

MAPPI

5 tâches,

- ▶ *Tâche 1 : Nouvelles structures d'index pour la recherche de motifs approchés*
- ▶ *Tâche 2 : Mapping pour la métagenomique et la métatranscriptomique*
- ▶ *Tâche 3 : Outils d'assemblage pour les NGS*
- ▶ *Tâche 4 : Assemblage guidé de données de métatranscriptomique*
- ▶ *Tâche 5 : Pipeline bioinformatique*

MAPPI

5 tâches, celles que je vais décrire dans le contexte Lillois

- ▶ *Tâche 1 : Nouvelles structures d'index pour la recherche de motifs approchés*
- ▶ *Tâche 2 : Mapping pour la métagenomique et la métatranscriptomique*
- ▶ *Tâche 3 : Outils d'assemblage pour les NGS*
- ▶ *Tâche 4 : Assemblage guidé de données de métatranscriptomique*
- ▶ *Tâche 5 : Pipeline bioinformatique*

Tâche 1 : Nouvelles structures d'index pour la recherche de motifs approchés

Contexte : Read Mapping

```
perl5.12  bash  bash  bash  ssh
5441 5451 5461 5471 5481 5491 5501 5511 5521 5531 5541 5551
TGTTCGTGACGTTTAAACACCACTCCGCATTTCAGACGCTTCTTCATCAAGAAATAACCTCCGGATTAAGTCCCTTAACTGTCTGGCTCTGTCAAGCGATTACAATAATAMACGAC
-----T-----
CGCTCGTAACGTTTAAACACCAACCCCGCA*TTGTAGAGGAICTTCAT          CATGTGAC*G*TCGGTTGGTCAGCGGATTTCAATAATGAAACGAC
CGCTCGTAACGTTTAAACACCAACCCCGCA*TTGTAGAGGAICTTCAT          tgaactcgg*tgatttagcagcgatttcaataatgaacgac
tgcctgtoncgttttaaacacccaaccccgcca*ttgttagggaaicttcac      tctcggcctggtca*gcgattttcaataatgaacgac
GCCTGTAAGTTTAAACACCAACCCCGCA*TTGTAGAGGAICTTCATCA         tctgg*ttcgtcagcgatttcaataatgaacgac
tccgtoncgttttaaacacccaaccccgcca*ttgttagggaaicttcacaa     ctggcctggtca*gcgattttcaataatgaacgac
tctonaacgttttaaacacccaaccccgcca*ttgttagggaaicttcacaa     ttgittg*a*agccatttttcaataatgaacgac
CGTAATGTTTAAACACCAACCCCGCA*TTGTAGAGGAICTTCATCAAG        GGTTTGGTCAGCGGATTTCAATAATGAAACGAC
gpcgttttaaacacccaaccccgcca*ttgttagggaaicttcacaagaa       ggtttagcagcgatttttcaataatgaacgac
ACGTTTAAACACCAACCCCGCA*TTGTAGAGGAICTTCATCAAGAAGT        CTGGTCAGCGGATTTCAATAATGAAACGAC
GTTAAACACCAACCCCGCA*TTGTAGAGGAICTTCATCAAGAAGTAA        ttgtagcagcgatttttcaataatgaacgac
ttttaaacacccaaccccgcca*ttgttagggaaicttcacaagaaagtaac    tgttcagcgatttttcaataatgaacgac
ttttaaacacccaaccccgcca*ttgttagggaaicttcacaagaaagtaacc    ctgagcgatttttcaataatgaacgac
caaacacccaaccccgcca*ttgttagggaaicttcacaagaaagtaaccctcgc  TCAGCGGATTAACAATAATGAAACGAC
CCAACCCCGCA*TTGTAGAGGAICTTCATCAAGAAGTAACTCCGGA        agccgattttcaataatgaacgac
aaccccgcca*ttgttagggaaicttcacaagaaagtaaccctcgcatt       GCCGATTTCAATAATGAAACGAC
ACCCCGCA*TTGTAGAGGAICTTCATCAAGAAGTAACTCCGCTAGTA       gtagcacaataatgaacgac
TCGCCA*TTGTAGAGGAICTTCATCAAGAAGTAACTCCGCTACAAT*G     CGATTTCAATAATGAAACGAC
CGCCA*TTGTAGAGGCTACTTCATCAAGAAGTAACTCCGGAATTA       gatcccccaataatgaacgac
gccca*ctgttagggaaicttcacaagaaagtaaccctcgcattaaagt     ctcaataatgaacgac
cca*ttgttagggaaicttcacaagaaagtaaccctcgcattaaagatt     caataatgaacgac
tgtaaac*gtctcttcacaagaaagtaaccctcgcattaaagattgtgt
aggctctcttcacaagaaagtaaccctcgcattaaagattgtgacacct
GCTACTTCATCAAGAAGTAACTCCGCTAGTAAAGT*CACTGAGCTTC
ctcttcacaagaaagtaaccctcgcattaaagattgtgactccgctc
```


Tâche 1 : Nouvelles structures d'index pour la recherche de motifs approchés

Contexte : *Read Mapping*

Réalisé : 1. Portage de l'algorithme de Wu-Mamber sur GPU

[Bit-Parallel Multiple Pattern Matching. T. T. Tran, M. Giraud, J.-S. Varré PPAM / PBC 2011.]

2. Indexation des voisinages des k -mers

But : profiter de l'efficacité du cache GPU/Processeur

Deux méthodes d'indexation envisagées :

- ▶ indexation directe (tri des mots → recherche dichotomique)
- ▶ hachage parfait

+ non encore publié mais des résultats :

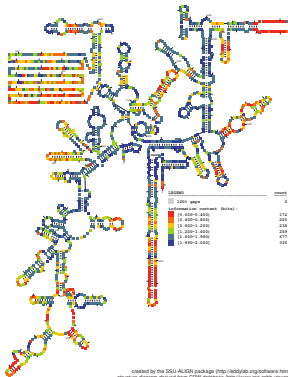
- ▶ mise en oeuvre en OpenCL (fonctionnelle sur CPU et GPU)
- ▶ gain en performance entre x10 et x60
- ▶ prototype de readmapper en cours

Tâche 4 : Assemblage guidé de données de métatranscriptomique

Contexte : identification d'ARN ribosomiques (16S/18S,23S/28S...)

- Buts :
- ▶ élimination
 - ▶ classification

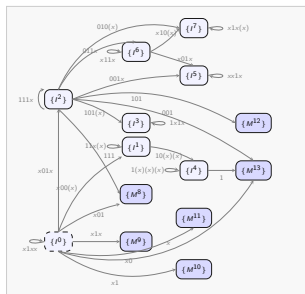
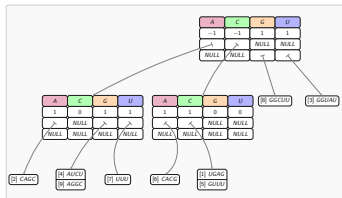
? : Problème nouveau sur données de *métatranscriptomique*



Tâche 4 : Assemblage guidé de données de métatranscriptomique

Contexte : identification d'ARN ribosomiques

- Réalisé :
- ▶ conception d'un filtre efficace pour la sélection des familles d'ARNr (SortMeRNA)
 - ▶ travail basé sur le *Burst Trie* et *l'automate de Levenstein*



Tâche 4 : Assemblage guidé de données de métatranscriptomique

Contexte : identification d'ARN ribosomiques

- En cours :
- ▶ communication aux *London Stringology Days*
 - ▶ publication en cours de soumission
 - ▶ séjour prévue au Génomètre pour la transition
fin Tâche 4 / début Tâche 2, 10-13 avril