

Algorithmes de mapping

Matthias Zytnicki
INRA MIAT

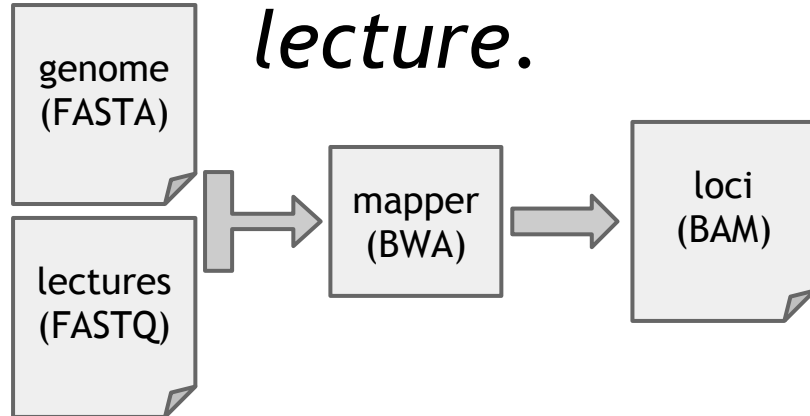
Plan

- Introduction
- Algorithmes pour l'ADN
- Algorithmes pour l'ARN
- Quasi-mapping
- Discussion

Qu'est-ce que le mapping ?

Définition

*Prédiction du locus
d'où vient une
lecture.*



Implémentation

Liste des régions
comportant le moins
de différences.

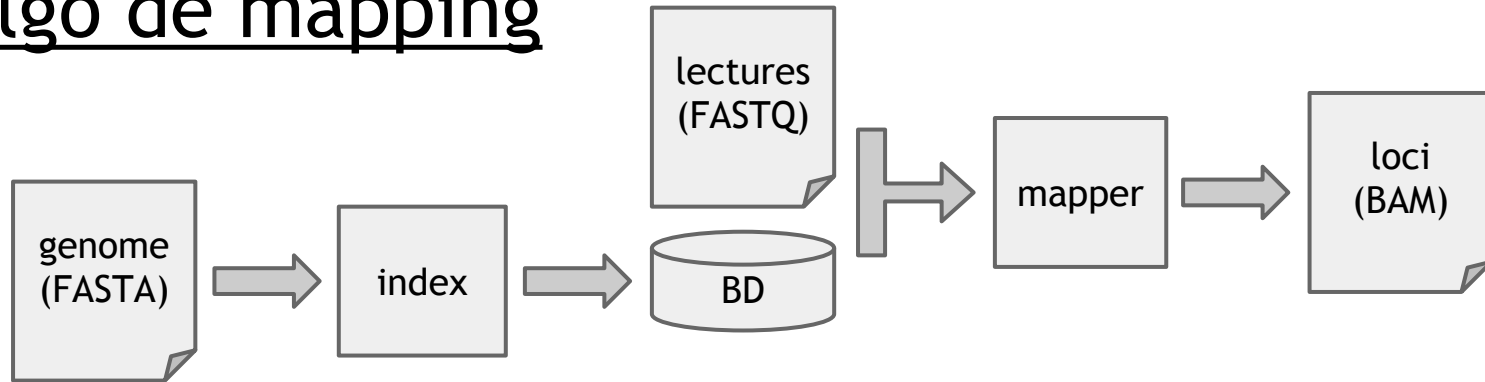
Mapping \neq alignement

ACGGGGTACGTACGT
---ACGTACGTACGT

Qu'est-ce qu'un algorithme ?

C'est une méthode implantée dans un ordinateur pour répondre à une question.

Algo de mapping



Plan

- Introduction
- Algorithmes pour l'ADN
- Algorithmes pour l'ARN
- Quasi-mapping
- Discussion

Algo 1 : *Seed and extend* (MAQ)

1. Énumération des k -mers dans le génome
(construction de la base)

Génome : CAGCTAGCTAGAATC

K -mers : (\emptyset) CAGC - (1) AGCT - (2) GCTA - ...

Tri : AATC (11) - AGCT (1,5) - CAGC (\emptyset) - ...

Rangement dans une table de hashage.

Algo 1 : *Seed and extend* (MAQ)

2. Essai des k -mers d'une lecture (*seeding*)

Lecture: ATAGCGA

- ATAG : -
- TAGC : 4

AATC : 11
AGCT : 1, 5
CAGC : 0
CTAG : 3, 7
GAAT : 10
GCTA : 2, 6
TAGA : 8
TAGC : 4

Algo 1 : *Seed and extend* (MAQ)

3. Extension approché (*extension*)

Génome : CAGCTAGCTAGAATC

Lecture : ATAGCGA

Algo 1 : *Seed and extend* (MAQ)

Remarques

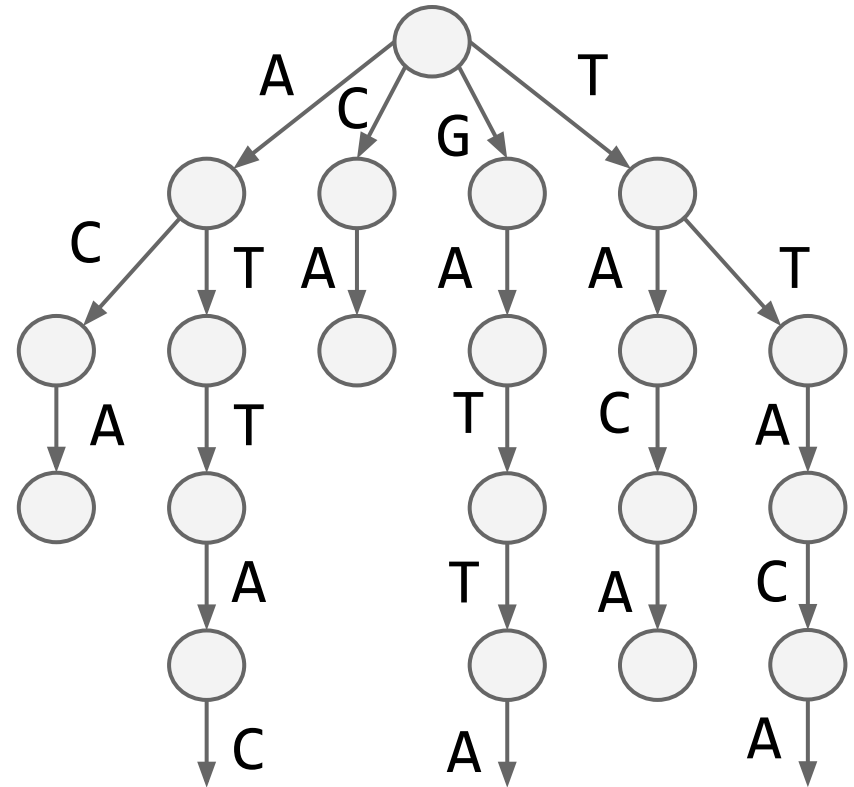
- Étape 1: lente
- Étape 2 : rapide
- Étape 3 : lente
- k grand : spécificité
- k petit : sensibilité
- k -mer très répété : non enregistré

CAGCTAGCTAGAATC
ATAGCGA

Algo 2: Burrows-Wheeler (BWA, Bowtie)

1. Transformer le génome en un “arbre”

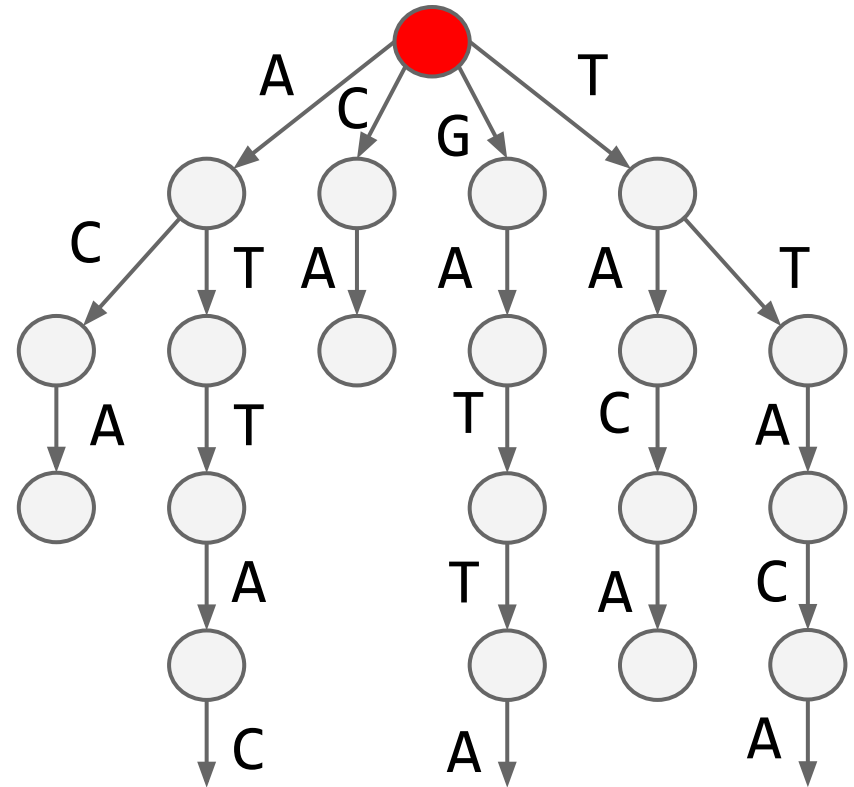
Génome : GATTACA



Algo 2: Burrows-Wheeler (BWA, Bowtie)

2. Explorer l'arbre en lisant une lecture

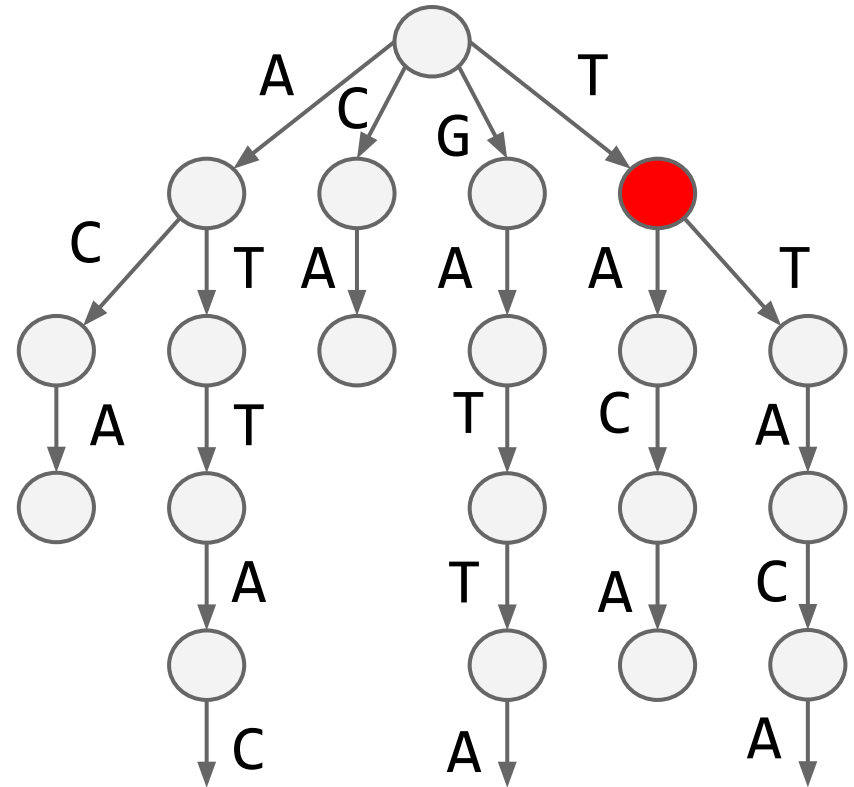
Lecture : TAC



Algo 2: Burrows-Wheeler (BWA, Bowtie)

2. Explorer l'arbre en lisant une lecture

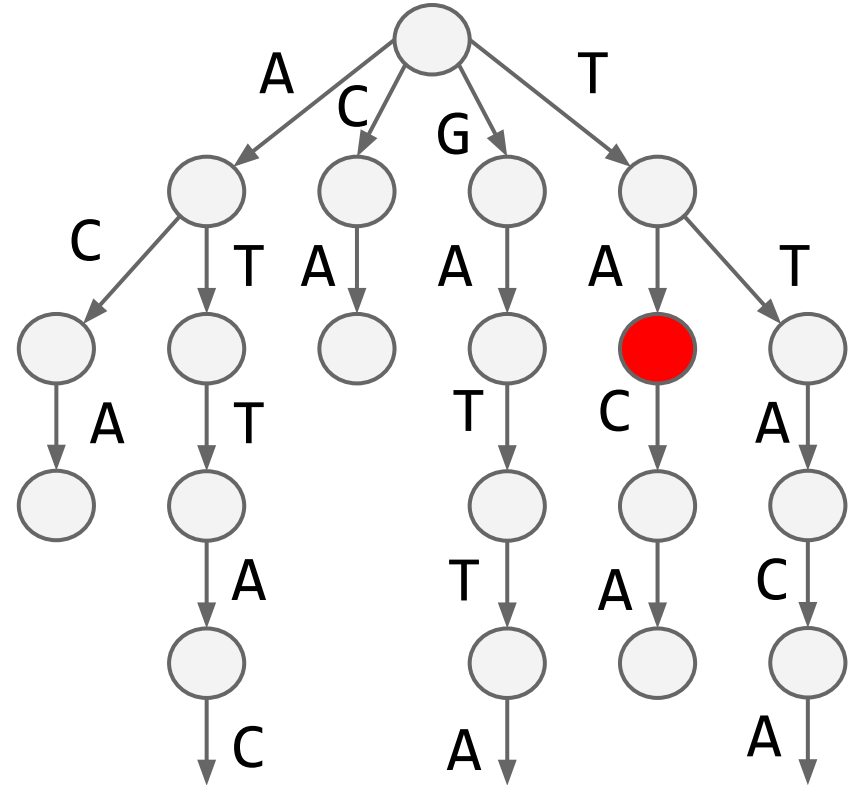
Lecture : TAC



Algo 2: Burrows-Wheeler (BWA, Bowtie)

2. Explorer l'arbre en lisant une lecture

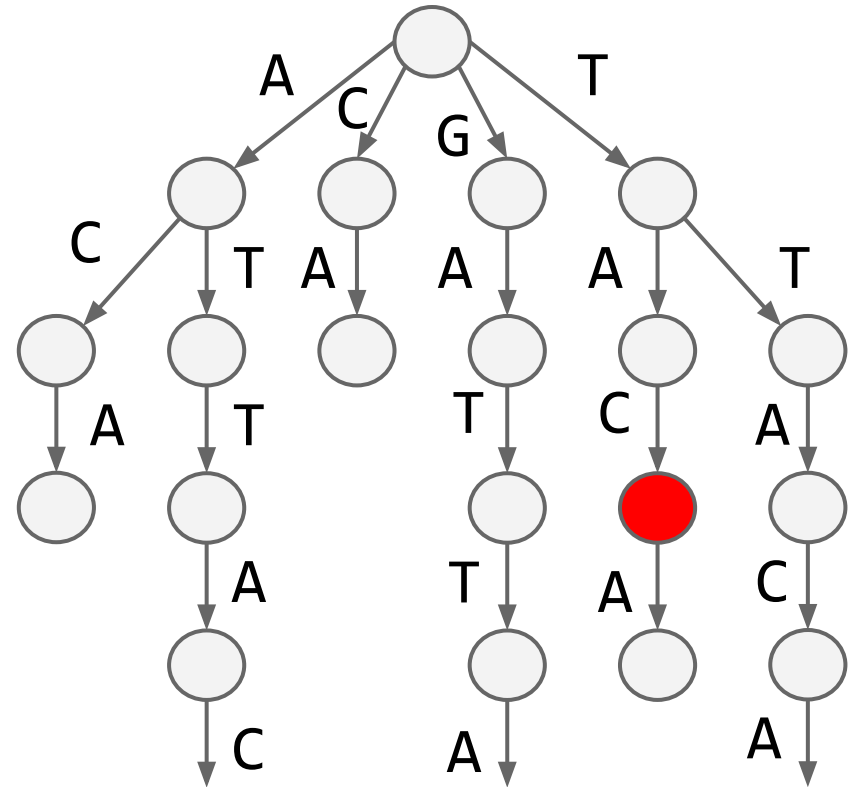
Lecture : TAC



Algo 2: Burrows-Wheeler (BWA, Bowtie)

2. Explorer l'arbre en lisant une lecture

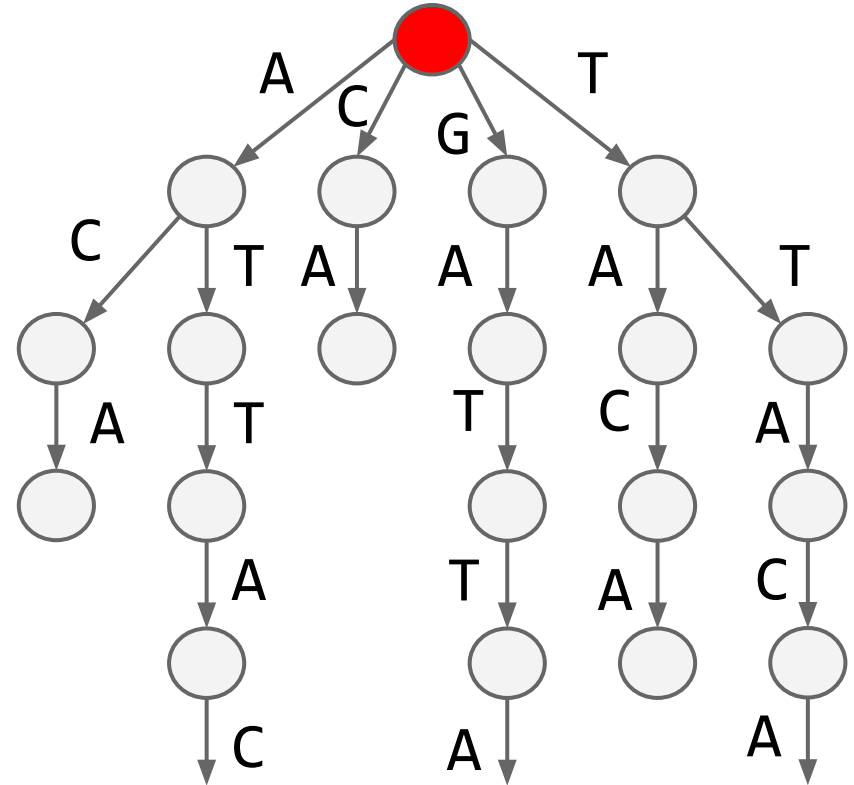
Lecture : TAC



Algo 2: Burrows-Wheeler (BWA, Bowtie)

2. Explorer l'arbre en lisant une lecture

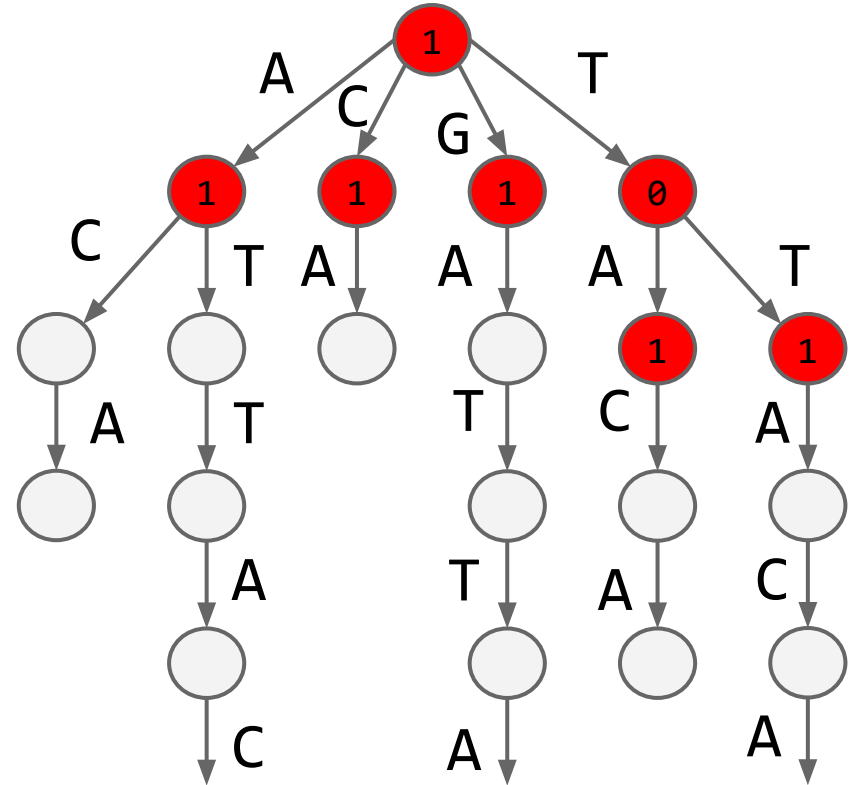
Lecture : TAT
(avec 1 erreur)



Algo 2: Burrows-Wheeler (BWA, Bowtie)

2. Explorer l'arbre en lisant une lecture

Lecture : TAT
(avec 1 erreur)

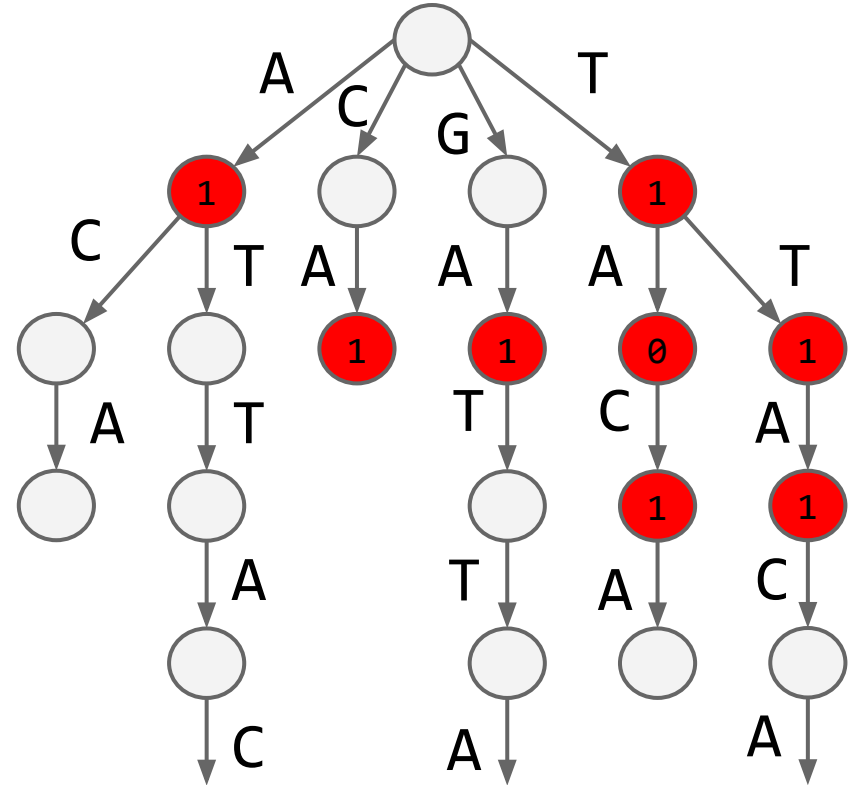


Algo 2: Burrows-Wheeler (BWA, Bowtie)

2. Explorer l'arbre en lisant une lecture

Lecture : **TAT**

(avec 1 erreur)

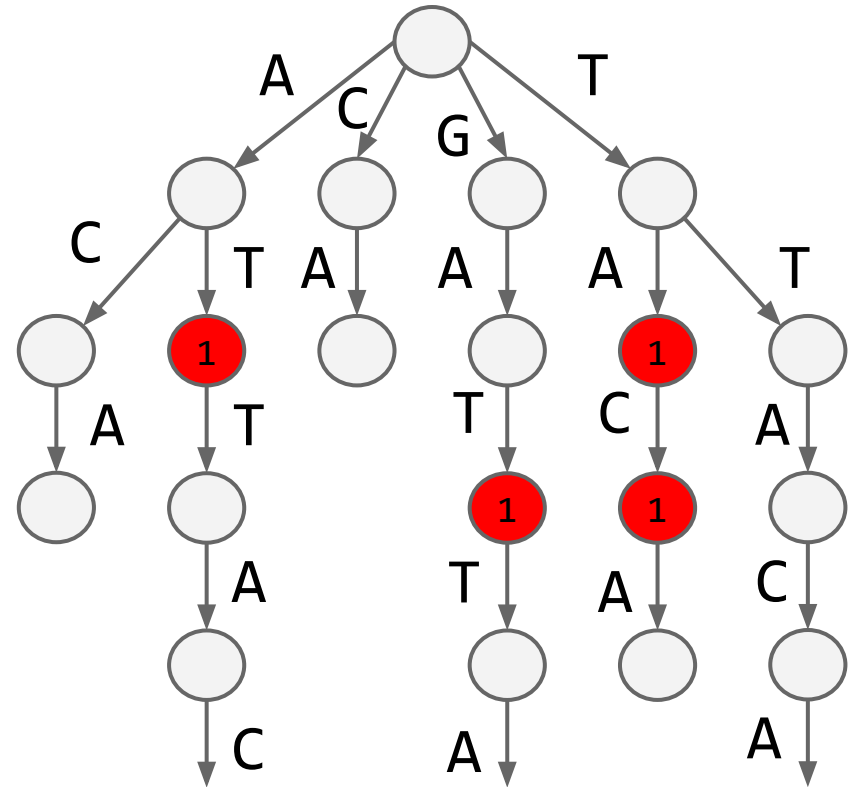


Algo 2: Burrows-Wheeler (BWA, Bowtie)

2. Explorer l'arbre en lisant une lecture

Lecture : **TAT**

(avec 1 erreur)



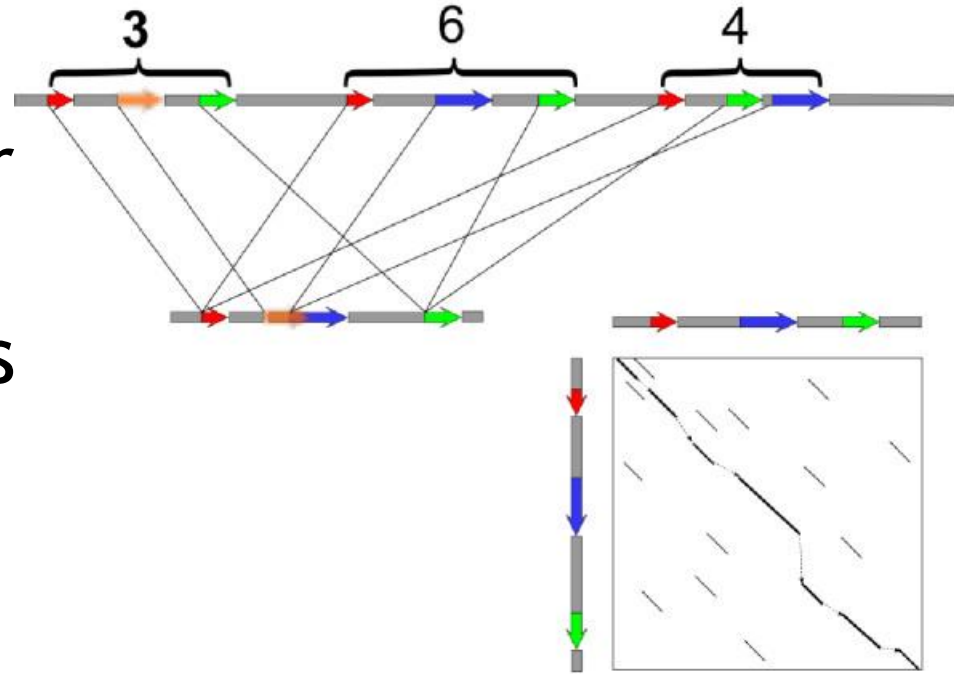
Algo 2: Burrows-Wheeler (BWA, Bowtie)

Remarques :

- Étape 1: lente
- Étape 2 sans erreur : rapide
- Étape 2 avec erreur : plutôt lent
- Pas de problème de répétition

Algo 3 : MEM (BWA-MEM, Blasr)

- Utilise Burrows-Wheeler pour des graines
- Recherche des régions “prometteuses”
- Aligne localement



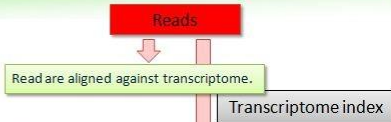
M. Chaisson & G. Tessler, *BMC Bioinformatics*, 2012

Plan

- Introduction
- Algorithmes pour l'ADN
- Algorithmes pour l'ARN
- Quasi-mapping
- Discussion

Algo 1 : TopHat2

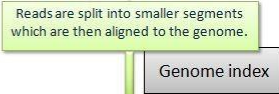
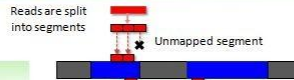
(1) Transcriptome alignment (optional)



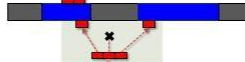
(2) Genome alignment



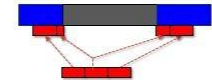
(3) Spliced alignment



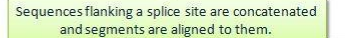
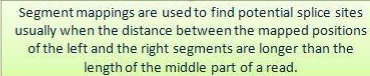
(3-1) Segment alignment to genome



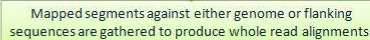
(3-2) Identification of splice sites (including indels and fusion break points)



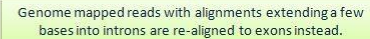
(3-3) Segments aligned to junction flanking sequences



(3-4) Segment alignments stitched together to form whole read alignments

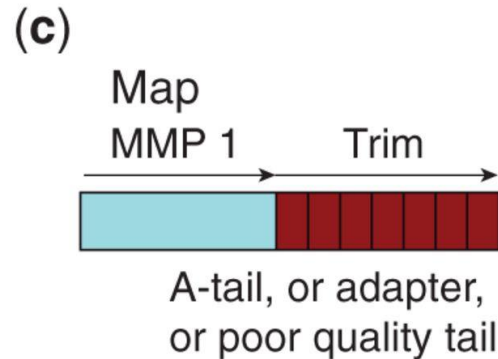
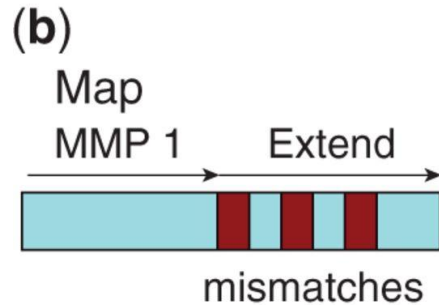
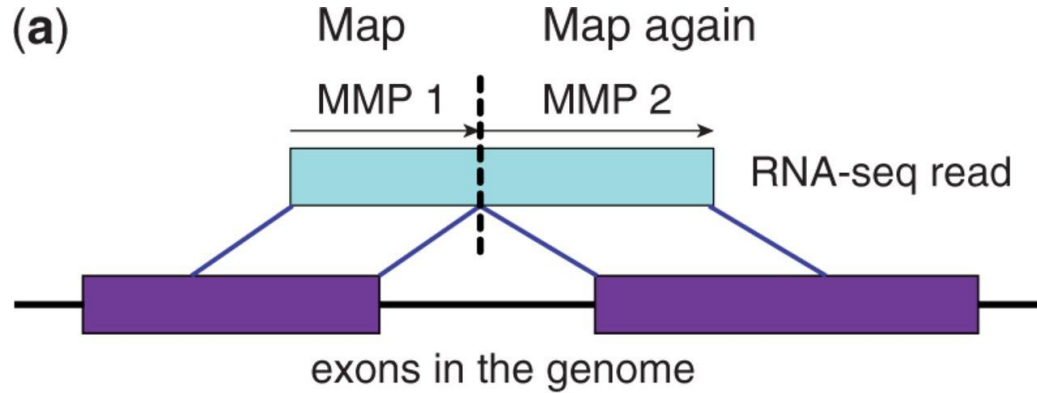


(3-5) Re-alignment of reads minimally overlapping introns



M. Kim & al., *Genome Biology*, 2013

Algo 2 : Star, HISAT



A. Dobin & *al.*,
Bioinformatics, 2013

Plan

- Introduction
- Algorithmes pour l'ADN
- Algorithmes pour l'ARN
- **Quasi-mapping**
- Discussion

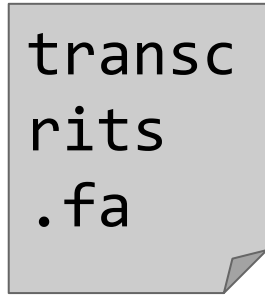
Contexte

- RNA-Seq, quantification de transcrits
- Méthode classique : mapping + quantification
- Et si l'on ne souhaite pas trouver de nouveaux transcrits ?

Question : Peut-on tout faire en 1 étape ?

Sailfish, Kallisto, Salmon

1. Indexer les k-mers des transcrits



transc
rits
.fa



AACGT : gene_32

ACACA : gene_74

ACCGT : gene_21

Plan

- Introduction
- Algorithmes pour l'ADN
- Algorithmes pour l'ARN
- Quasi-mapping
- Discussion

Commentaires

- Qualité du mapping: 0-40
- Lectures pairées: plus sensible/spécifique
- Encodage binaire: A=00, C=01, G=10, T=11

Quel est le meilleur outil ?

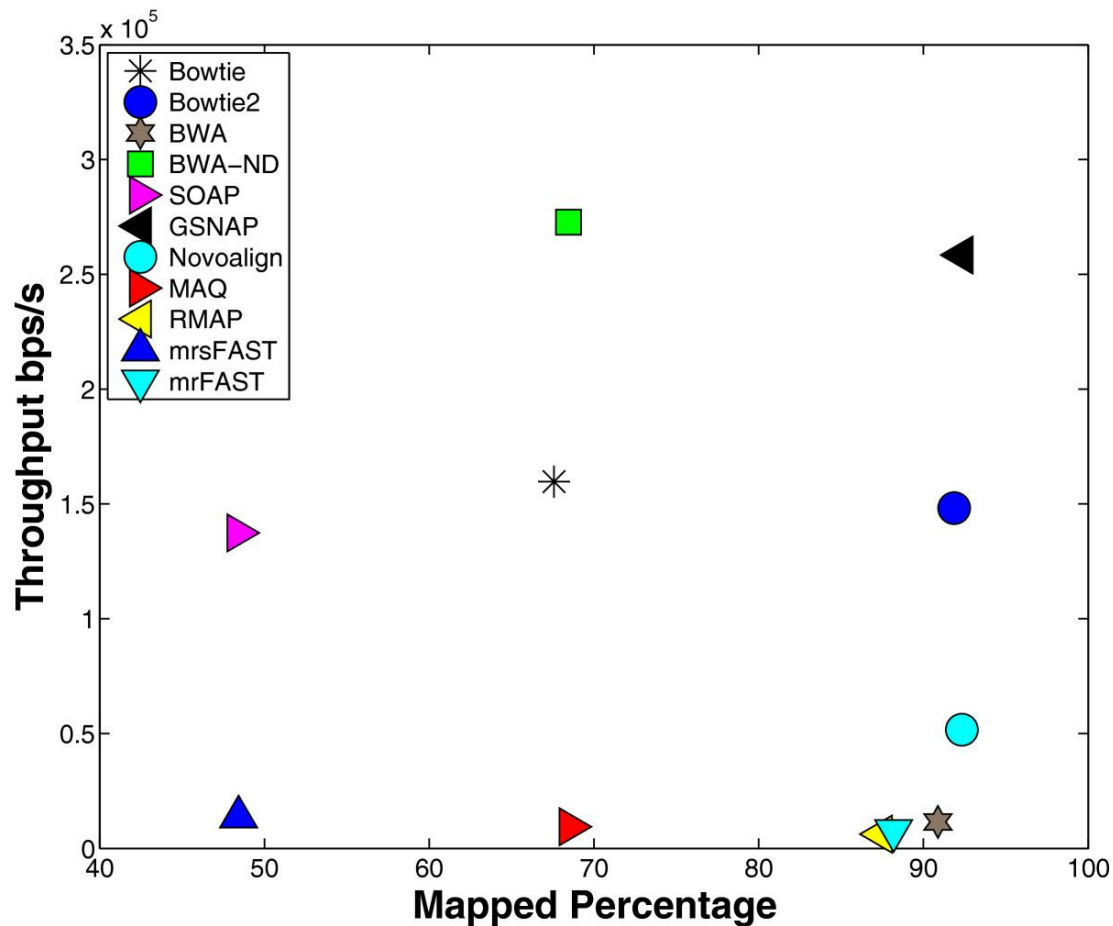
- À chaque séquençage son outil

“ For 70bp or longer Illumina, 454, Ion Torrent and Sanger reads, assembly contigs and BAC sequences, BWA-MEM is usually the preferred algorithm. For short sequences, BWA-backtrack may be better. BWA-SW may have better sensitivity when alignment gaps are frequent. ”

- On ne rejette pas un article qui utilise les outils classiques (Bowtie, BWA, TopHat, STAR)

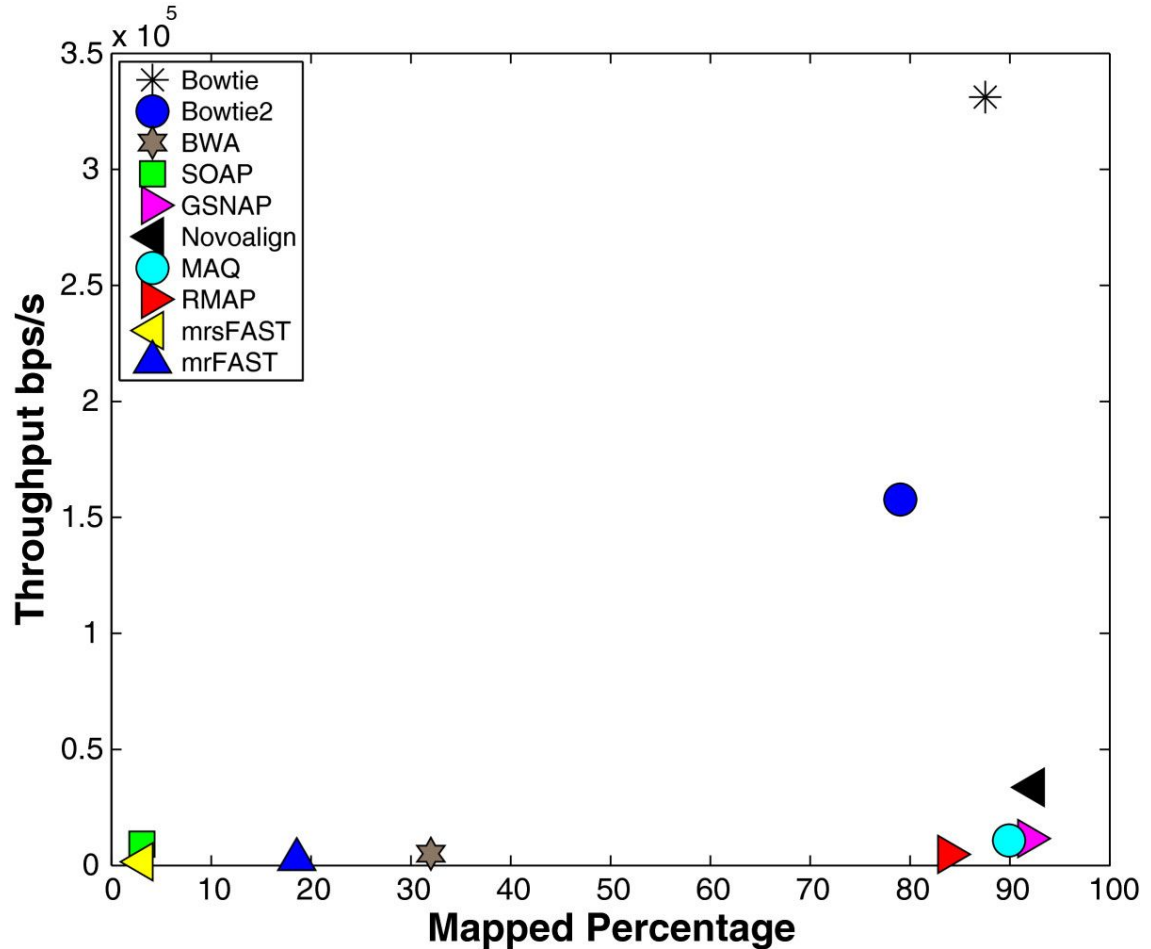
Benchmarks

A Hatem *et al.*,
BMC Bioinf., 2013



Benchmarks

A. Hatem & *al.*,
BMC Bioinf., 2013



Benchmarks

P. Engström &
al., Nat. Meth.,
 2013

