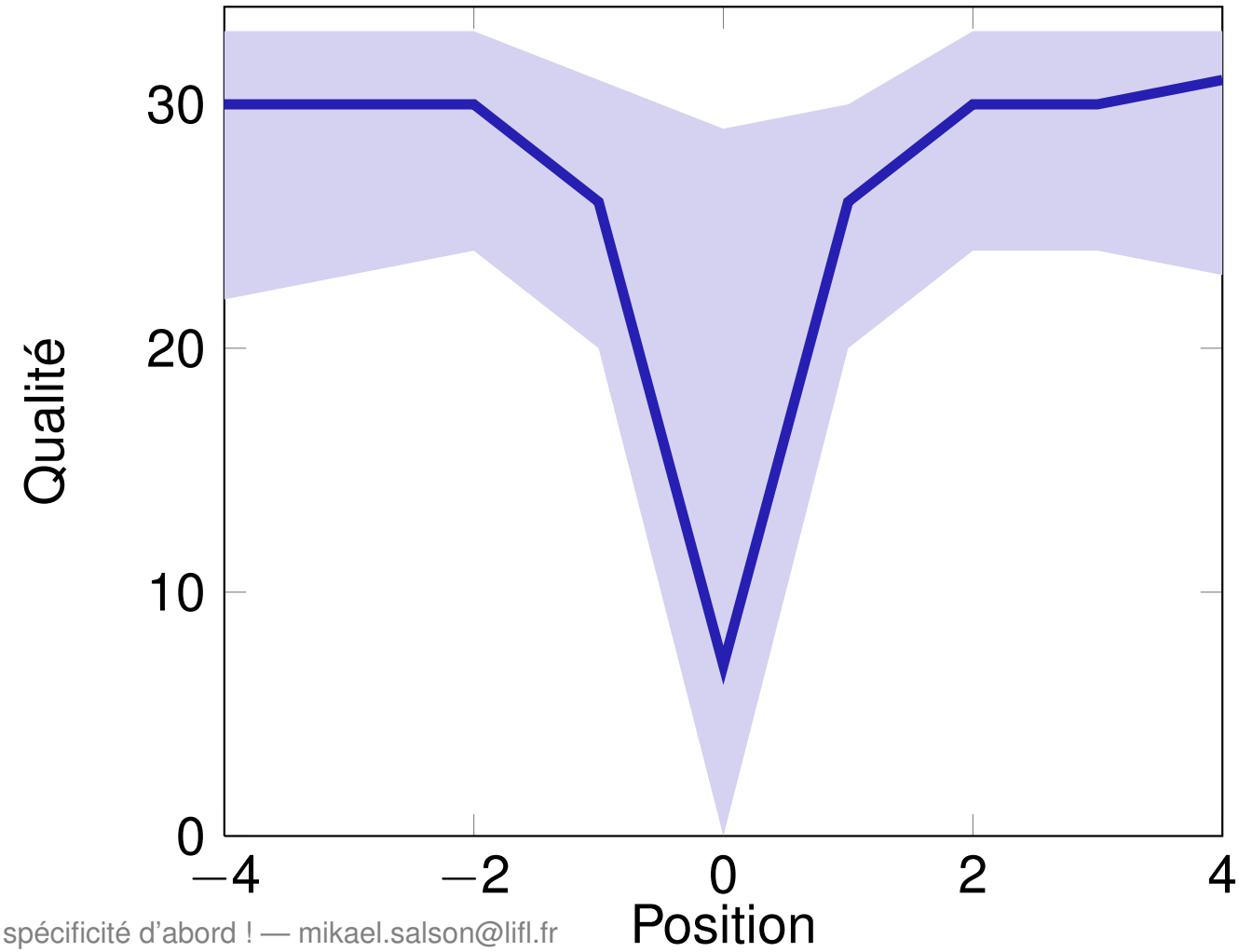
→ Alignement

Graines de taille 30

→ chr. 21
Substitution
(pos 35 à 42)
score $\geq$ 20
Filtre

1 307 lectures

→ Analyse qualités



Évaluation des performances

Évaluation des performances

SEAL

Ruffalo *et al*
Bioinformatics, 2011

Évaluation des performances

SEAL

Ruffalo *et al*
Bioinformatics, 2011

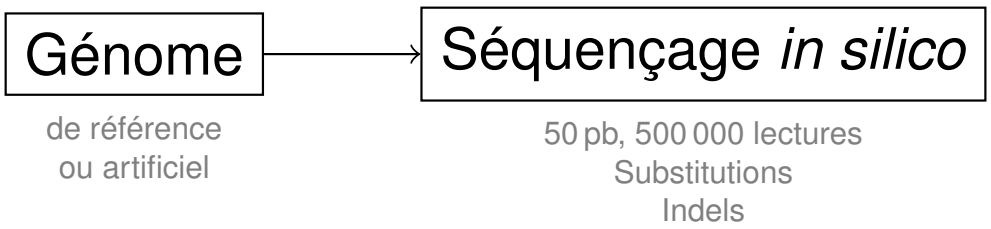
Génome

de référence
ou artificiel

Évaluation des performances

SEAL

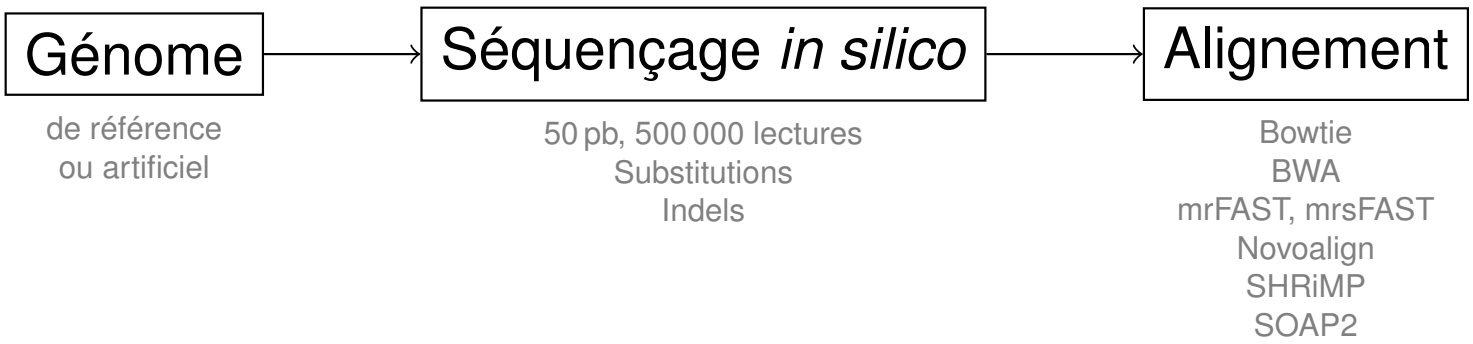
Ruffalo *et al*
Bioinformatics, 2011



Évaluation des performances

SEAL

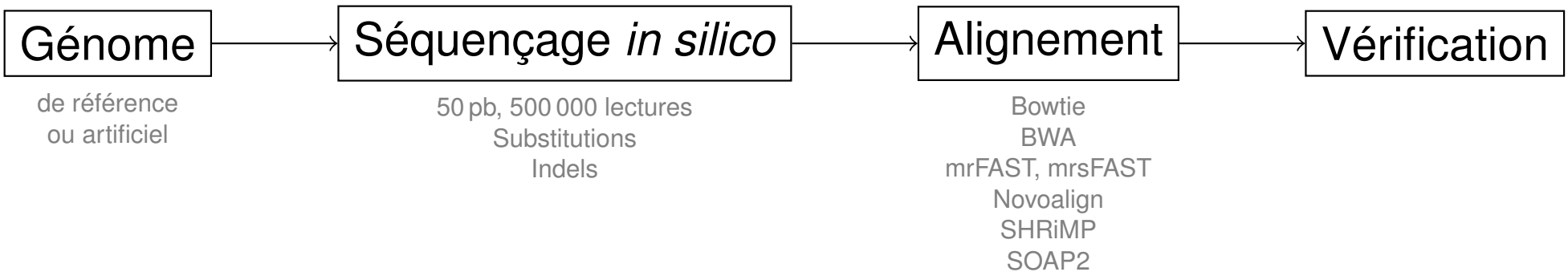
Ruffalo *et al*
Bioinformatics, 2011



Évaluation des performances

SEAL

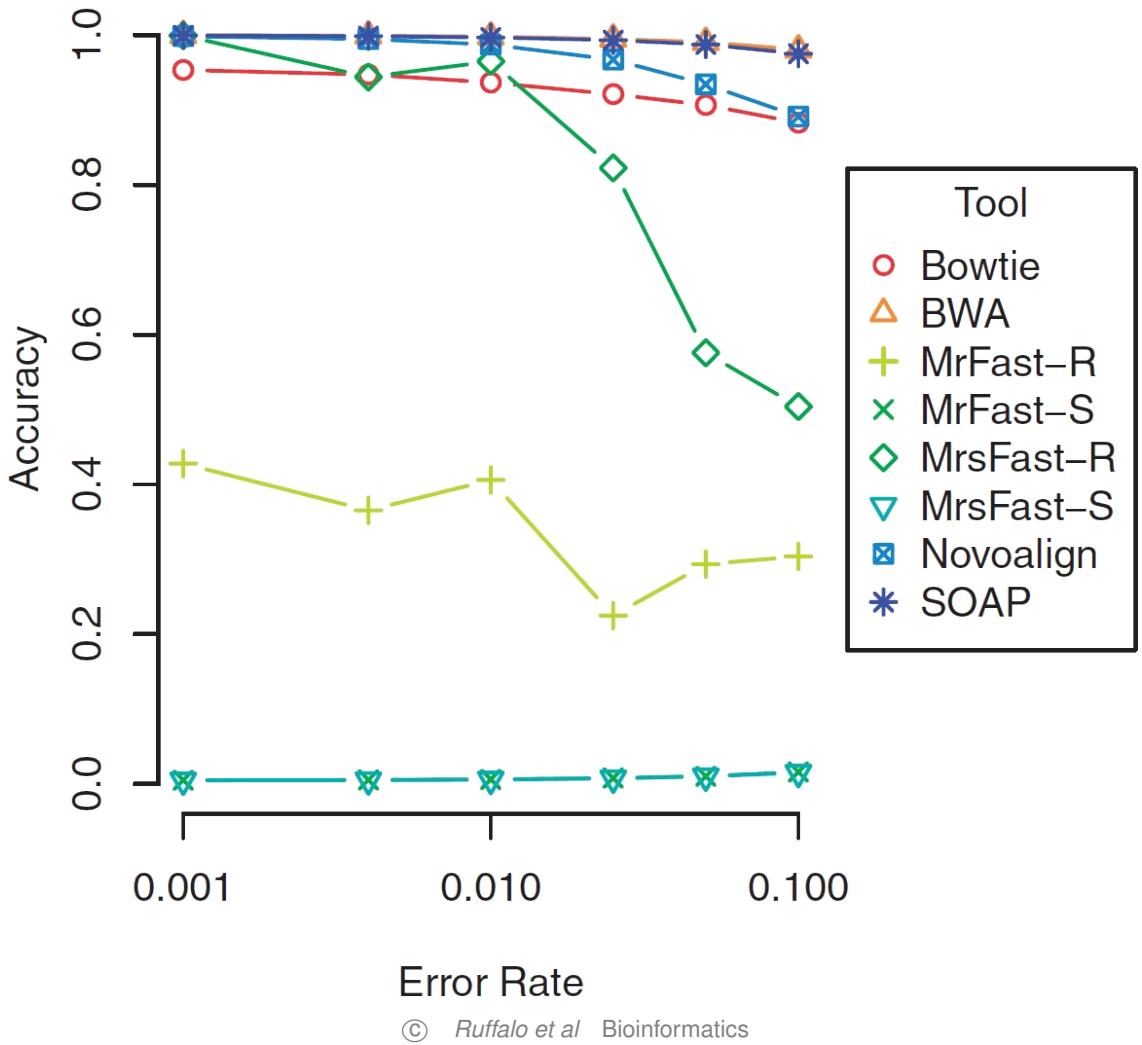
Ruffalo *et al*
Bioinformatics, 2011



Évaluation des performances

SEAL

Ruffalo *et al*
Bioinformatics, 2011

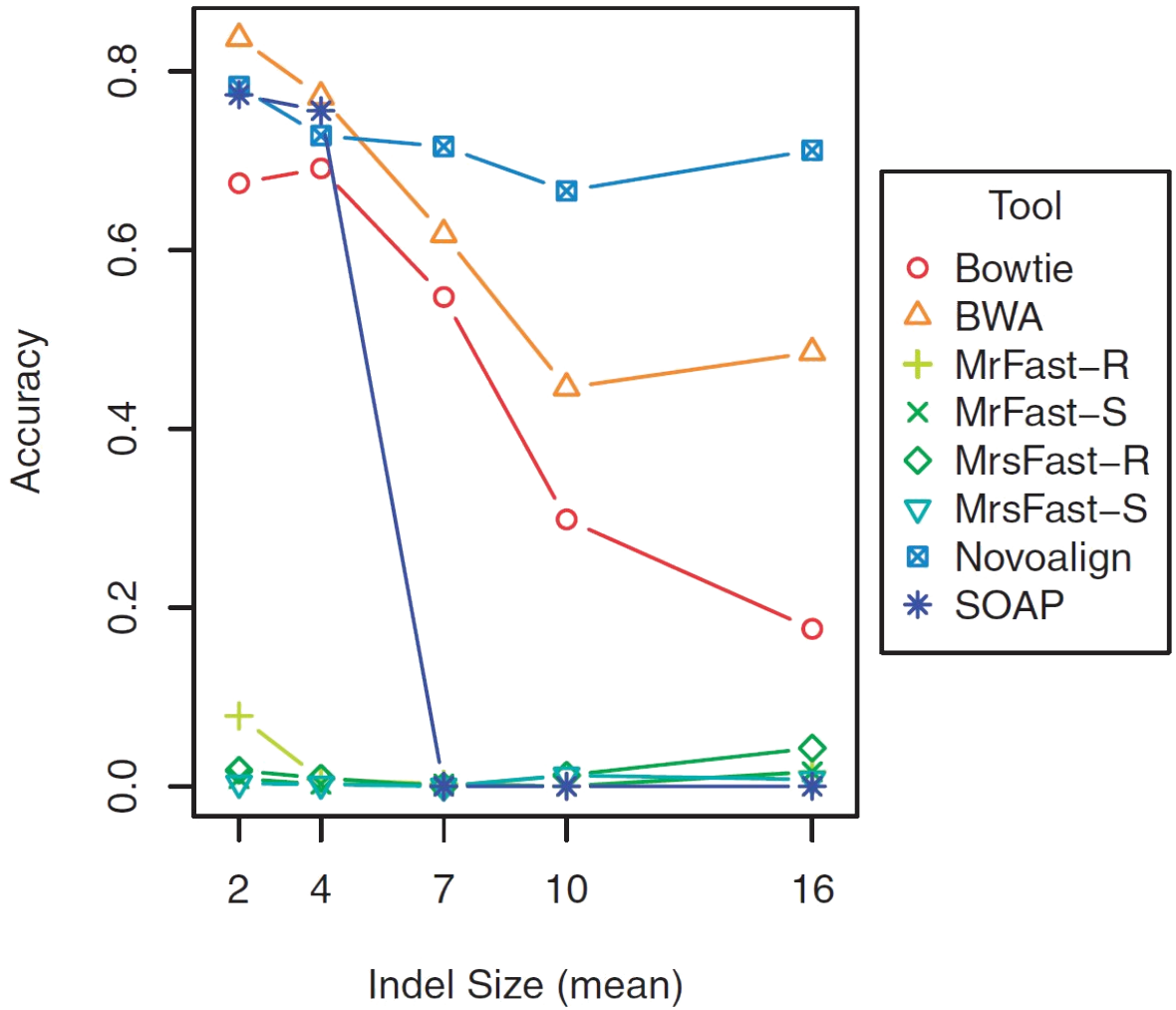


© *Ruffalo et al* Bioinformatics

Évaluation des performances

SEAL

Ruffalo *et al*
Bioinformatics, 2011



© *Ruffalo et al* Bioinformatics

Évaluation des performances

Rabema

Holtgrewe *et al*
BMC Bioinformatics, 2011

Évaluation des performances

Rabema

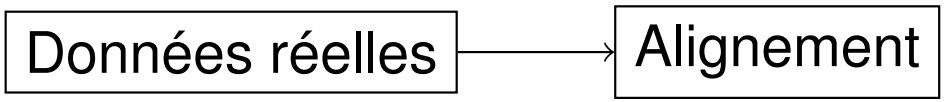
Holtgrewe *et al*
BMC Bioinformatics, 2011

Données réelles

Évaluation des performances

Rabema

Holtgrewe *et al*
BMC Bioinformatics, 2011



10 000 lectures sélectionnées
Sensibilité max (RazerS)

Évaluation des performances

Rabema

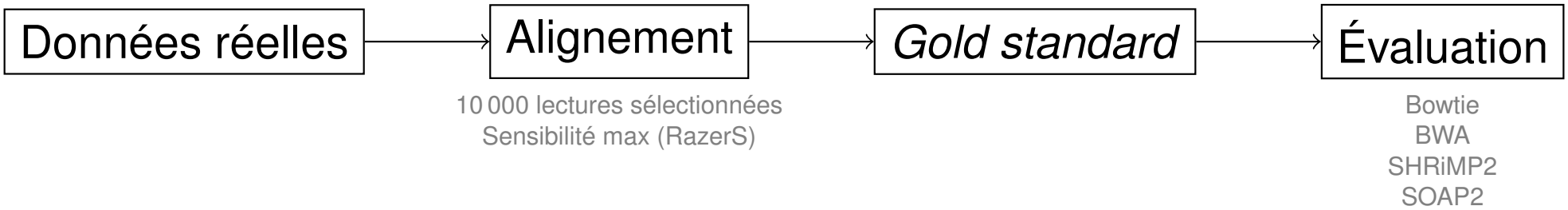
Holtgrewe *et al*
BMC Bioinformatics, 2011

10 000 lectures sélectionnées
Sensibilité max (RazerS)

Évaluation des performances

Rabema

Holtgrewe *et al*
BMC Bioinformatics, 2011

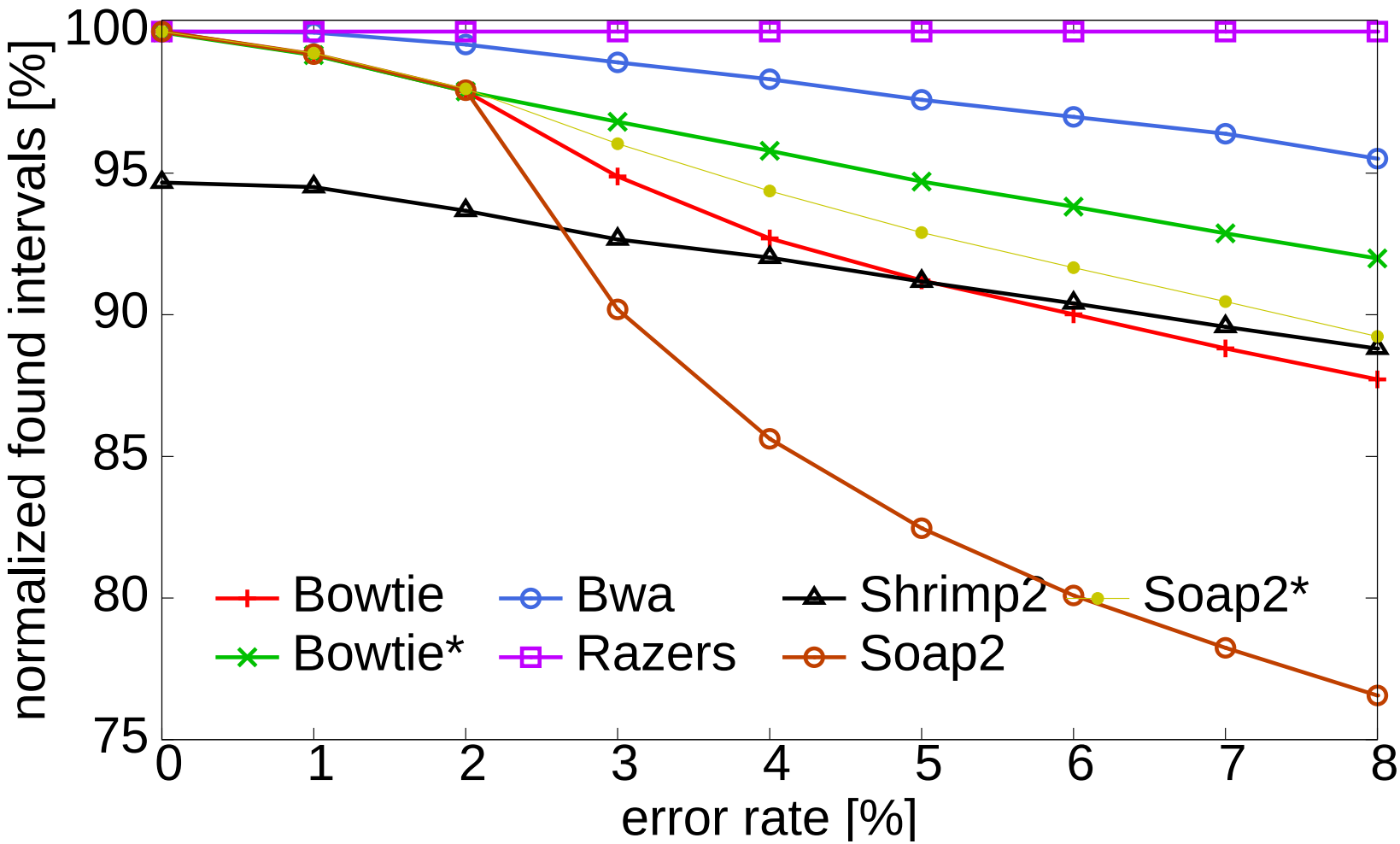


Évaluation des performances

Rabema

Holtgrewe *et al*
BMC Bioinformatics, 2011

D. melanogaster, lectures 100 pb Illumina

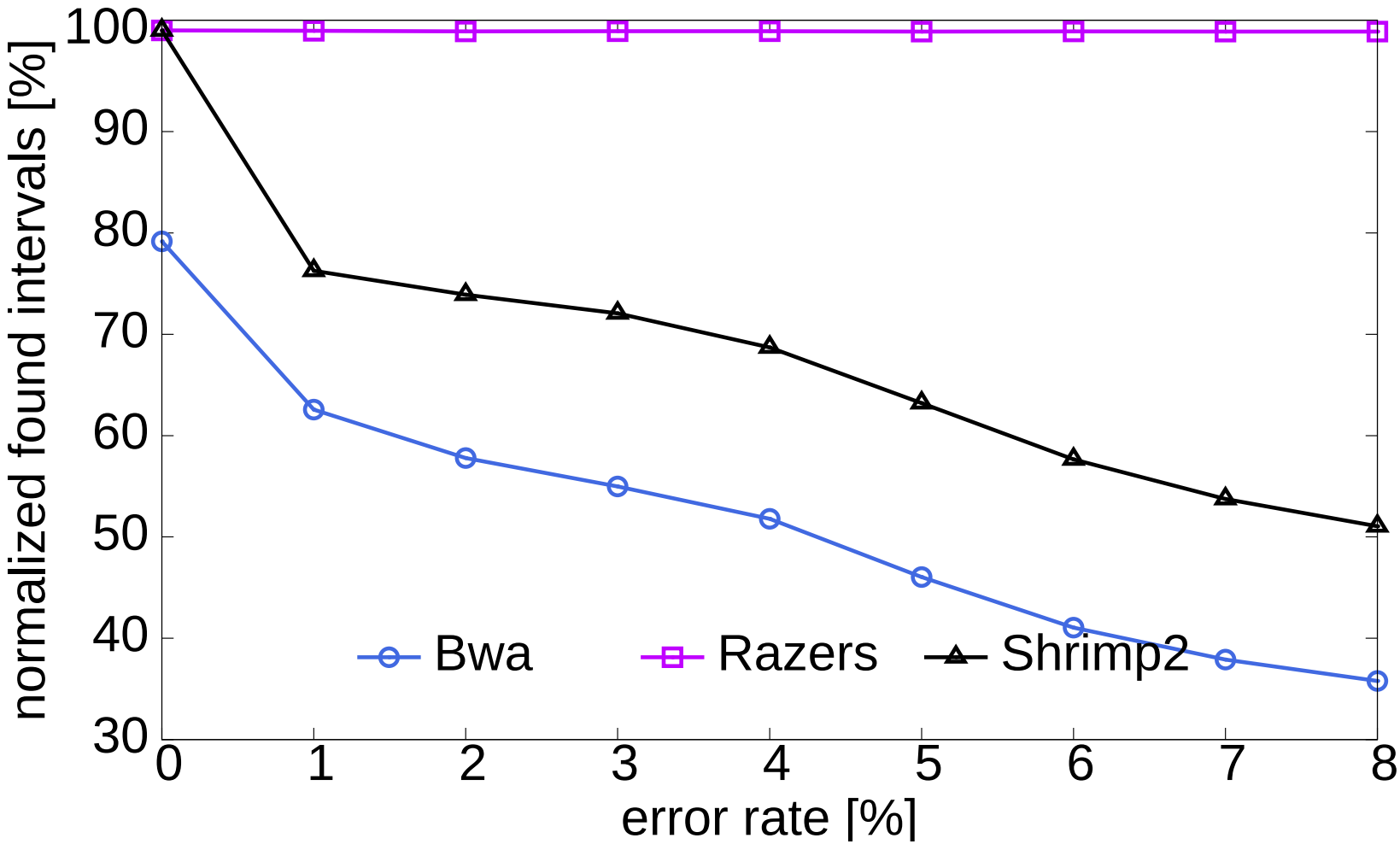


Évaluation des performances

Rabema

Holtgrewe *et al*
BMC Bioinformatics, 2011

D. melanogaster, lectures 454 (moyenne : 273 pb)



Évaluation des performances

Schbath *et al*

Journal of Computational Biology, 2012

Évaluation des performances

Schbath *et al*

Journal of Computational Biology, 2012

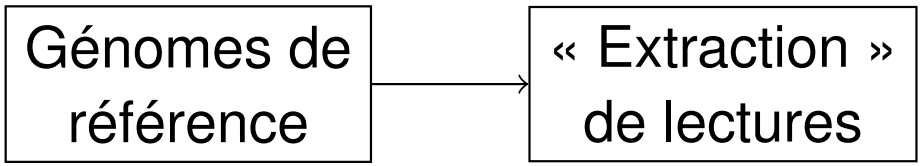
Génomes de référence

H. sapiens
904 génomes bactériens

Évaluation des performances

Schbath *et al*

Journal of Computational Biology, 2012



H. sapiens
904 génomes bactériens

10 M lectures de 40 pb
Ajout de trois substitutions

Évaluation des performances

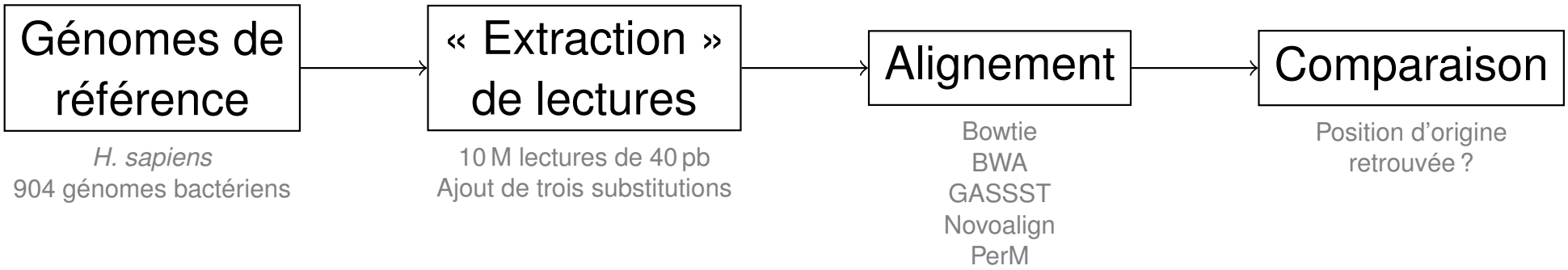
Schbath *et al*

Journal of Computational Biology, 2012

Évaluation des performances

Schbath *et al*

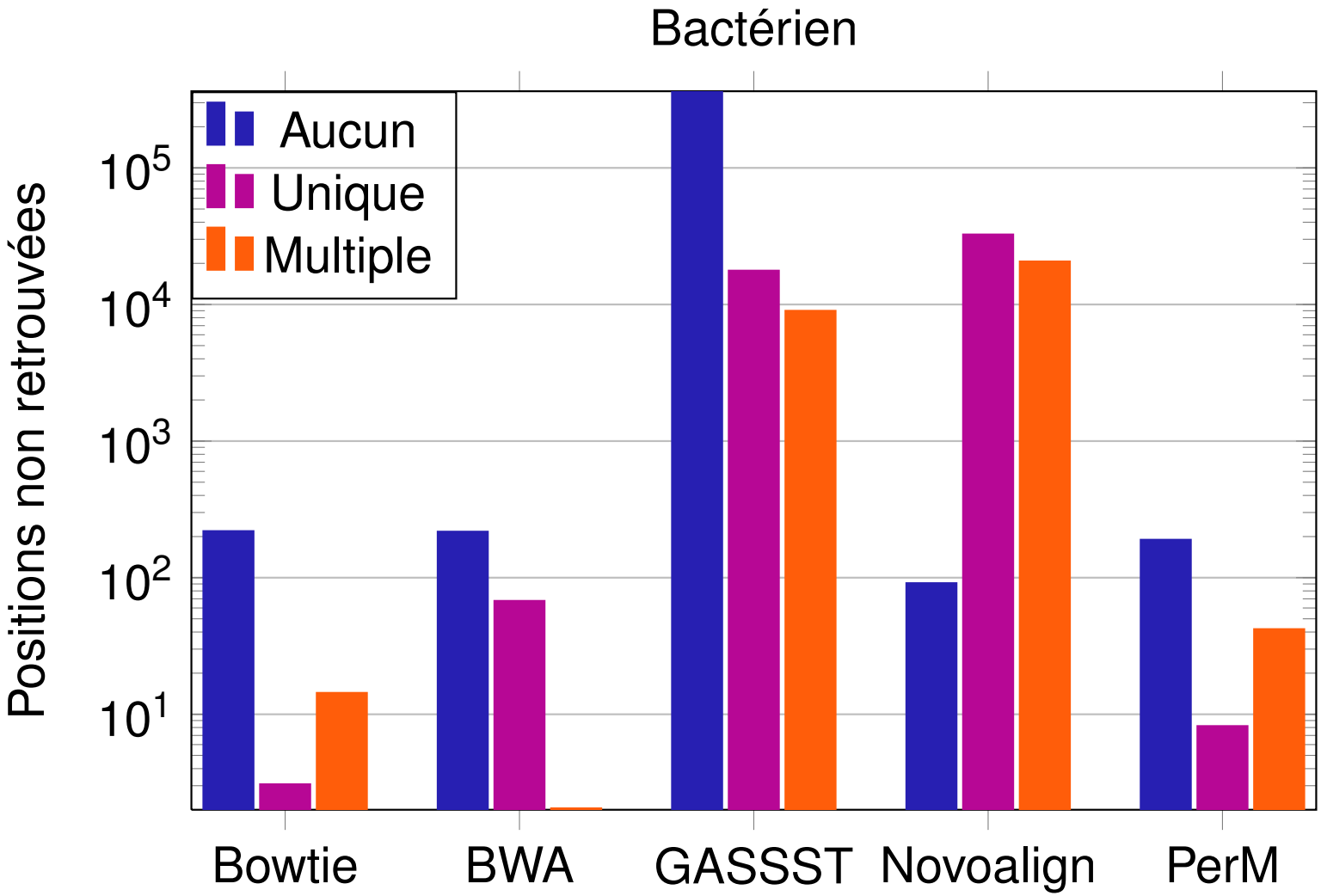
Journal of Computational Biology, 2012



Évaluation des performances

Schbath *et al*

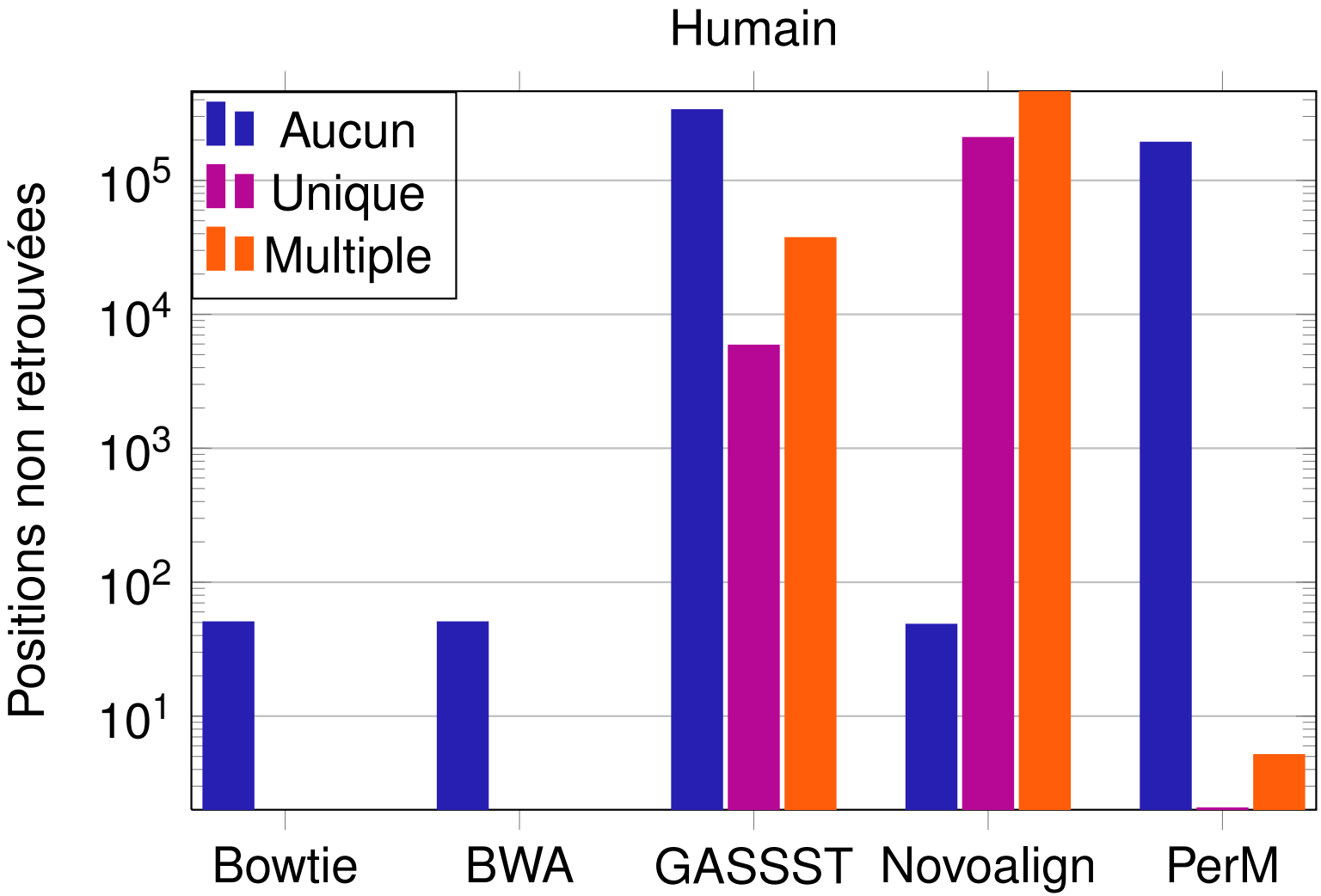
Journal of Computational Biology, 2012



Évaluation des performances

Schbath *et al*

Journal of Computational Biology, 2012



Évaluation des performances

Yu *et al*

BioData Mining, 2012

Évaluation des performances

Yu *et al*

BioData Mining, 2012

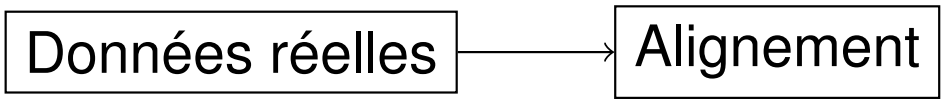
Données réelles

3 000 exons humains
7,4 M et 5,4 M lectures de 68pb

Évaluation des performances

Yu *et al*

BioData Mining, 2012



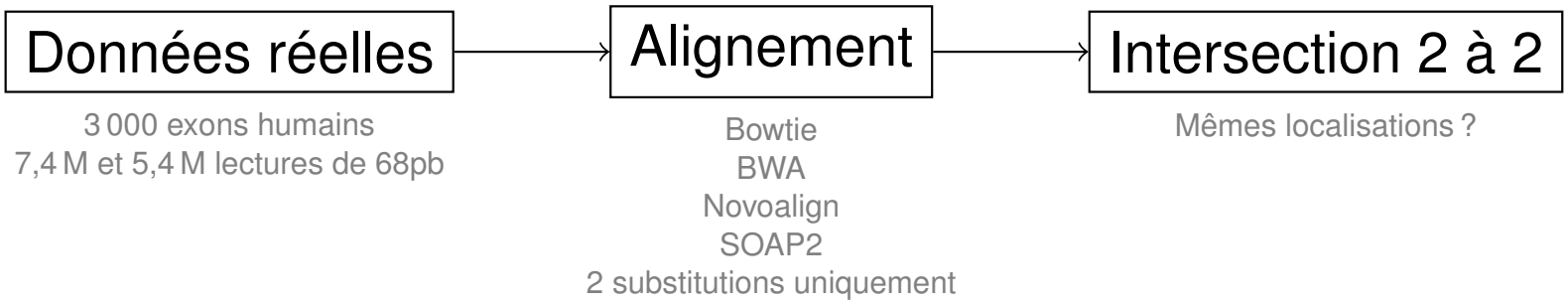
3 000 exons humains
7,4 M et 5,4 M lectures de 68pb

Bowtie
BWA
Novoalign
SOAP2
2 substitutions uniquement

Évaluation des performances

Yu *et al*

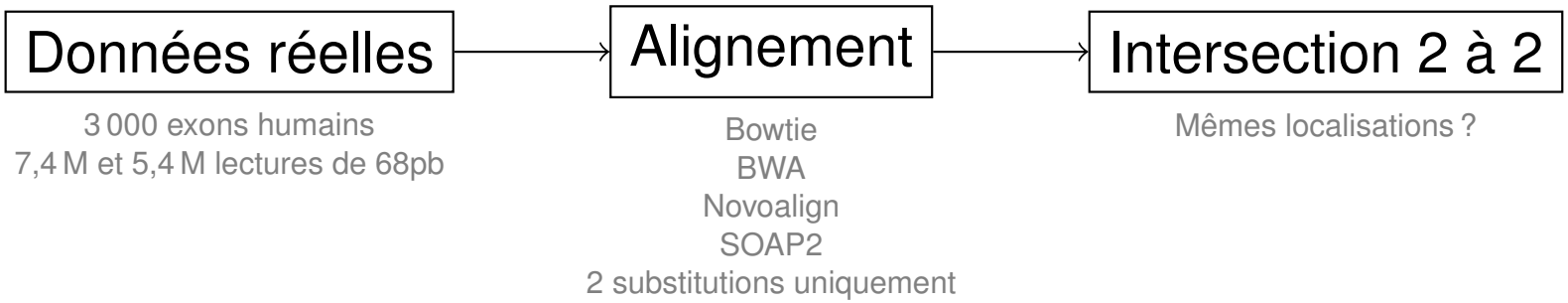
BioData Mining, 2012



Évaluation des performances

Yu *et al*

BioData Mining, 2012



$\sim$ 95 % d'identité entre Bowtie, BWA et SOAP2

$\sim$ 85 % pour Novoalign

Évalution. . . et le RNA-Seq ?

Jonctions exon-exon ?

Évalution. . . et le RNA-Seq ?

Jonctions exon-exon ?

Transcrits de fusion ?

CRAC, un outil pour le RNA-Seq

CRAC, un outil pour le RNA-Seq

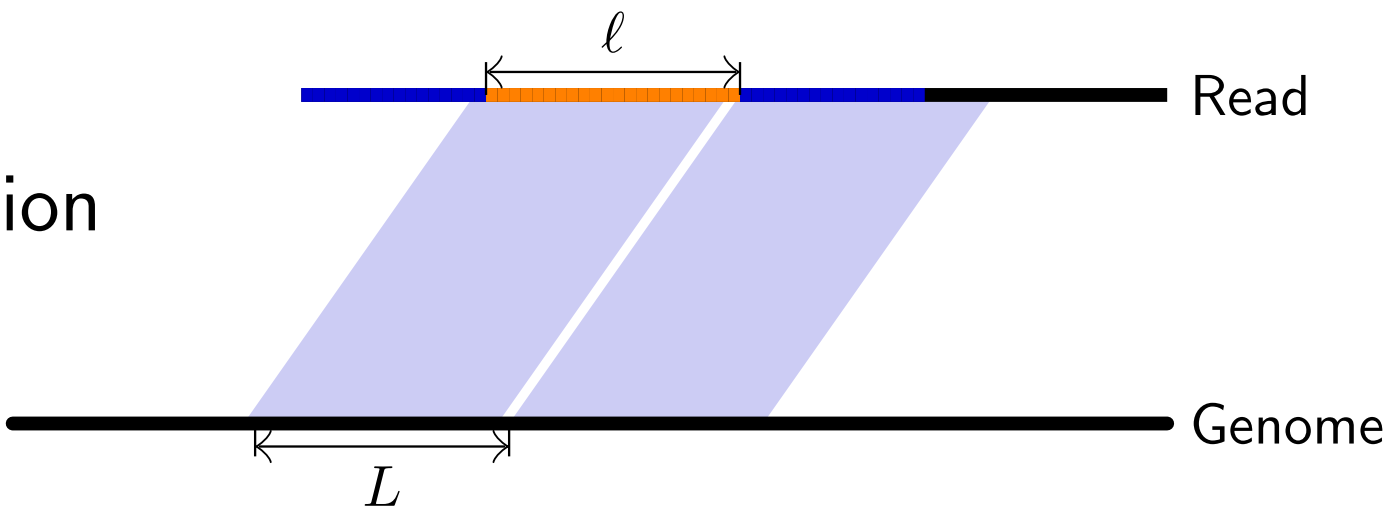
Avec Nicolas Philippe, Thérèse Commes, Éric Rivals

crac.gforge.inria.fr

CRAC, un outil pour le RNA-Seq

Substitution

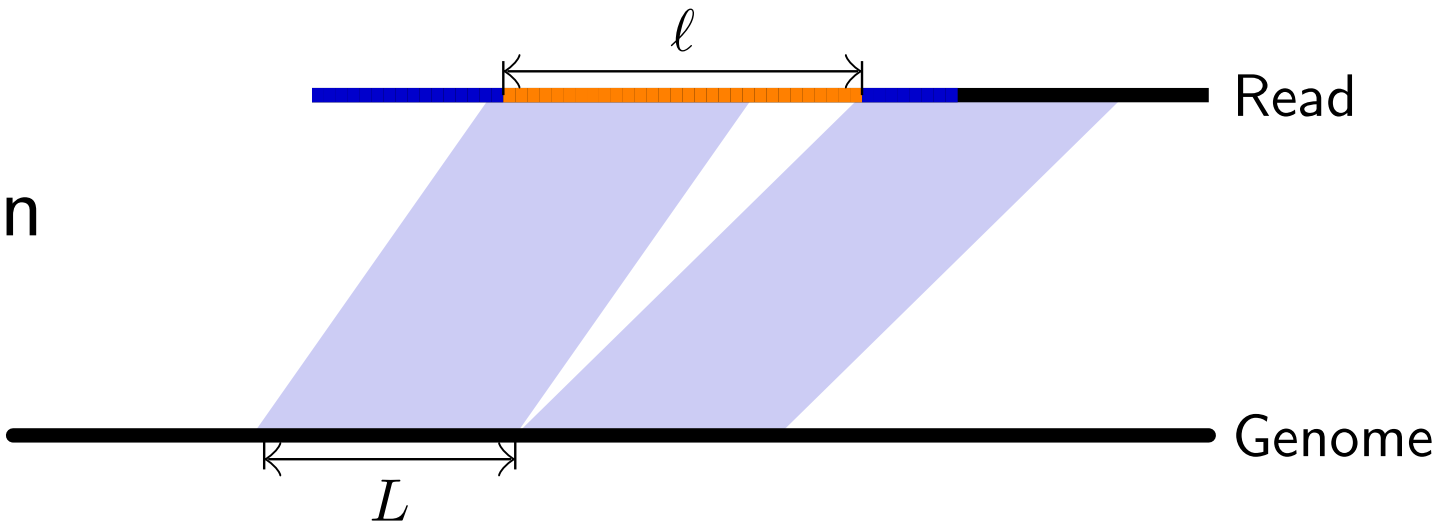
$\ell = L$



CRAC, un outil pour le RNA-Seq

Insertion

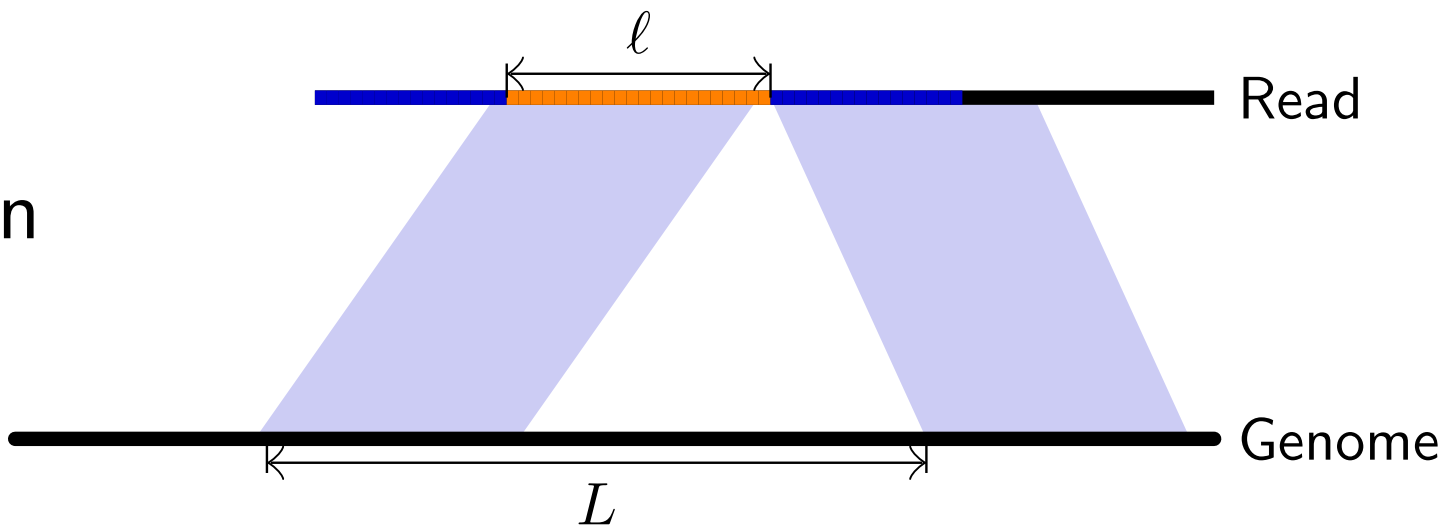
$\ell > L$



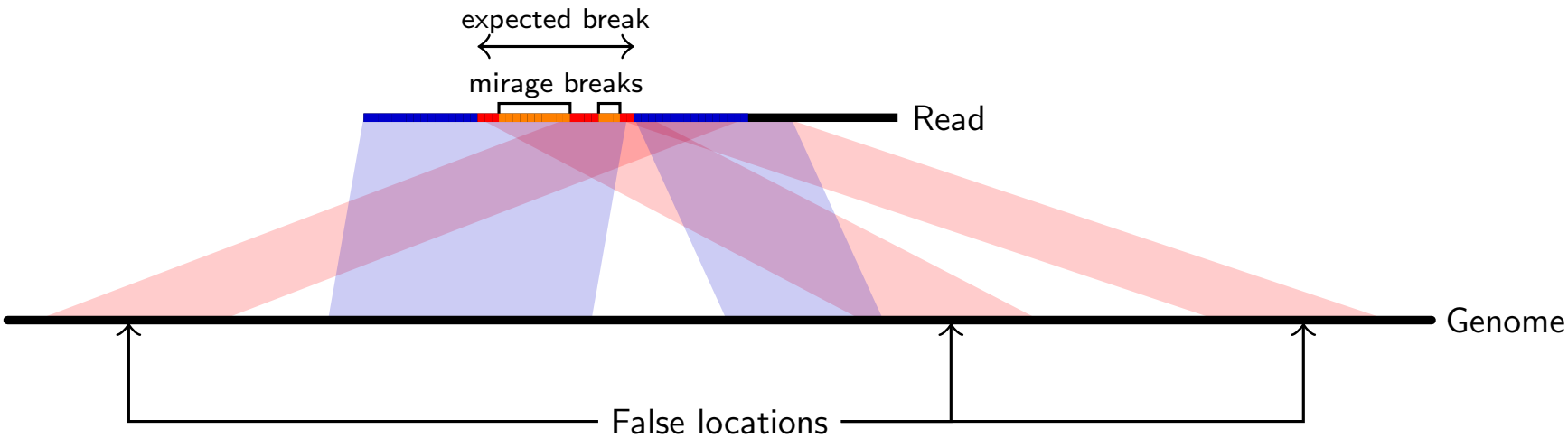
CRAC, un outil pour le RNA-Seq

Deletion

$\ell < L$



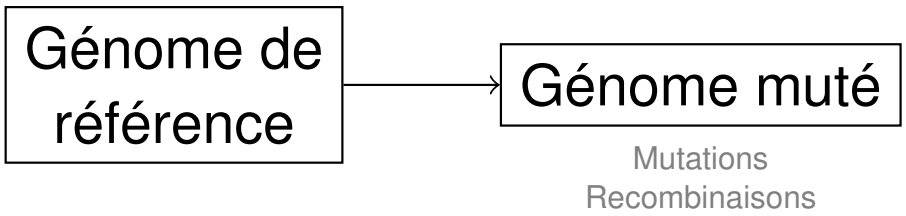
CRAC, un outil pour le RNA-Seq



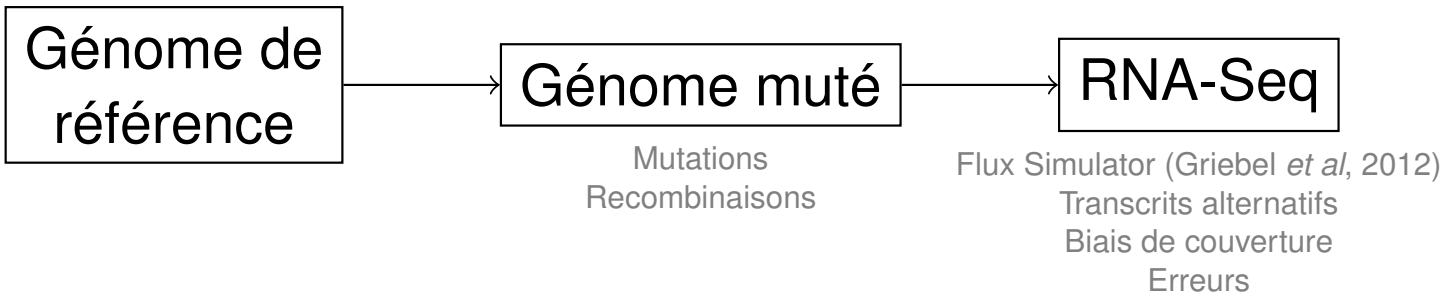
Données simulées en RNA-Seq

Génome de référence

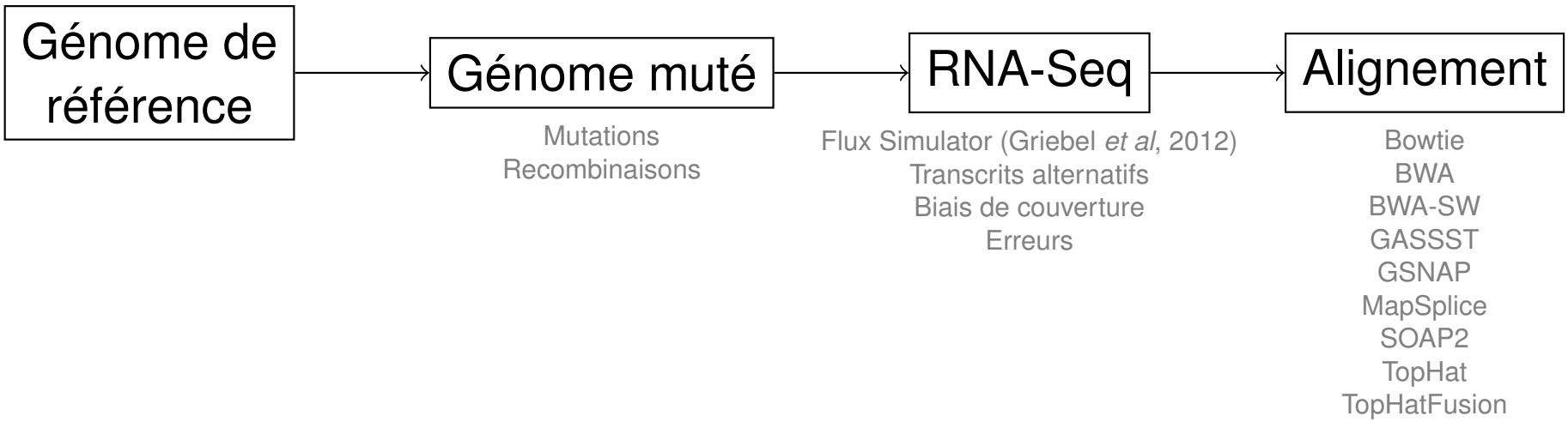
Données simulées en RNA-Seq



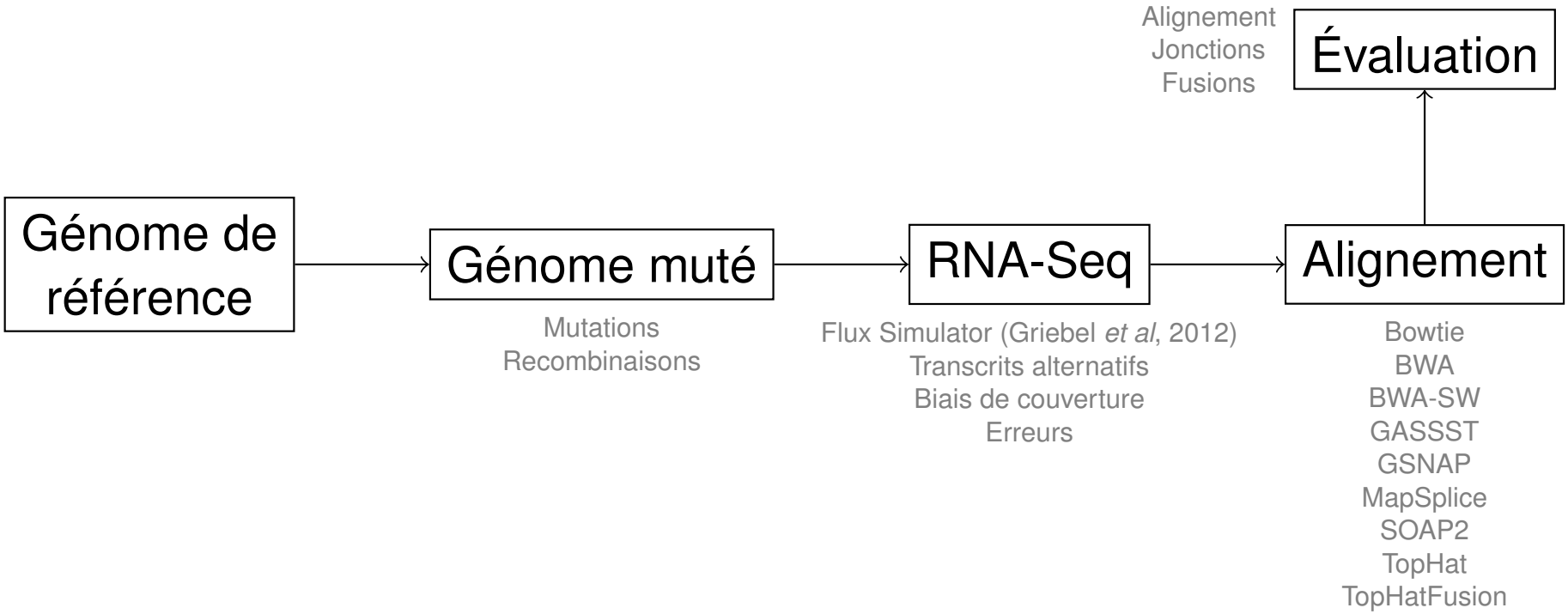
Données simulées en RNA-Seq



Données simulées en RNA-Seq



Données simulées en RNA-Seq



Données simulées

Type	SNV	Insertion	Délétion	Fusion	Erreurs
Génome complet	3 139 937	287 336	287 502	1 002	
45 M lectures (75 pb)	29 084	2 687	2 734	647	12 836 882
48 M lectures (200 pb)	52 971	4 810	4 901	914	38 840 045

Performances sur données simulées

Alignement (occurrences uniques)

Tool	75bp		200bp	
	Sensitivity	Precision	Sensitivity	Precision
Bowtie	75.42	99.59	55.72	**99.81**
BWA	79.29	99.13	68.66	96.86
CRAC	**94.51**	**99.72**	**95.89**	**99.78**
GASSST	70.73	99.09	59.43	97.86
GSNAP	**94.62**	**99.88**	**84.84**	99.28
SOAP2	77.6	99.52	56.08	99.78

Performances sur données simulées

Découverte de jonctions exon-exon

Tool	75bp		200bp	
	Sensitivity	Precision	Sensitivity	Precision
CRAC	79.43	**99.5**	**86.02**	**99.18**
GSNAP	**84.17**	97.03	72.94	97.09
MapSplice	79.89	**97.68**	**84.72**	**98.82**
TopHat	**84.96**	89.59	54.07	94.69

Performances sur données simulées

Découverte de transcrits de fusion

Tool	75bp		200bp	
	Sensitivity	Precision	Sensitivity	Precision
CRAC	**53.89**	**93.84**	**64.86**	**90.18**
MapSplice	2.33	0	**2.63**	**0.01**
TopHatFusion	**32.73**	42.02		
TopHatFusionPost	12.26	**97.22**		

Jonctions dans données réelles

ERR030856

RNA-Seq, Hi-Seq 2000
100 pb orienté
76 M lectures

Jonctions dans données réelles

ERR030856

RNA-Seq, Hi-Seq 2000
100 pb orienté
76 M lectures

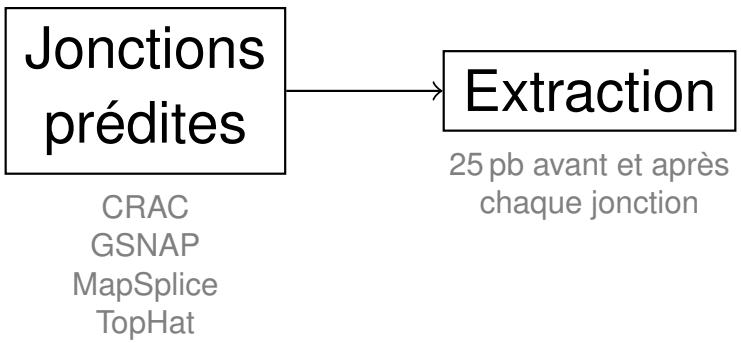
Jonctions prédites

CRAC
GSNAP
MapSplice
TopHat

Jonctions dans données réelles

ERR030856

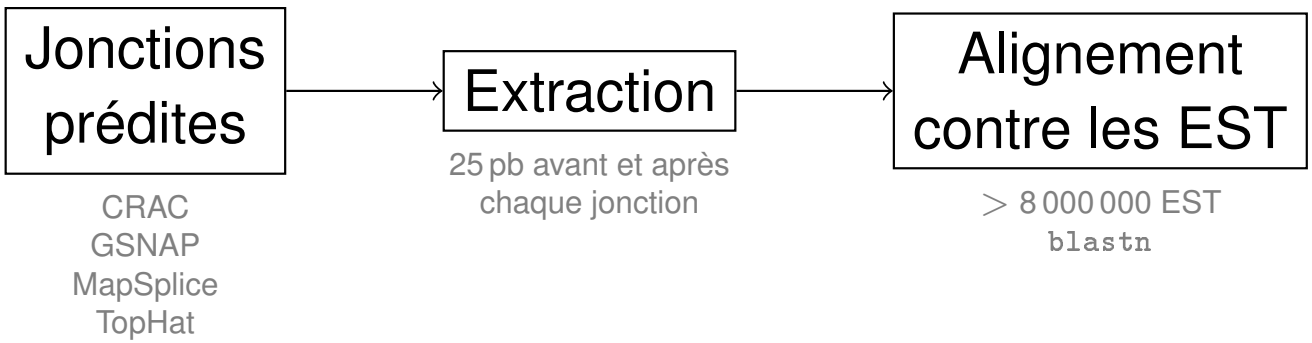
RNA-Seq, Hi-Seq 2000
100 pb orienté
76 M lectures



Jonctions dans données réelles

ERR030856

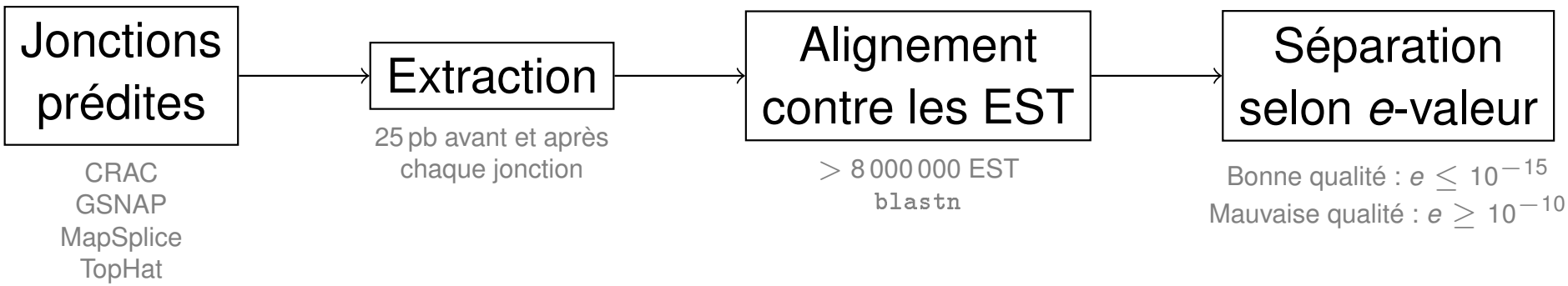
RNA-Seq, Hi-Seq 2000
100 pb orienté
76 M lectures



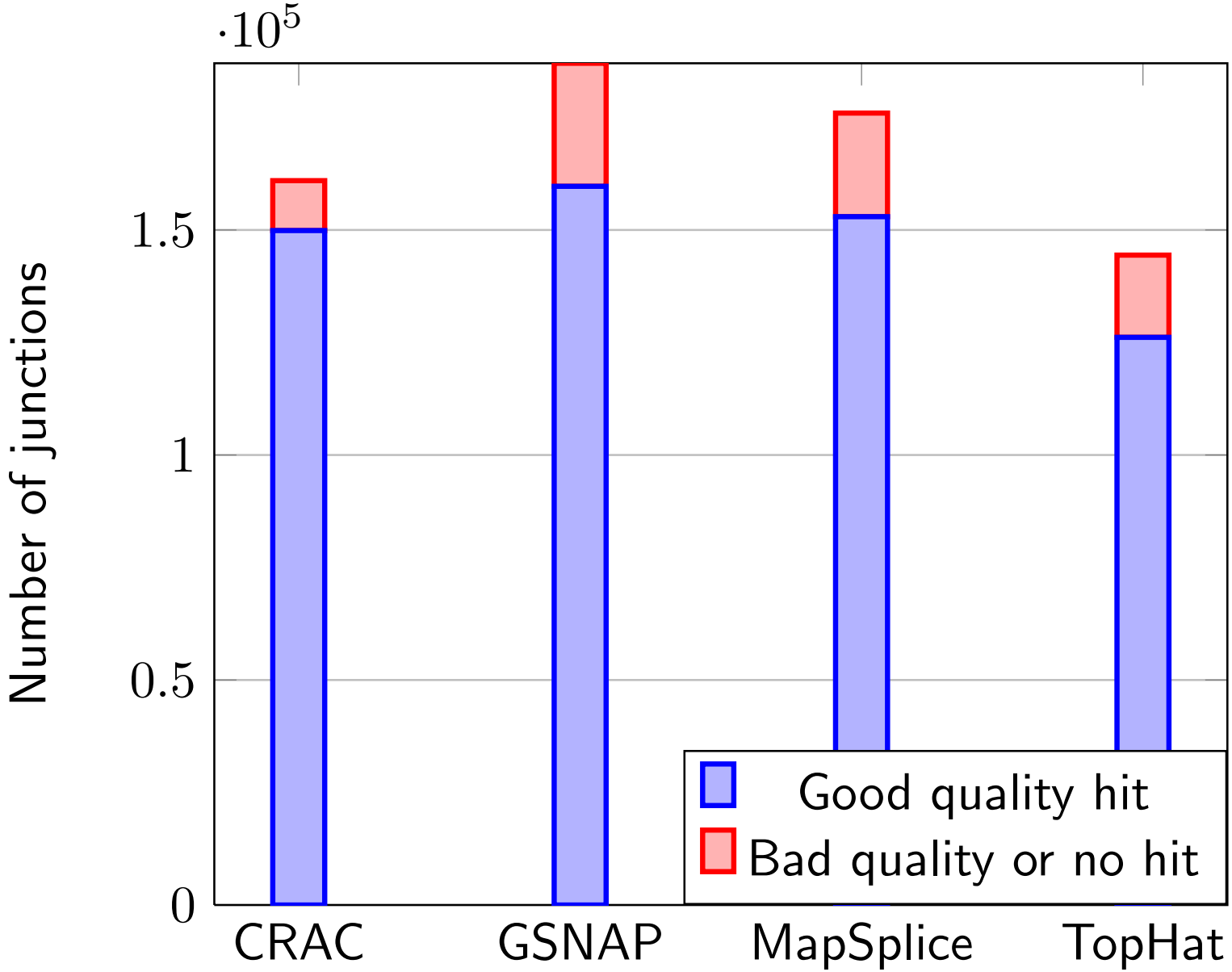
Jonctions dans données réelles

ERR030856

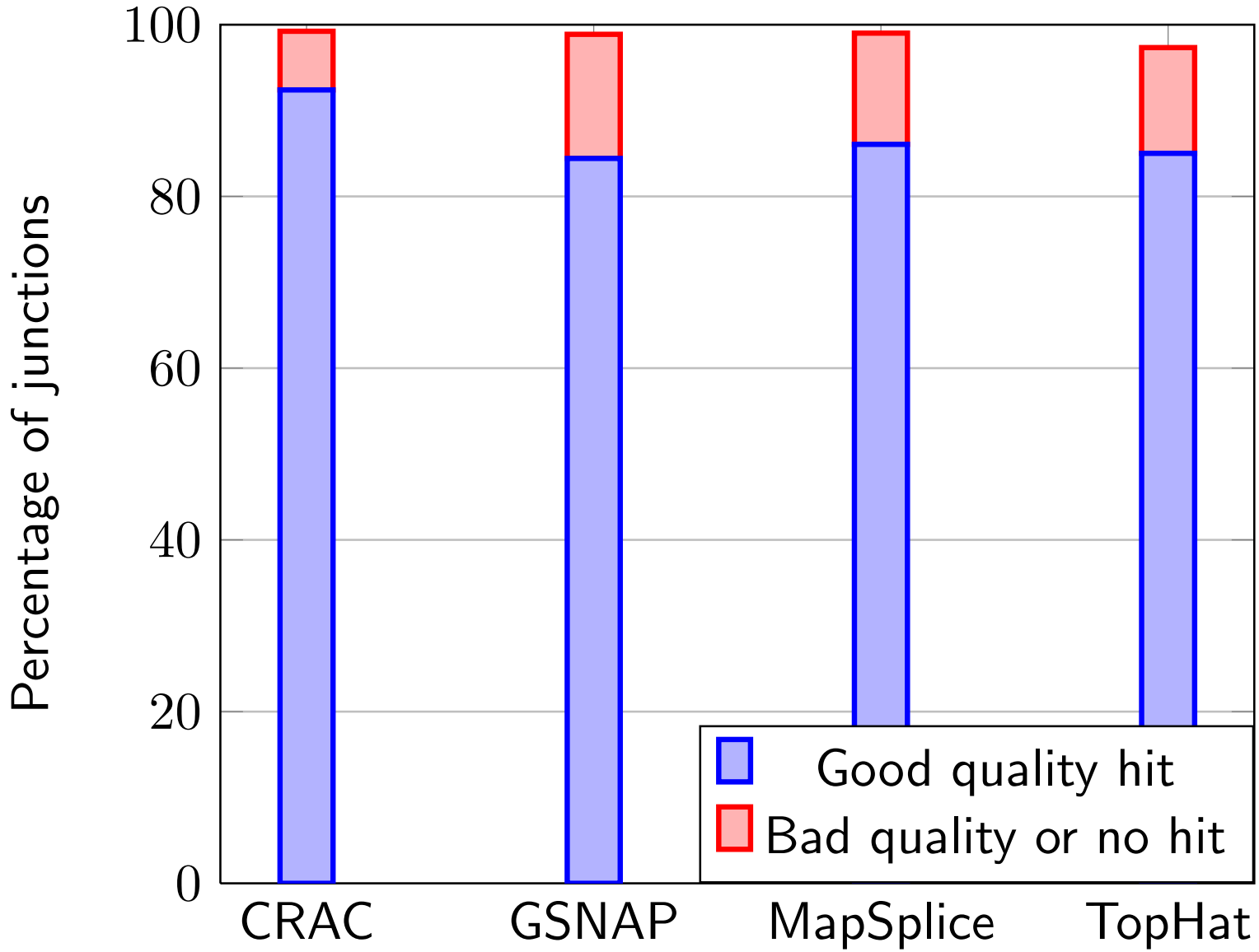
RNA-Seq, Hi-Seq 2000
100 pb orienté
76 M lectures



Jonctions dans données réelles



Jonctions dans données réelles

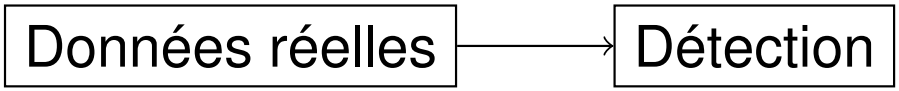


Fusions dans données réelles

Données réelles

Cellules de cancer du sein
27 fusions connues
Lectures 50 pb
Edgren *et al*, Genome Biology, 2011

Fusions dans données réelles



Cellules de cancer du sein
27 fusions connues
Lectures 50 pb
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

Fusions dans données réelles

Cellules de cancer du sein
27 fusions connues
Lectures 50 pb
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

Fusions dans données réelles

Cellules de cancer du sein
27 fusions connues
Lectures 50 pb
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

CRAC
455

Fusions dans données réelles

Cellules de cancer du sein
27 fusions connues
Lectures 50 pb
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

CRAC	**TopHatFusion**
455	193 163

Fusions dans données réelles

Cellules de cancer du sein
27 fusions connues
Lectures 50 pb
Edgren *et al*, Genome Biology, 2011

TopHatFusion
CRAC

CRAC	**TopHatFusion**
455	193 163
Validées : 20	Validées : 21

Perspectives. . .

Qualité : vraiment informative ?

Perspectives. . .

Qualité : vraiment informative ?

Nécessité d'évaluation plus complète

Bonus

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$$

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$$

Permutations circulaires

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $

$ C G A G A C G A A

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $

$ C G A G A C G A A

A $ C G A G A C G A

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $

$ C G A G A C G A A

A $ C G A G A C G A

A A $ C G A G A C G

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $
$ C G A G A C G A A
A $ C G A G A C G A
A A $ C G A G A C G
G A A $ C G A G A C
C G A A $ C G A G A
A C G A A $ C G A G
G A C G A A $ C G A
A G A C G A A $ C G
G A G A C G A A $ C

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $
$ C G A G A C G A A
A $ C G A G A C G A
A A $ C G A G A C G
G A A $ C G A G A C
C G A A $ C G A G A
A C G A A $ C G A G
G A C G A A $ C G A
A G A C G A A $ C G
G A G A C G A A $ C

Permutations triées

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $

$ C G A G A C G A A

A $ C G A G A C G A

A A $ C G A G A C G

G A A $ C G A G A C

C G A A $ C G A G A

A C G A A $ C G A G

G A C G A A $ C G A

A G A C G A A $ C G

G A G A C G A A $ C

Permutations triées

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires	Permutations triées
C G A G A C G A A $	$ C G A G A C G A A
$ C G A G A C G A A	
A $ C G A G A C G A	
A A $ C G A G A C G	
G A A $ C G A G A C	
C G A A $ C G A G A	
A C G A A $ C G A G	
G A C G A A $ C G A	
A G A C G A A $ C G	
G A G A C G A A $ C	

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $

$ C G A G A C G A A A

A $ C G A G A C G A

A A $ C G A G A C G

G A A $ C G A G A C

C G A A $ C G A G A

A C G A A $ C G A G

G A C G A A $ C G A

A G A C G A A $ C G

G A G A C G A A $ C

Permutations triées

$ C G A G A C G A A

A $ C G A G A C G A

A A $ C G A G A C G

C G A A $ C G A G A

C G A G A C G A A $

G A A $ C G A G A C

G A C G A A $ C G A

G A G A C G A A $ C

A G A C G A A $ C G

A C G A A $ C G A G

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $

$ C G A G A C G A A

A $ C G A G A C G A

A A $ C G A G A C G

G A A $ C G A G A C

C G A A $ C G A G A

A C G A A $ C G A G

G A C G A A $ C G A

A G A C G A A $ C G

G A G A C G A A $ C

Permutations triées

$ C G A G A C G A A

A $ C G A G A C G A

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $

$ C G A G A C G A A A

A $ C G A G A C G A

A A $ C G A G A C G

G A A $ C G A G A C

C G A A $ C G A G A

A C G A A $ C G A G

G A C G A A $ C G A

A G A C G A A $ C G

G A G A C G A A $ C

Permutations triées

$ C G A G A C G A A

A $ C G A G A C G A

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$$

Permutations circulaires

C G A G A C G A A $

$ C G A G A C G A A

A $ C G A G A C G A

A A $ C G A G A C G

G A A $ C G A G A C

C G A A $ C G A G A

A C G A A $ C G A G

G A C G A A $ C G A

A G A C G A A $ C G

G A G A C G A A $ C

Permutations triées

$ C G A G A C G A A

A $ C G A G A C G A

A A $ C G A G A C G

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $
$ C G A G A C G A A
A $ C G A G A C G A
A A $ C G A G A C G
G A A $ C G A G A C
C G A A $ C G A G A
A C G A A $ C G A G
G A C G A A $ C G A
A G A C G A A $ C G
G A G A C G A A $ C

Permutations triées

$ C G A G A C G A A
A $ C G A G A C G A
A A $ C G A G A C G

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{C\,G\,A\,G\,A\,C\,G\,A\,A\,\$}}$$

Permutations circulaires

C G A G A C G A A $
$ C G A G A C G A A
A $ C G A G A C G A
A A $ C G A G A C G
G A A $ C G A G A C
C G A A $ C G A G A
A C G A A $ C G A G
G A C G A A $ C G A
A G A C G A A $ C G
G A G A C G A A $ C

Permutations triées

$ C G A G A C G A A
A $ C G A G A C G A
A A $ C G A G A C G
A C G A A $ C G A G

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$$

Permutations circulaires	Permutations triées
C G A G A C G A A $	$ C G A G A C G A A
$ C G A G A C G A A	A $ C G A G A C G A
A $ C G A G A C G A	A A $ C G A G A C G
A A $ C G A G A C G	A C G A A $ C G A G
G A A $ C G A G A C	A G A C G A A $ C G
C G A A $ C G A G A	C G A A $ C G A G A
A C G A A $ C G A G	C G A G A C G A A $
G A C G A A $ C G A	G A A $ C G A G A C
A G A C G A A $ C G	G A C G A A $ C G A
G A G A C G A A $ C	G A G A C G A A $ C

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$$

Permutations circulaires	Permutations triées
C G A G A C G A A $	$ C G A G A C G A A
$ C G A G A C G A A	A $ C G A G A C G A
A $ C G A G A C G A	A A $ C G A G A C G
A A $ C G A G A C G	A C G A A $ C G A G
G A A $ C G A G A C	A G A C G A A $ C G
C G A A $ C G A G A	C G A A $ C G A G A
A C G A A $ C G A G	C G A G A C G A A $
G A C G A A $ C G A	G A A $ C G A G A C
A G A C G A A $ C G	G A C G A A $ C G A
G A G A C G A A $ C	G A G A C G A A $ C

Transformée de Burrows-Wheeler

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$$

Permutations circulaires	Permutations triées
C G A G A C G A A $	$ C G A G A C G A A
$ C G A G A C G A A	A $ C G A G A C G A
A $ C G A G A C G A	A A $ C G A G A C G
A A $ C G A G A C G	A C G A A $ C G A G
G A A $ C G A G A C	A G A C G A A $ C G
C G A A $ C G A G A	C G A A $ C G A G A
A C G A A $ C G A G	C G A G A C G A A $
G A C G A A $ C G A	G A A $ C G A G A C
A G A C G A A $ C G	G A C G A A $ C G A
G A G A C G A A $ C	G A G A C G A A $ C

$$\mathrm{TBW}(T) = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{AAGGGA\$CAC}}$$

Fonction LF

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$$

F									L
$	C	G	A	G	A	C	G	A	A
A	$	C	G	A	G	A	C	G	A
A	A	$	C	G	A	G	A	C	G
A	C	G	A	A	$	C	G	A	G
A	G	A	C	G	A	A	$	C	G
C	G	A	A	$	C	G	A	G	A
C	G	A	G	A	C	G	A	A	$
G	A	A	$	C	G	A	G	A	C
G	A	C	G	A	A	$	C	G	A
G	A	G	A	C	G	A	A	$	C

Fonction LF

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$$

F	L
$	A
A	A
A	G
A	G
A	G
C	A
C	$
G	C
G	A
G	C

Fonction LF

$$T = \overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\underset{1\ 1\ 1\ 2\ 2\ 2\ 3\ 3\ 4\ 1}{\mathrm{CGAGACGAA\$}}}$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; C_1 \; G_1 \; A_1 \; G_2 \; A_2 \; C_2 \; G_3 \; A_3 \; A_4 \; \$_1$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; {}_1 \; {}_1 \; {}_2 \; {}_2 \; {}_2 \; {}_3 \; {}_3 \; {}_4$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

Fonction LF

$$T = \overset{0}{\mathrm{C}_1}\overset{1}{\mathrm{G}_1}\overset{2}{\mathrm{A}_1}\overset{3}{\mathrm{G}_2}\overset{4}{\mathrm{A}_2}\overset{5}{\mathrm{C}_2}\overset{6}{\mathrm{G}_3}\overset{7}{\mathrm{A}_3}\overset{8}{\mathrm{A}_4}\overset{9}{\$_1}$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

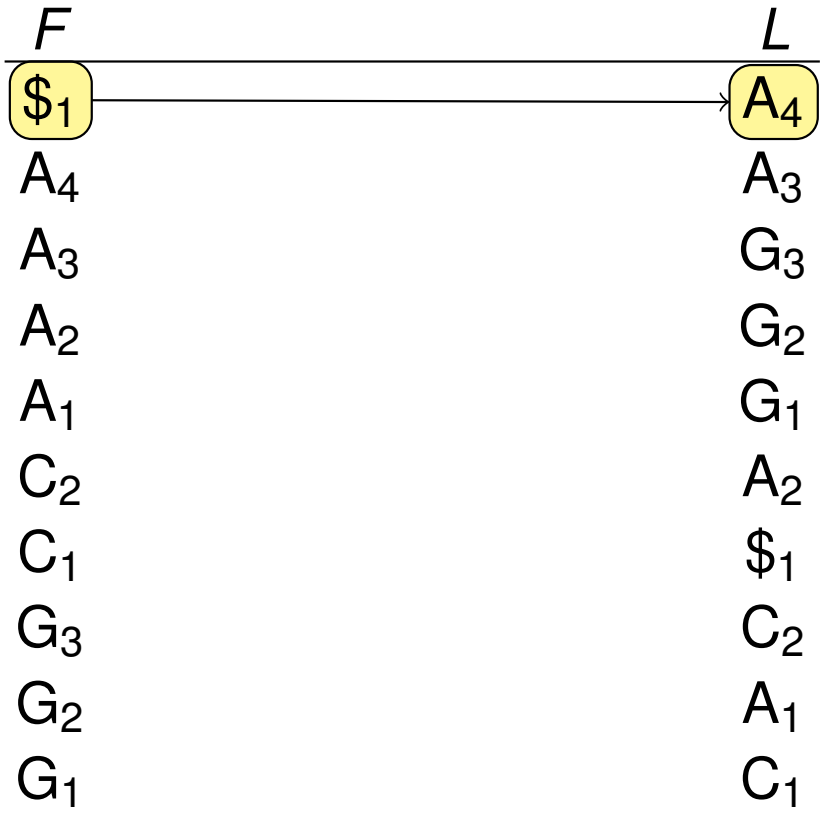
À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; C_1 \; G_1 \; A_1 \; G_2 \; A_2 \; C_2 \; G_3 \; A_3 \; A_4$$



Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

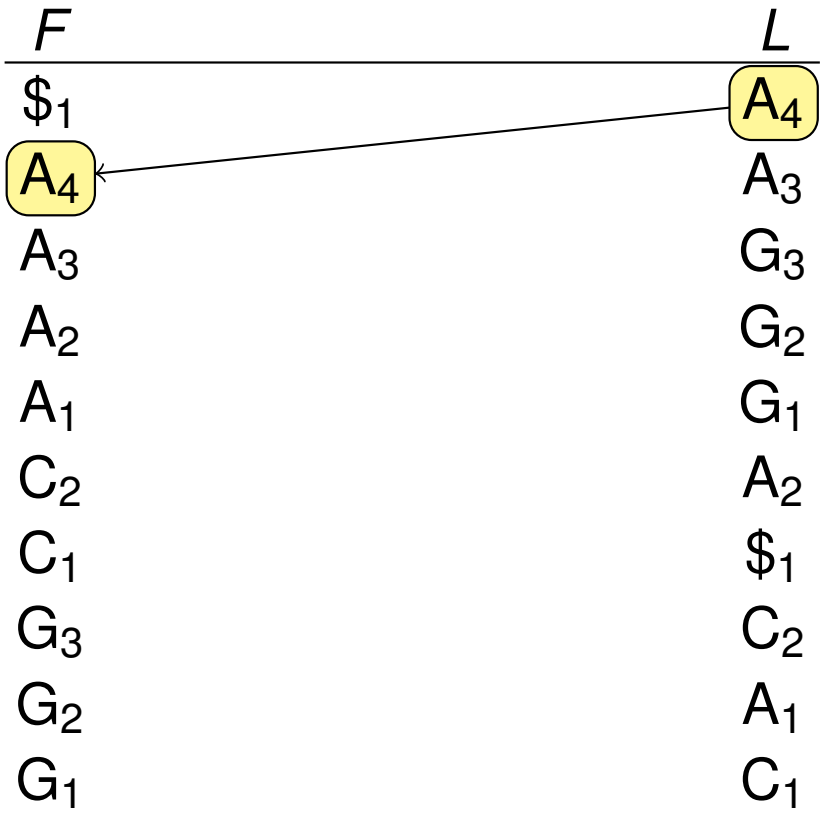
À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$A

Fonction LF

$$T = \overset{0}{C}_{1}\overset{1}{G}_{1}\overset{2}{A}_{1}\overset{3}{G}_{2}\overset{4}{A}_{2}\overset{5}{C}_{2}\overset{6}{G}_{3}\overset{7}{A}_{3}\overset{8}{A}_{4}\overset{9}{\$}_{1}$$



Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

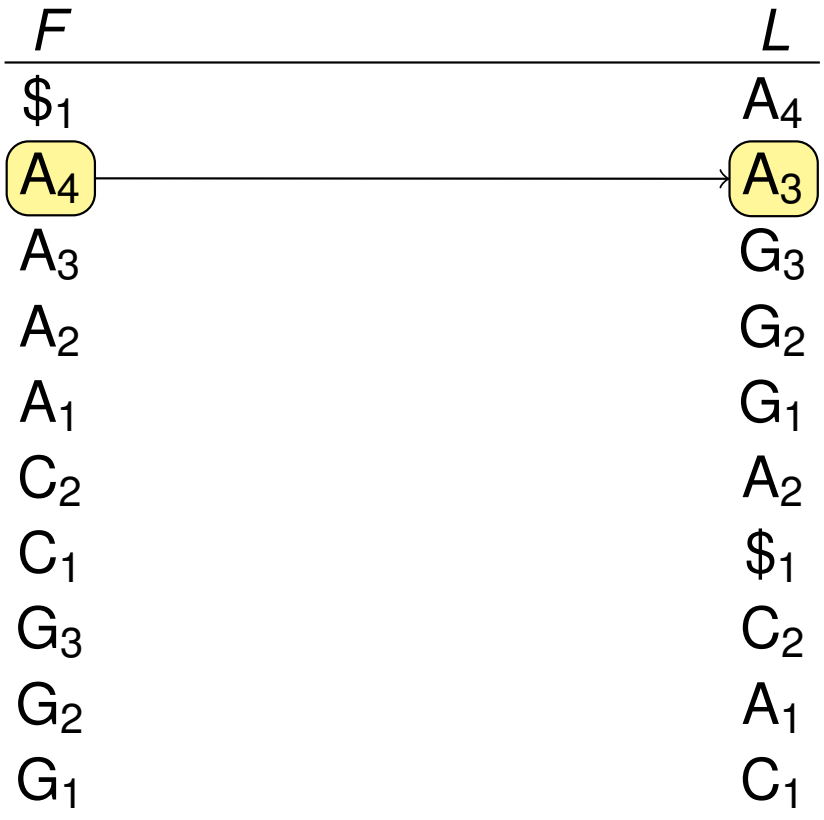
À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$A

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; C_1 \; G_1 \; A_1 \; G_2 \; A_2 \; C_2 \; G_3 \; A_3 \; A_4 \; \$_1$$



Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

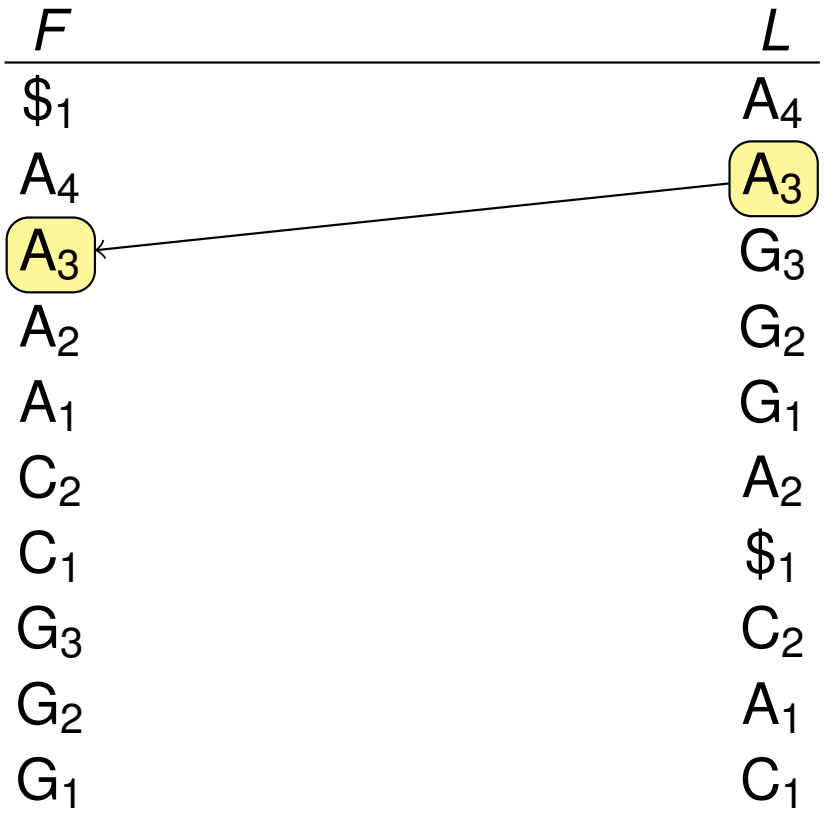
À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$AA

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; {}_1 \; {}_1 \; {}_2 \; {}_2 \; {}_2 \; {}_3 \; {}_3 \; {}_4 \; {}_1$$



Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$AA

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; C_1 \; G_1 \; A_1 \; G_2 \; A_2 \; C_2 \; G_3 \; A_3 \; A_4 \; \$_1$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

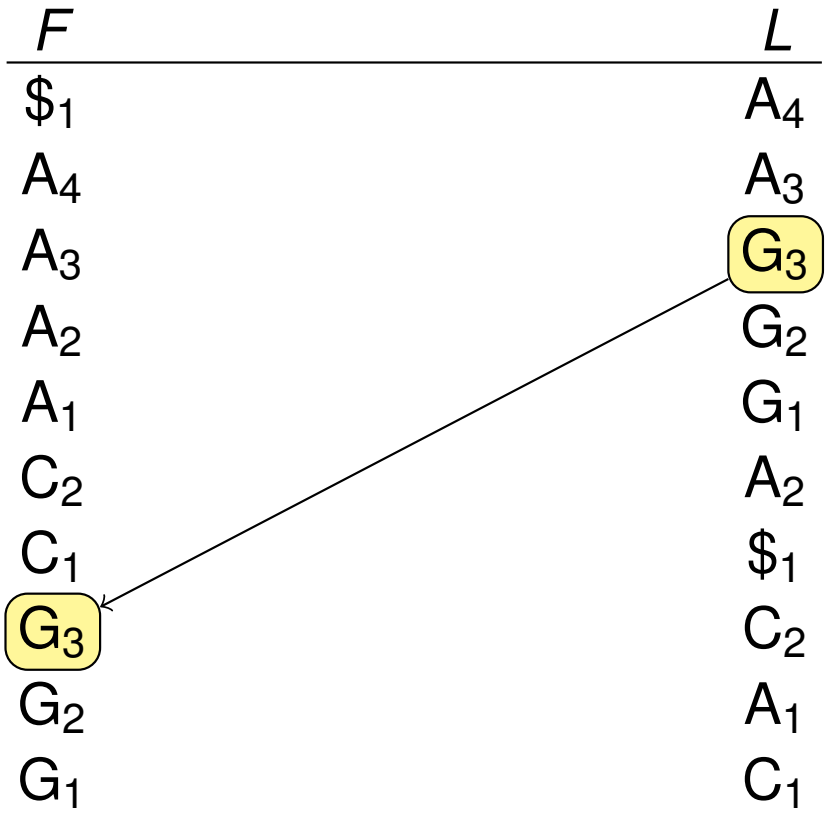
À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

\$AAG

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; C_1 \; G_1 \; A_1 \; G_2 \; A_2 \; C_2 \; G_3 \; A_3 \; A_4$$



Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$AAG

Fonction LF

$$T = \overset{0}{C}_{1}\overset{1}{G}_{1}\overset{2}{A}_{1}\overset{3}{G}_{2}\overset{4}{A}_{2}\overset{5}{C}_{2}\overset{6}{G}_{3}\overset{7}{A}_{3}\overset{8}{A}_{4}\overset{9}{\$}_{1}$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

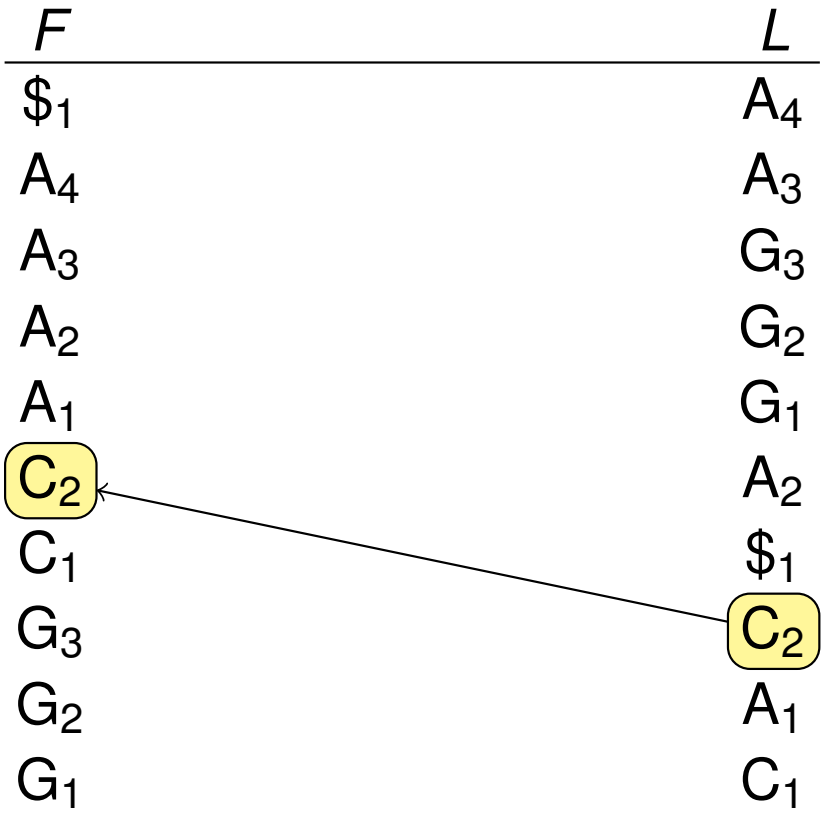
À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$AAGC

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; {}_1 \; {}_1 \; {}_2 \; {}_2 \; {}_2 \; {}_3 \; {}_3 \; {}_4$$



Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$AAGC

Fonction LF

$$T = \overset{0}{C}\overset{1}{G}\overset{2}{A}\overset{3}{G}\overset{4}{A}\overset{5}{C}\overset{6}{G}\overset{7}{A}\overset{8}{A}\overset{9}{\$}_1 \; C_1 \; G_1 \; A_1 \; G_2 \; A_2 \; G_3 \; A_3 \; A_4 \; \$_1$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

\$AAGCA

Fonction LF

$$T = \overset{0}{C}_{1}\overset{1}{G}_{1}\overset{2}{A}_{1}\overset{3}{G}_{2}\overset{4}{A}_{2}\overset{5}{C}_{2}\overset{6}{G}_{3}\overset{7}{A}_{3}\overset{8}{A}_{4}\overset{9}{\$}_{1}$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

$AAGCA

Fonction LF

$$T = \overset{0}{C}_{1}\overset{1}{G}_{1}\overset{2}{A}_{1}\overset{3}{G}_{2}\overset{4}{A}_{2}\overset{5}{C}_{2}\overset{6}{G}_{3}\overset{7}{A}_{3}\overset{8}{A}_{4}\overset{9}{\$}_{1}$$

F	L
$\$_1$	A_4
A_4	A_3
A_3	G_3
A_2	G_2
A_1	G_1
C_2	A_2
C_1	$\$_1$
G_3	C_2
G_2	A_1
G_1	C_1

Propriété

Les lettres identiques sont dans le même ordre dans F et dans L.

Fonction LF

À partir de L et F, passage d'une lettre dans T à la précédente.

Récupération du texte

\$AAGCA $\cdots$

FM-index

- Introduit par Ferragina et Manzini (2000) ;

FM-index

- Introduit par Ferragina et Manzini (2000) ;
- structure d'indexation compressée basée sur la transformée de Burrows-Wheeler ;

FM-index

- Introduit par Ferragina et Manzini (2000) ;
- structure d'indexation compressée basée sur la transformée de Burrows-Wheeler ;
- utilise :
  - la transformée de Burrows-Wheeler (lettres) ;
  - un échantillon de la table des suffixes (positions).

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\text{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\text{CGAGACGAA\$}}$

F									L
\$	C	G	A	G	A	C	G	A	A
A	\$	C	G	A	G	A	C	G	A
A	A	\$	C	G	A	G	A	C	G
A	C	G	A	A	\$	C	G	A	G
A	G	A	C	G	A	A	\$	C	G
C	G	A	A	\$	C	G	A	G	A
C	G	A	G	A	C	G	A	A	\$
G	A	A	\$	C	G	A	G	A	C
G	A	C	G	A	A	\$	C	G	A
G	A	G	A	C	G	A	A	\$	C

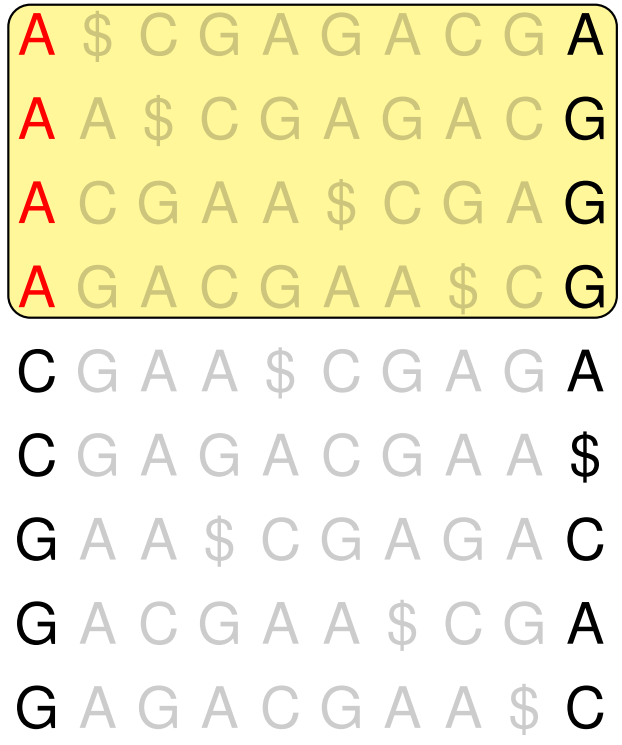
Recherche de motifs

Recherchons $P =$ $\overset{0}{\mathrm{A}}\overset{1}{\mathrm{G}}\overset{2}{\mathrm{A}}$ dans $T =$ $\overset{0}{\mathrm{C}}\overset{1}{\mathrm{G}}\overset{2}{\mathrm{A}}\overset{3}{\mathrm{G}}\overset{4}{\mathrm{A}}\overset{5}{\mathrm{C}}\overset{6}{\mathrm{G}}\overset{7}{\mathrm{A}}\overset{8}{\mathrm{A}}\overset{9}{\$}$

F									L
$	C	G	A	G	A	C	G	A	A
A	$	C	G	A	G	A	C	G	A
A	A	$	C	G	A	G	A	C	G
A	C	G	A	A	$	C	G	A	G
A	G	A	C	G	A	A	$	C	G
C	G	A	A	$	C	G	A	G	A
C	G	A	G	A	C	G	A	A	$
G	A	A	$	C	G	A	G	A	C
G	A	C	G	A	A	$	C	G	A
G	A	G	A	C	G	A	A	$	C

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\text{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\text{CGAGACGAA\$}}$

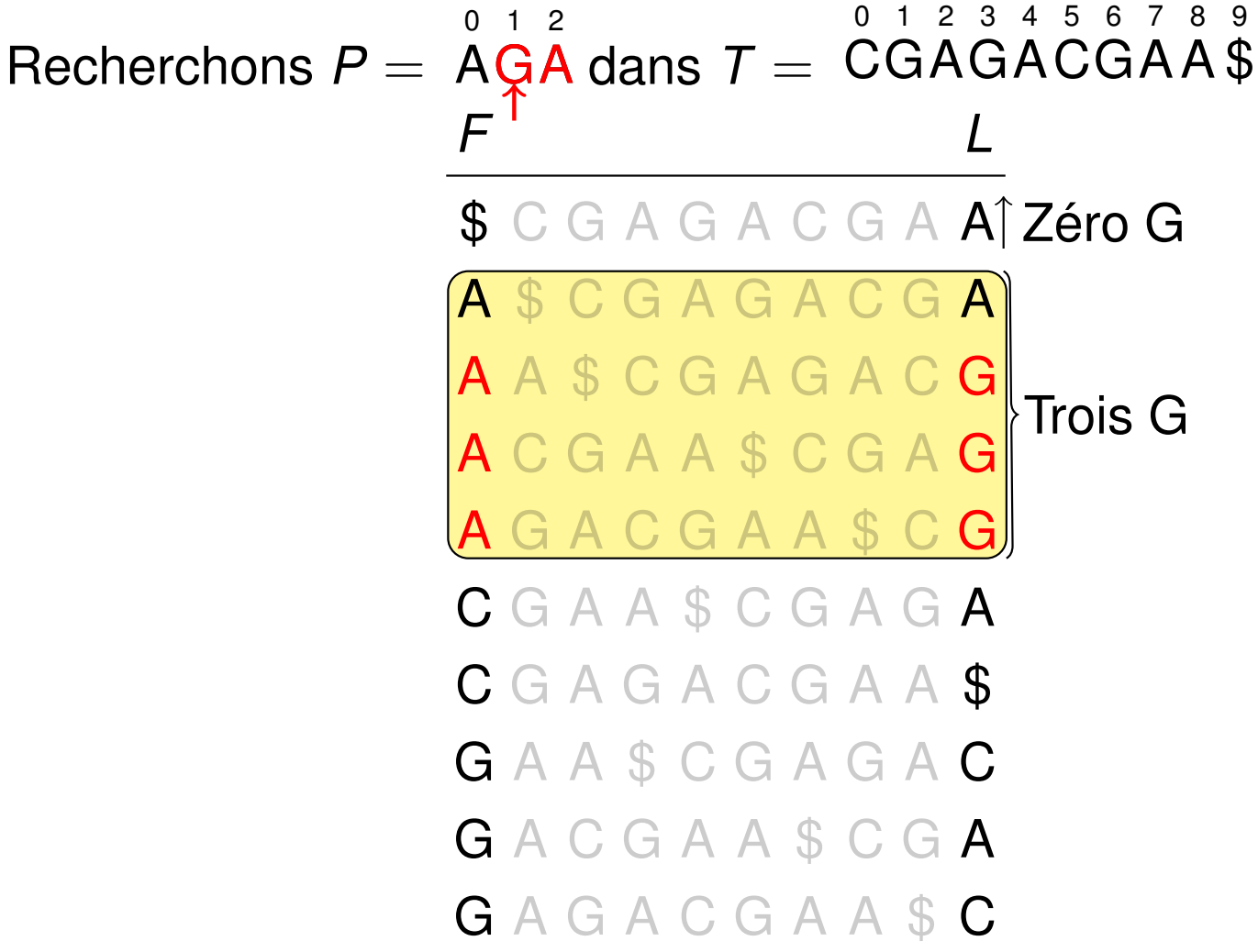


Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\mathrm{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$

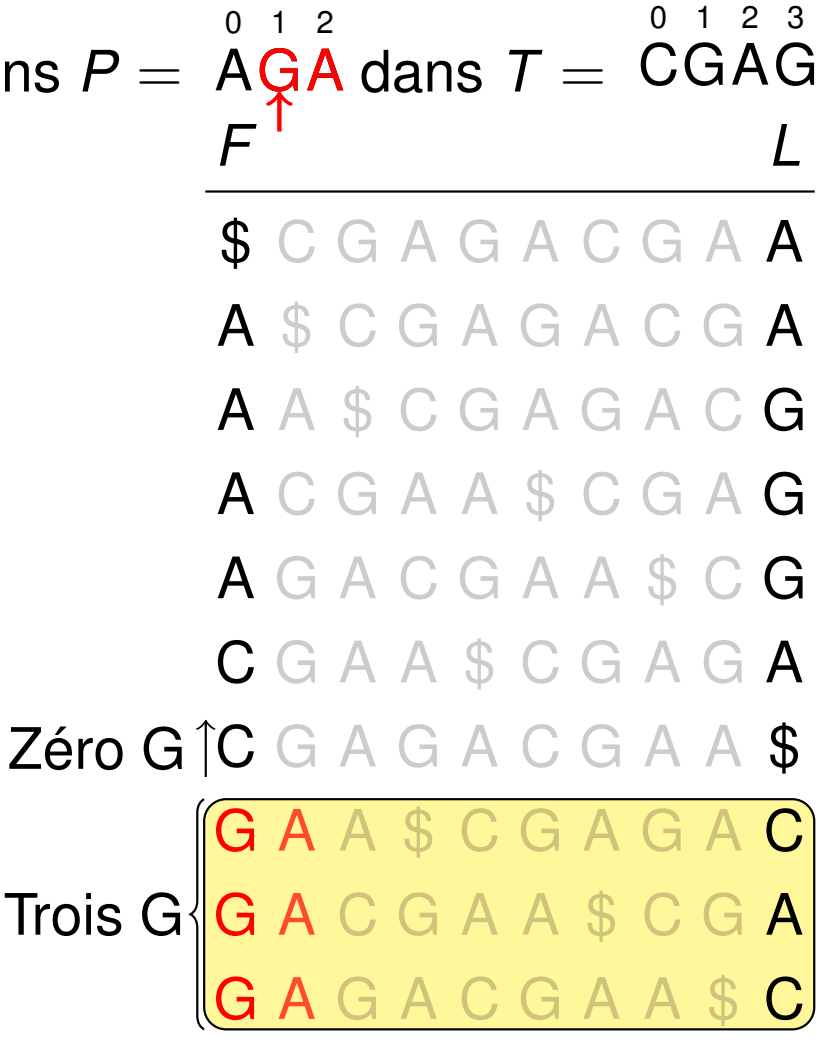
F	L
$	A
A	$
A	A
A	C
A	G
C	A
C	A
G	$
G	C
G	A
G	C

Recherche de motifs



Recherche de motifs

Recherchons $P =$ A G A dans $T =$ C G A G A C G A A $



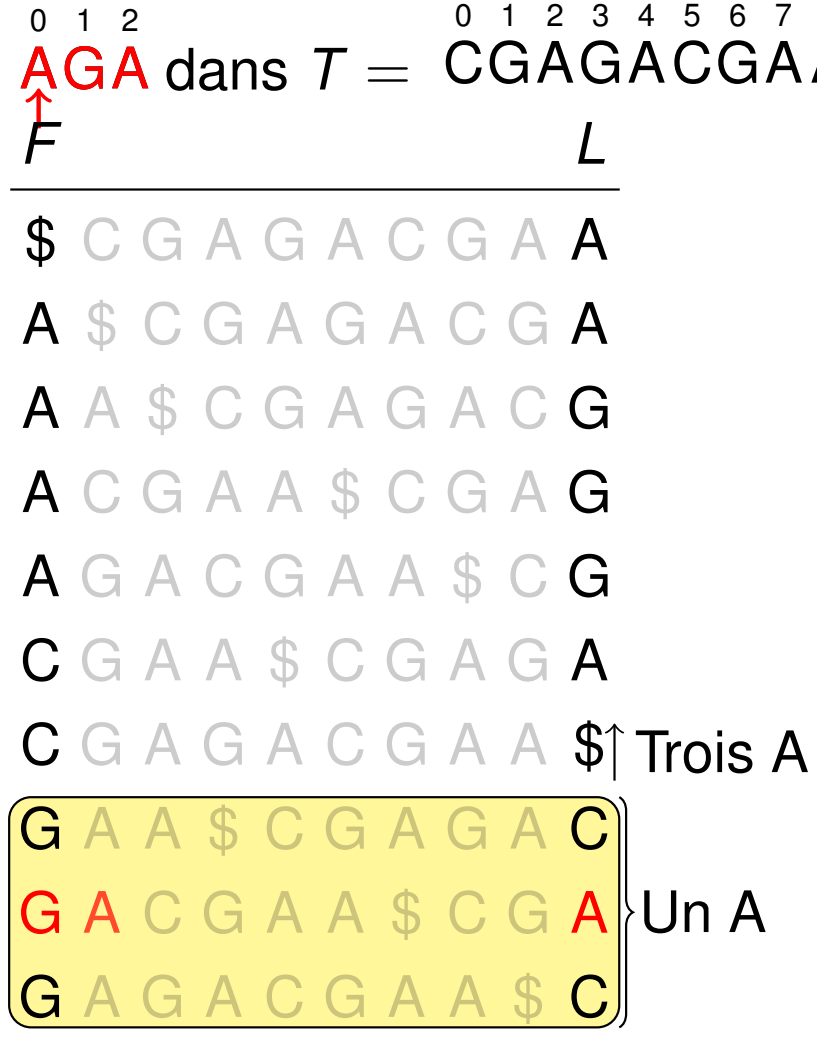
Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\mathrm{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$

F									L
$	C	G	A	G	A	C	G	A	A
A	$	C	G	A	G	A	C	G	A
A	A	$	C	G	A	G	A	C	G
A	C	G	A	A	$	C	G	A	G
A	G	A	C	G	A	A	$	C	G
C	G	A	A	$	C	G	A	G	A
C	G	A	G	A	C	G	A	A	$
G	A	A	$	C	G	A	G	A	C
G	A	C	G	A	A	$	C	G	A
G	A	G	A	C	G	A	A	$	C

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\mathrm{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\mathrm{CGAGACGAA\$}}$



Recherche de motifs

Recherchons $P =$ A G A dans $T =$ C G A G A C G A A $

	F									L
	$	C	G	A	G	A	C	G	A	A
	A	$	C	G	A	G	A	C	G	A
	A	A	$	C	G	A	G	A	C	G
Trois A	A	C	G	A	A	$	C	G	A	G
Un A	A	G	A	C	G	A	A	$	C	G
	C	G	A	A	$	C	G	A	G	A
	C	G	A	G	A	C	G	A	A	$
	G	A	A	$	C	G	A	G	A	C
	G	A	C	G	A	A	$	C	G	A
	G	A	G	A	C	G	A	A	$	C

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\text{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\text{CGAGACGAA\$}}$

F									L
\$	C	G	A	G	A	C	G	A	A
A	\$	C	G	A	G	A	C	G	A
A	A	\$	C	G	A	G	A	C	G
A	C	G	A	A	\$	C	G	A	G
A	G	A	C	G	A	A	\$	C	G
C	G	A	A	\$	C	G	A	G	A
C	G	A	G	A	C	G	A	A	\$
G	A	A	\$	C	G	A	G	A	C
G	A	C	G	A	A	\$	C	G	A
G	A	G	A	C	G	A	A	\$	C

Il y a une seule occurrence de AGA dans T.

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\text{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\text{CGAGACGAA\$}}$

	F	L
9	\$ C G A G A C G A	A
8	A \$ C G A G A C G	A
7	A A \$ C G A G A C	G
4	A C G A A \$ C G A	G
2	A G A C G A A \$ C	G
5	C G A A \$ C G A G	A
0	C G A G A C G A A	\$
6	G A A \$ C G A G A	C
3	G A C G A A \$ C G	A
1	G A G A C G A A \$	C

Il y a une seule occurrence de AGA dans T.

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\text{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\text{CGAGACGAA\$}}$

	F	L
9	\$ C G A G A C G A	A
8	A \$ C G A G A C G	A
7	A A \$ C G A G A C	G
4	A C G A A \$ C G A	G
2	A G A C G A A \$ C	G
5	C G A A \$ C G A G	A
0	C G A G A C G A A	\$
6	G A A \$ C G A G A	C
3	G A C G A A \$ C G	A
1	G A G A C G A A \$	C

Il y a une seule occurrence de AGA dans T.

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{AGA}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{CGAGACGAA\$}$

Il y a une seule occurrence de AGA dans T.

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{AGA}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{CGAGACGAA\$}$

Il y a une seule occurrence de AGA dans T.

Recherche de motifs

Recherchons $P =$ $\overset{0\ 1\ 2}{\text{AGA}}$ dans $T =$ $\overset{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9}{\text{CGAGACGAA\$}}$

	F	L
9	\$ C G A G A C G A	A
	A \$ C G A G A C G	A
	A A \$ C G A G A C	G
	A C G A A \$ C G A	G
	A G A C G A A \$ C	G
	C G A A \$ C G A G	A
0	C G A G A C G A A	\$
6	G A A \$ C G A G A	C
3	G A C G A A \$ C G	A
	G A G A C G A A \$	C

Il y a une seule occurrence de AGA dans T.