

Analyse de génome bactérien : procédures et outils bio-informatiques

Jean-Guillaume E-R

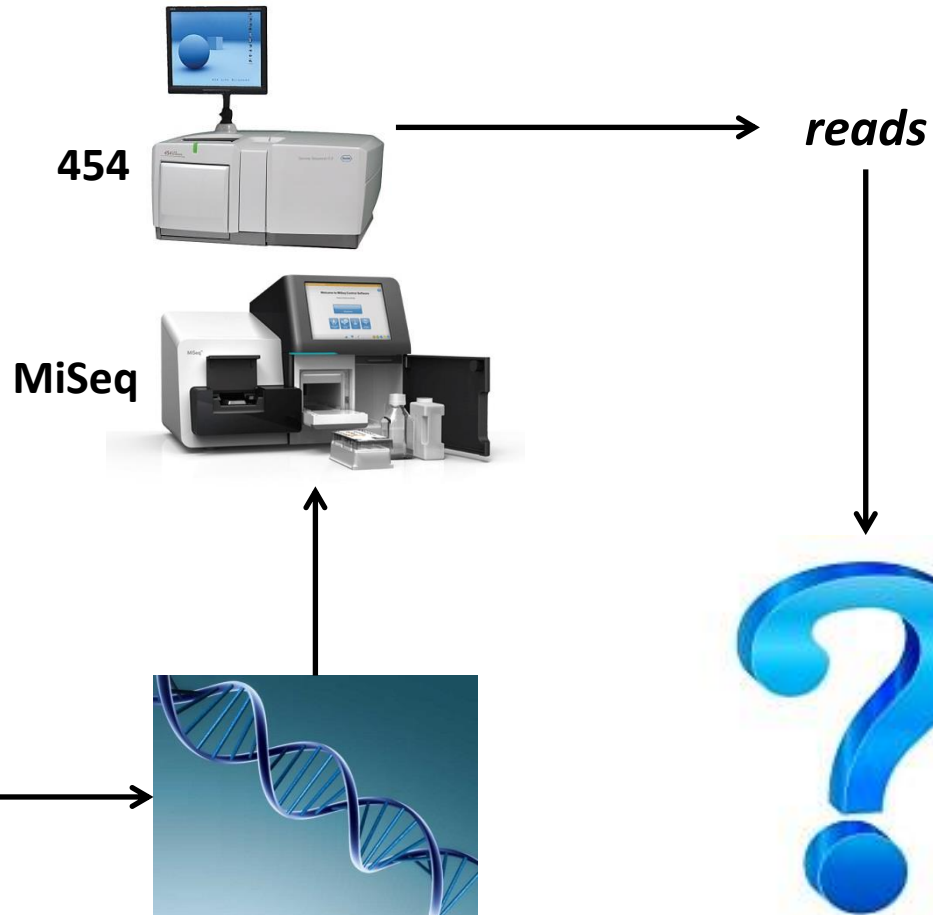
Laboratoire de Steve Charette



Objectifs de cette présentation

- 1) Présenter une procédure d'analyse
- 2) Présentation de différents outils bio-informatiques

De l'infection, jusqu'à l'écran



Technologies de séquençage

Company	Instrument	Average reads length (bp)	Output (Gb)
Roche	454 GS FLX+	700	0,7
Illumina	MiSeq	2 x 300	15
Illumina	HiSeq	2 x 125	1000
Life Technologies	Ion Proton	200	10
Life Technologies	Ion PGM	200-400	0,3-2 ^a
Pacific Biosciences	PacBio RS II ^c	14 000 ^d	0,5-1

Résultats de séquençage

➤ Dans un fichier en **Format.fastq**

- 1) Nom de la lecture (*reads*) et de son descriptif
- 2) Séquence de la lecture en pb
- 3) Répétition du nom de la lecture et de son descriptif
- 4) Score de qualité en code ASCII

Exemple d'un fichier.fastq:

```
1) @SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
2) GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
3) +SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
4) !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!9IG9IC
```

Traitement des données

- Important d'effectuer une filtration des lectures pour:
 - 1) Enlever les séquences adaptatrices ajoutées pour le séquençage
 - 2) Les lectures de mauvaises qualités
- **Trimmomatic** : logiciel de filtration
- Programmes d'**assemblage *de novo*** :
 - Ray, A5 pipeline, Newbler, CLC Genomics Workbench
- Contre vérifier l'assemblage par un deuxième programme

Logiciels libres

Logiciels propriétaires

Assemblage *de novo* des reads

Chromosome

Séquençage

reads

reads

reads

Contigs_1

Contigs_2

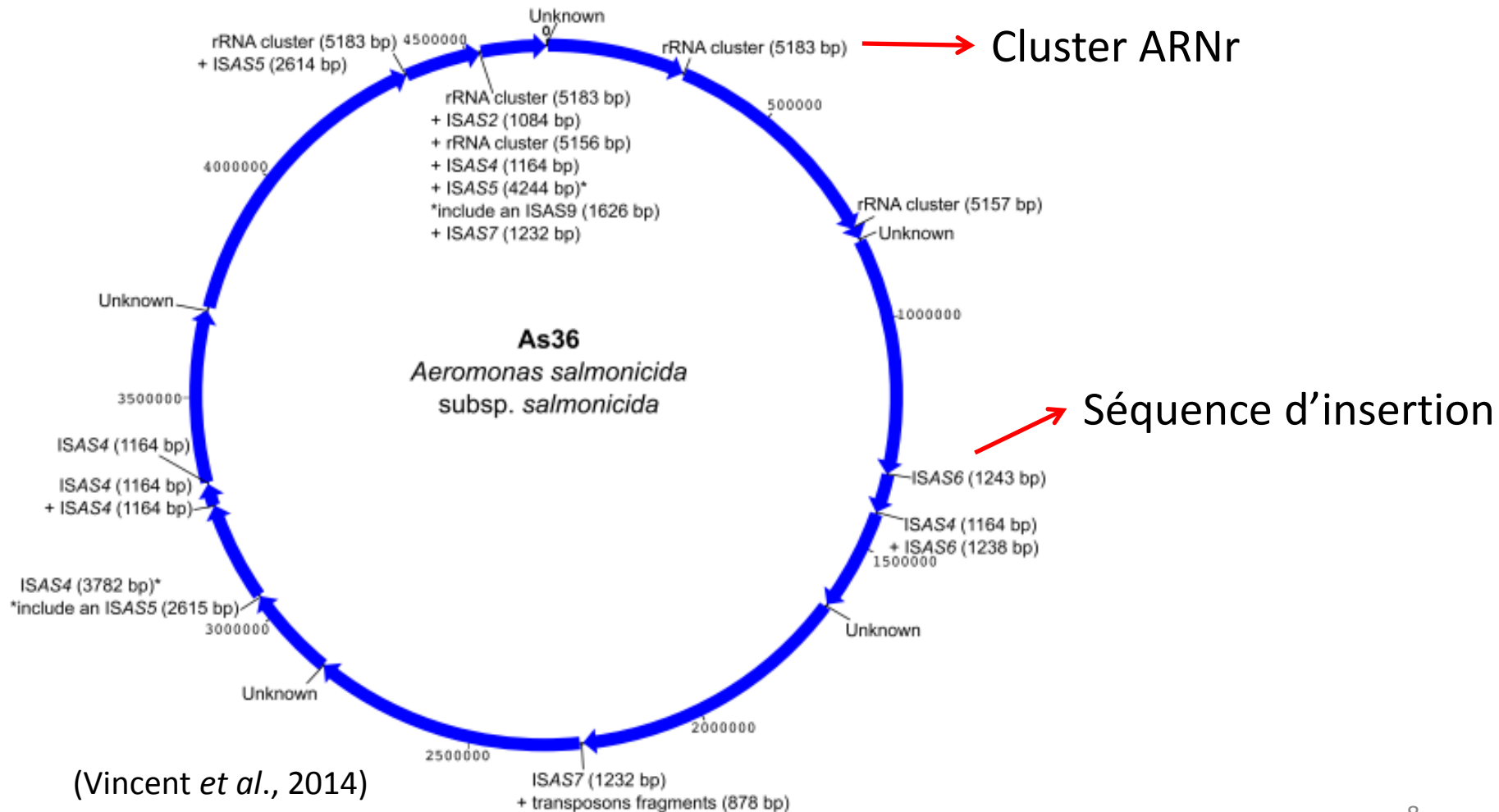
Contigs_3

Contigs_4

Séquence répétée quatre fois dans le chromosome

Bris de contigage

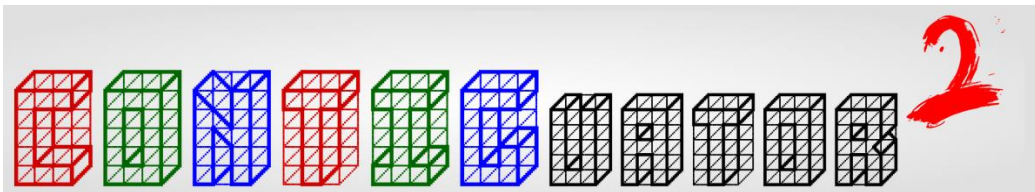
- Causés par les éléments répétés dans les génomes



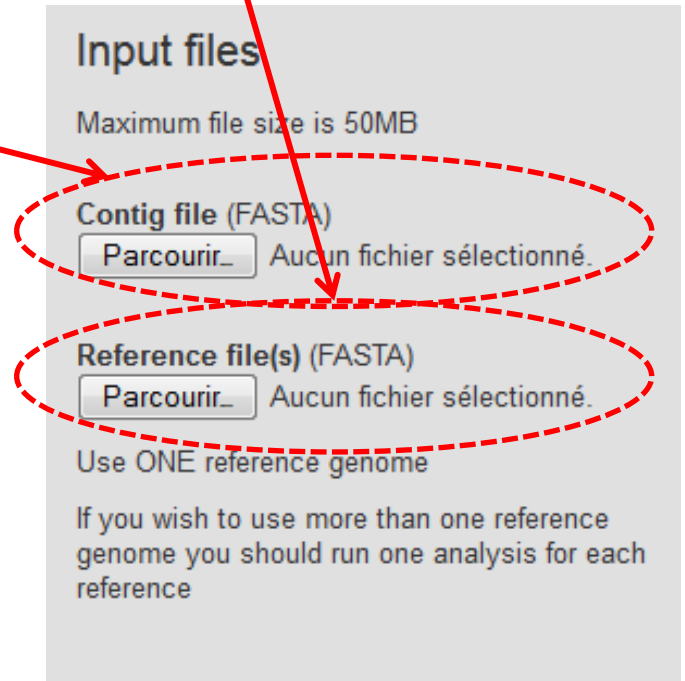
Analyses des résultats d'assemblage

- Connaitre la souche bactérienne de référence
 - Même espèce, même environnement
- Fichier multi-fasta contenant tous les contigs

Utilisation de serveur web **CONTIGuator**



<http://contiguator.sourceforge.net/>

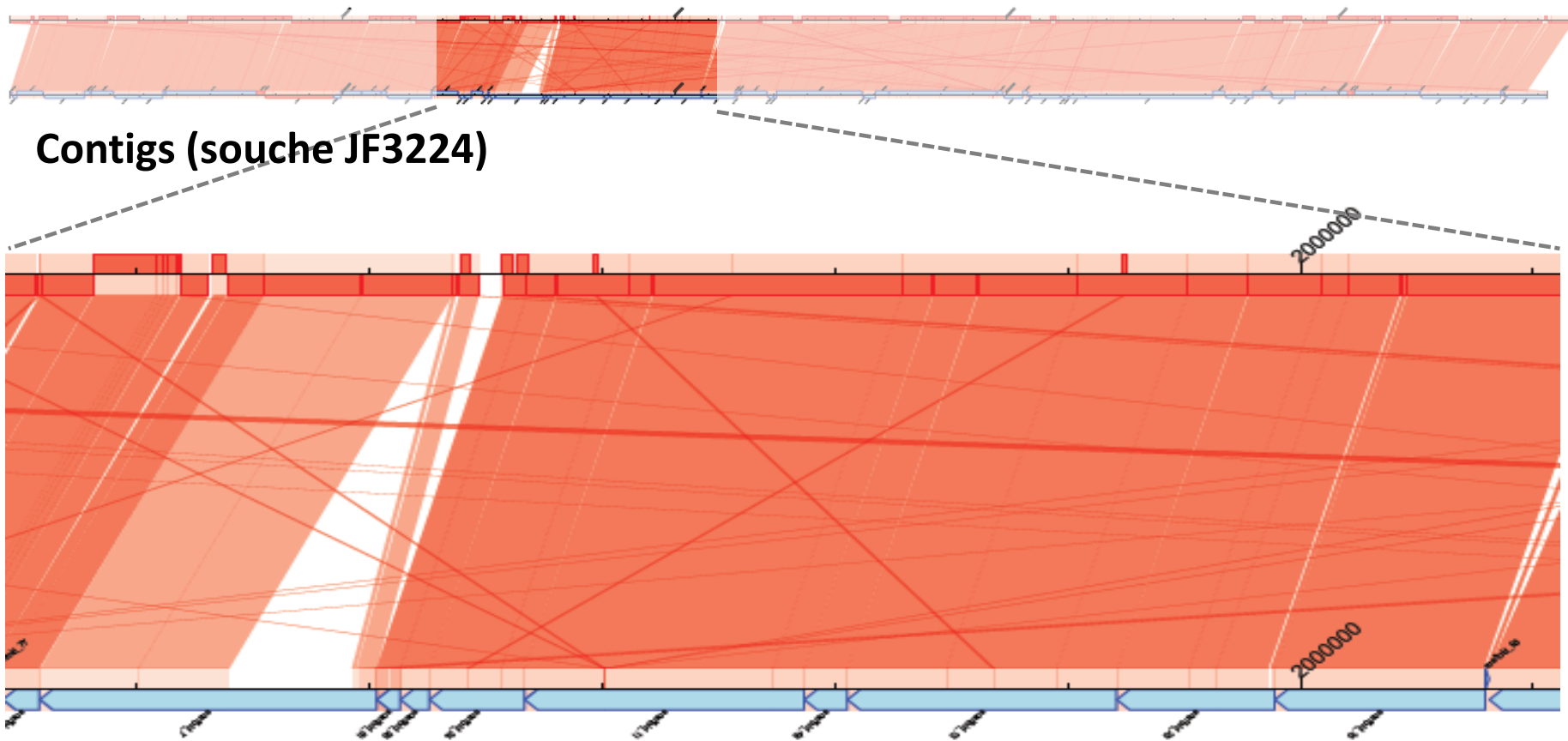


The screenshot shows the web interface of the CONTIGuator server. It has a light gray background. At the top, it says 'Input files' in bold. Below that, it says 'Maximum file size is 50MB'. There are two main input sections, each with a red dashed oval around it. The first section is 'Contig file (FASTA)' with a 'Parcourir...' button and the text 'Aucun fichier sélectionné.' The second section is 'Reference file(s) (FASTA)' with a 'Parcourir...' button and the text 'Aucun fichier sélectionné.' Below these sections, it says 'Use ONE reference genome' and then a paragraph: 'If you wish to use more than one reference genome you should run one analysis for each reference'.

Alignement CONTIGuator

Chromosome de la référence (souche A449)

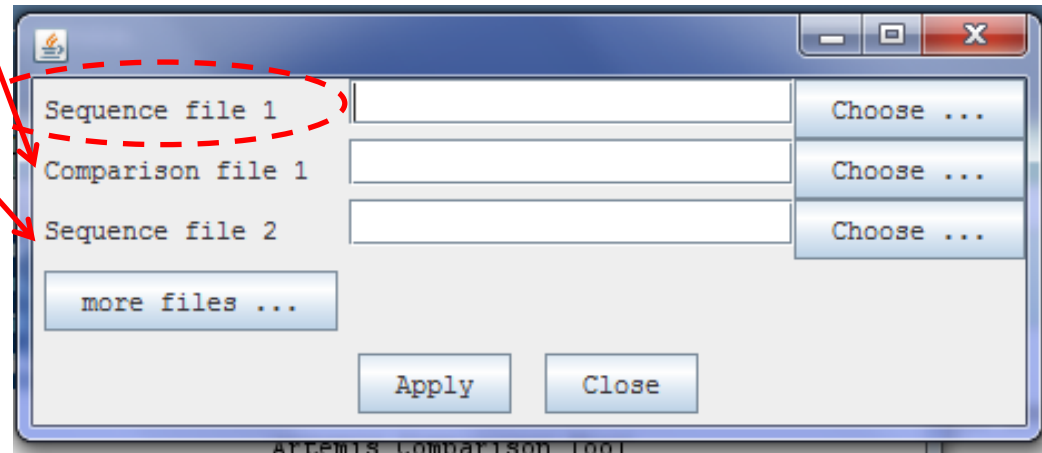
Contigs (souche JF3224)



Output de CONTIGuator

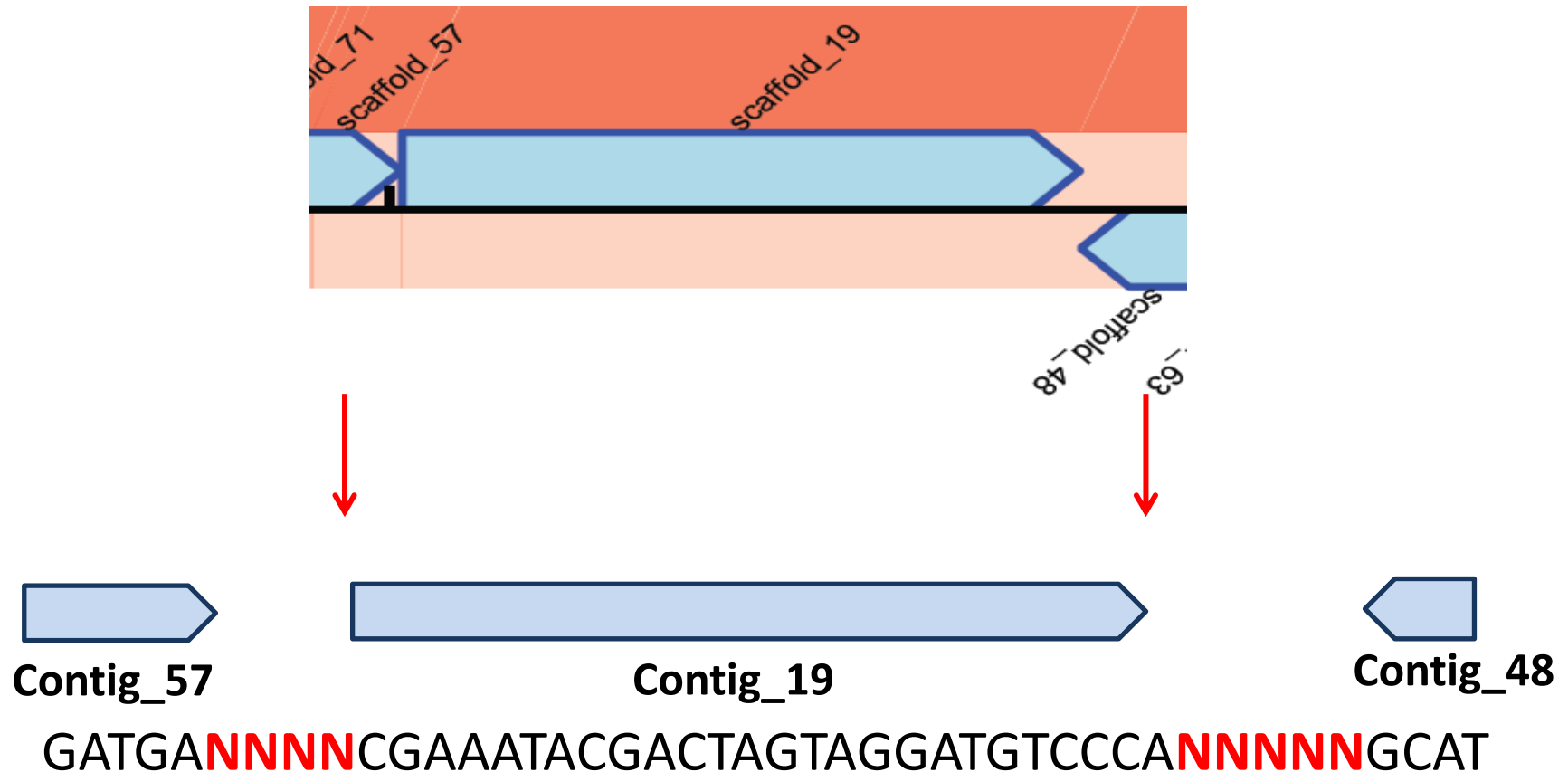
- Représentation de l'alignement en format.pdf
- Fichiers nécessaire pour visualiser l'alignement dans ACT
 - PseudoContigs.CRUNCH
 - PseudoContigs.FSA

Chromosome de la référence
(souche A449)



PseudoContigs.FSA

Synonyme: PseudoChromosome.FSA

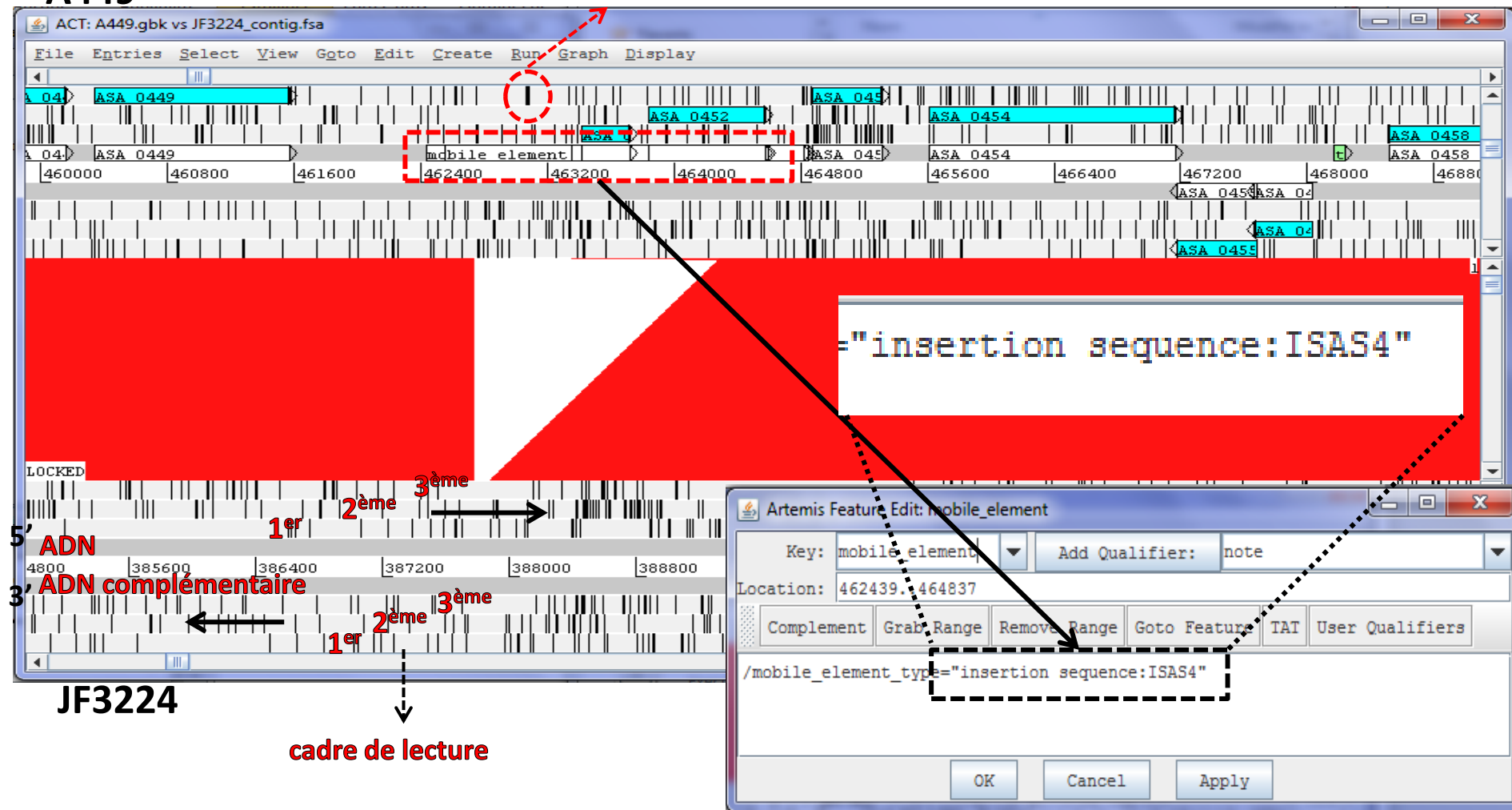


ACT

- Logiciel interactif pour visualiser un alignement

A449

Codon stop



Analyses des contigs *unmapped*

- Vérification de la nature des contigs *unmapped*
 - Peuvent être des plasmides, des phages, etc
- BLASTn des contigs *unmapped* sur le serveur du NCBI

The screenshot shows the NCBI BLASTn suite interface. At the top, there's a blue header with the BLAST logo and the text "Basic Local Alignment Search Tool". Below this is a navigation bar with links: Home, Recent Results, Saved Strategies, and Help. The main content area is titled "Standard Nucleotide BLAST" and includes a sub-header "NCBI/ BLAST/ blastn suite". There are tabs for different BLAST programs: blastn (selected), blastp, blastx, tblastn, and tblastx. The "Enter Query Sequence" section contains a large text input field for "Enter accession number(s), gi(s), or FASTA sequence(s)". To the right of this field is a "Clear" link and a "Query subrange" section with "From" and "To" input fields. Below the main input field, there's a section for "Or, upload file" with a "Parcourir..." button and the text "Aucun fichier sélectionné."

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

Analyse du Pseudo Chromosome

Pseudo Chromosome de JF3224 généré par CONTIGuator



Moteur d'annotation automatique

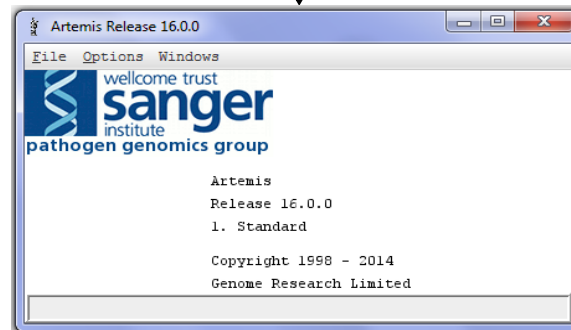
<http://rast.nmpdr.org/rast.cgi>



Format GenBank



<http://www.pathogenomics.sfu.ca/islandviewer/query.php>



[Home](#) [Documentation](#) [Input](#) [Output](#) [Instructions](#) [Databases](#) [Contact](#)

<http://phast.wishartlab.com/>



- Web serveur d'annotation de génome bactérien
- Identifie les ARNt, les ARNr et les séquences codantes (CDS)
 - **tRNAscan-SE, search_for_rnas, Glimmer 3**
- Annotation des CDS d'après les bases de données
 - **FIGfams**
- Propose plusieurs formats de *Output* (multi-fasta, GenBank, Excel, etc)



RAST Rapid Annotation using Subsystem Technology version 2.0

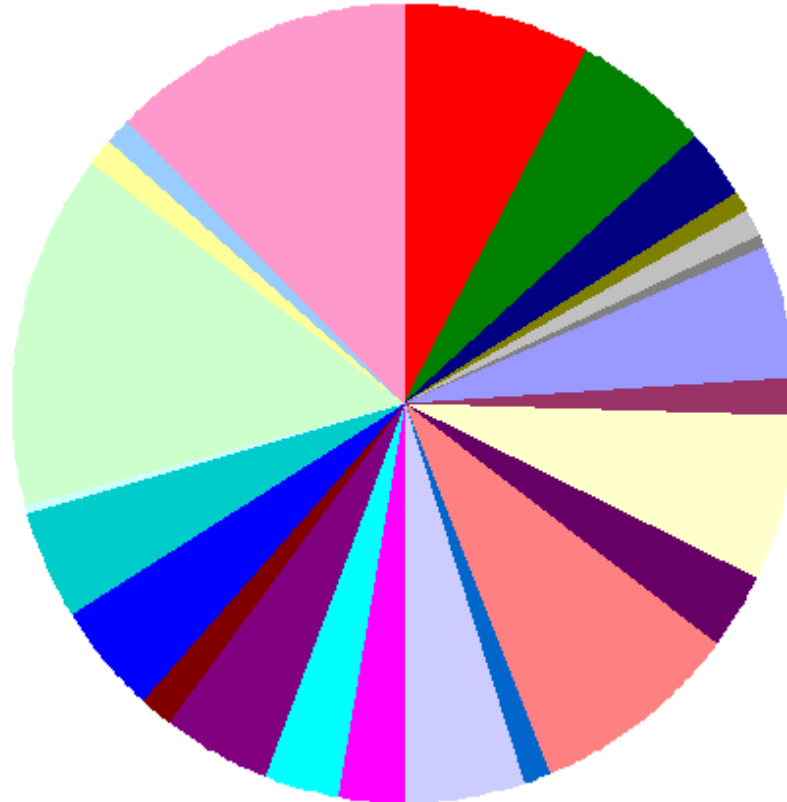
The NMPDR, SEED-based, prokaryotic genome annotation service.
For more information about The SEED please visit theSEED.org.

➤ Détermine les sous-systèmes

Subsystem Coverage



Subsystem Category Distribution



Subsystem Feature Counts

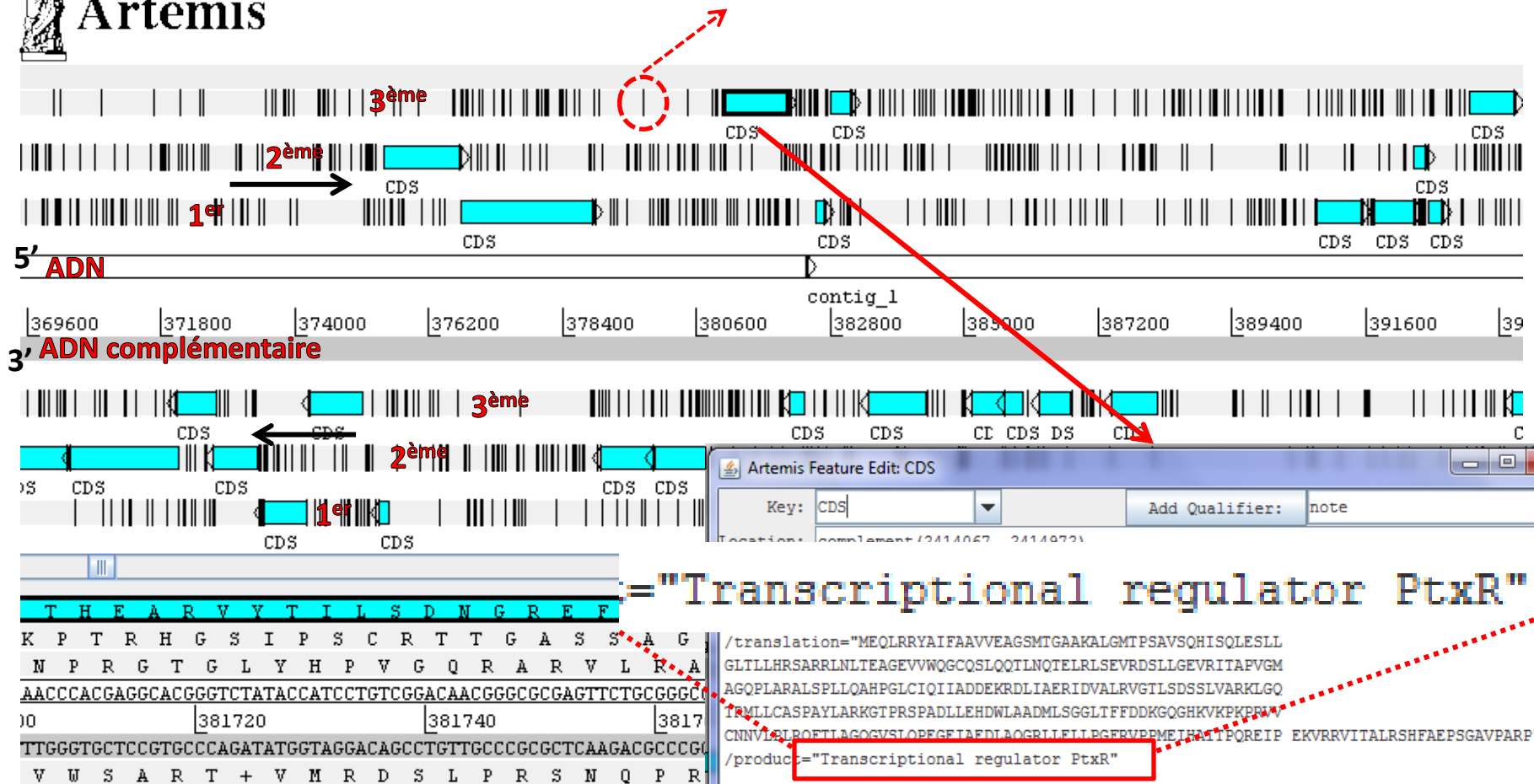
⊕	Cofactors, Vitamins, Prosthetic Groups, Pigments (288)
⊕	Cell Wall and Capsule (195)
⊕	Virulence, Disease and Defense (94)
⊕	Potassium metabolism (39)
⊕	Photosynthesis (0)
⊕	Miscellaneous (36)
⊕	Phages, Prophages, Transposable elements, Plasmids (18)
⊕	Membrane Transport (193)
⊕	Iron acquisition and metabolism (52)
⊕	RNA Metabolism (245)
⊕	Nucleosides and Nucleotides (107)
⊕	Protein Metabolism (310)
⊕	Cell Division and Cell Cycle (41)
⊕	Motility and Chemotaxis (180)
⊕	Regulation and Cell signaling (98)
⊕	Secondary Metabolism (6)
⊕	DNA Metabolism (111)
⊕	Fatty Acids, Lipids, and Isoprenoids (151)
⊕	Nitrogen Metabolism (51)
⊕	Dormancy and Sporulation (2)
⊕	Respiration (159)
⊕	Stress Response (163)
⊕	Metabolism of Aromatic Compounds (12)
⊕	Amino Acids and Derivatives (513)
⊕	Sulfur Metabolism (38)
⊕	Phosphorus Metabolism (39)
⊕	Carbohydrates (439)

Visualisation de l'annotation du Pseudo Chromosome de JF3224

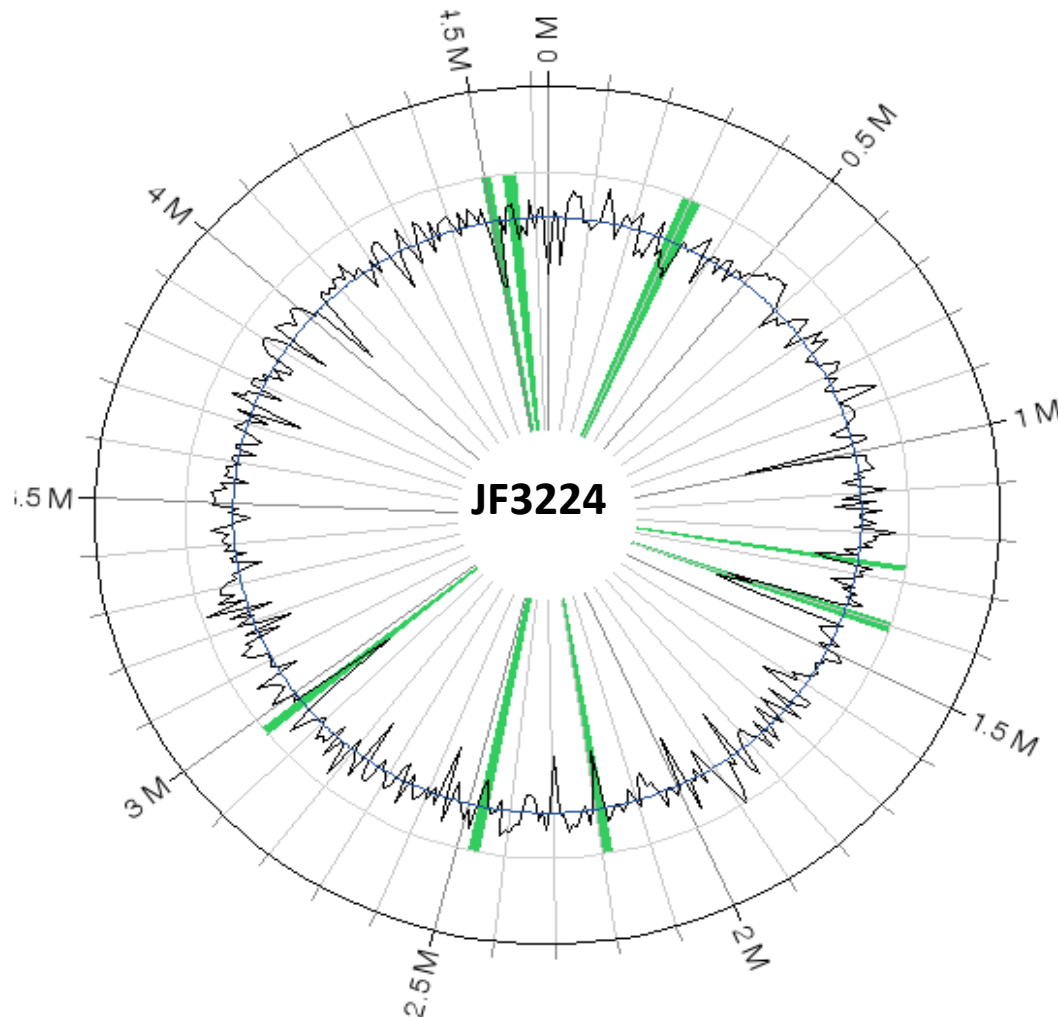


Artemis

Codon stop



➤ Détermination de transfert horizontale



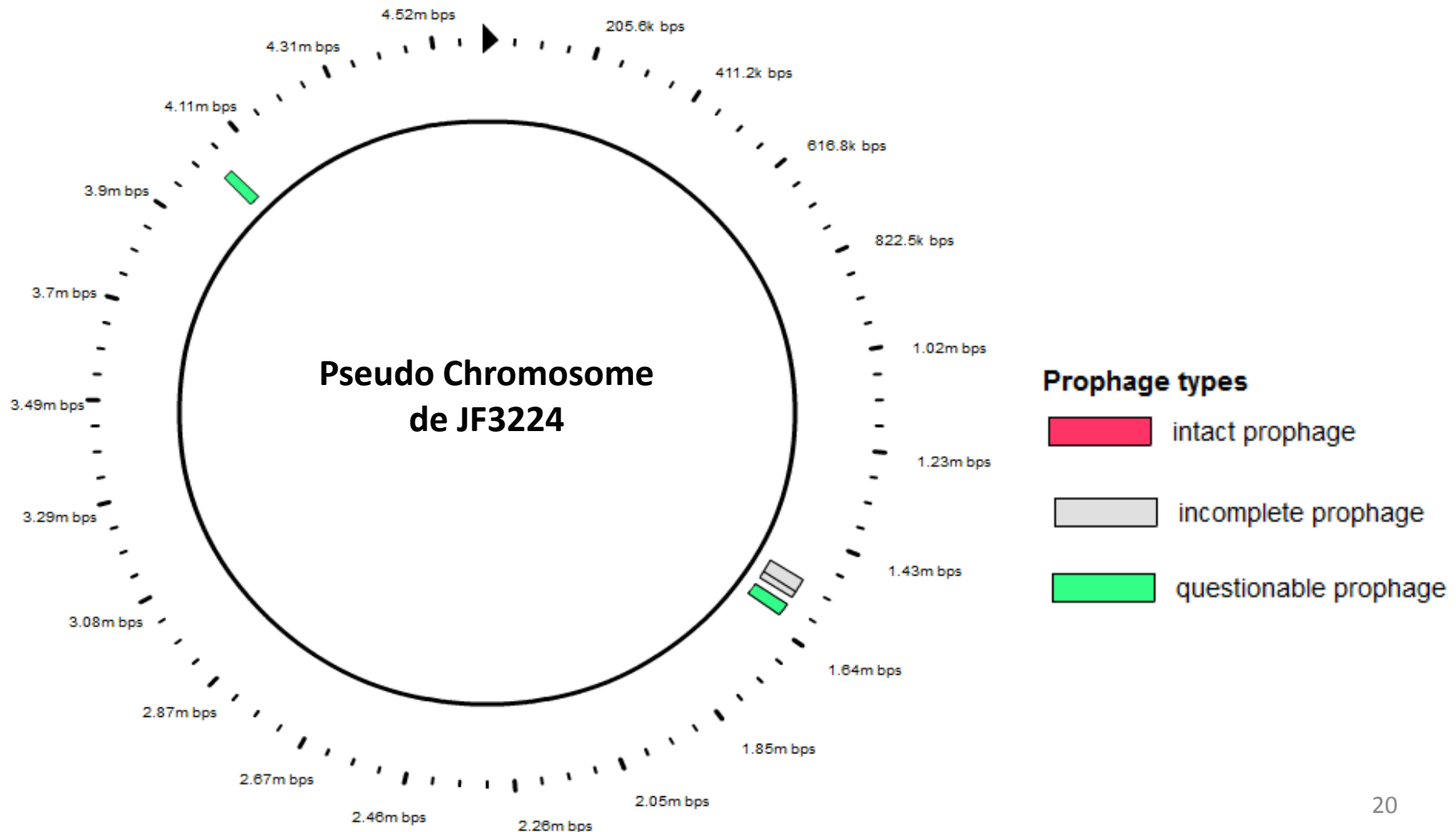
Legend (Help)

- ☐ Integrated
- ☒ IslandPick
- ☐ SIGI-HMM
- ☐ IslandPath-DIMOB



[Home](#)
[Documentation](#)
[Input](#)
[Output](#)
[Instructions](#)
[Databases](#)
[Contact](#)

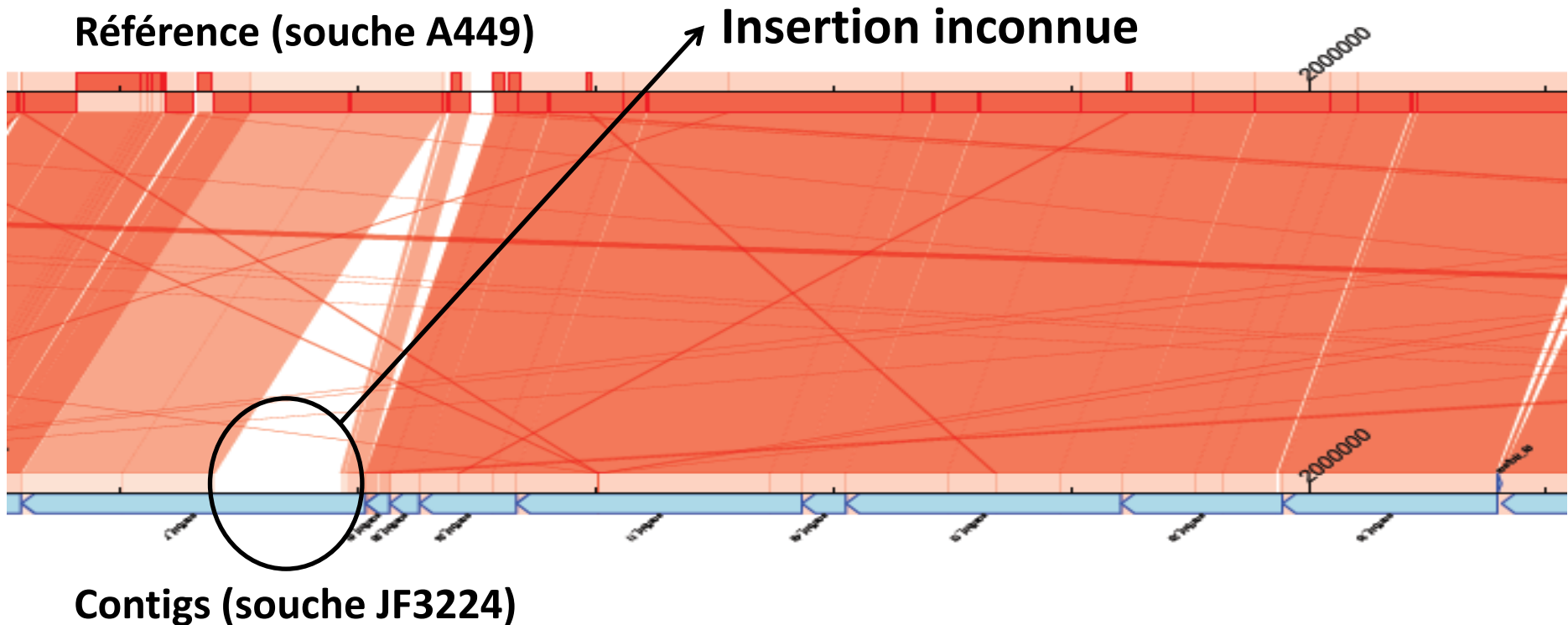
➤ Web serveur permettant l'identifier de prophages



Limites de l'annotation RAST

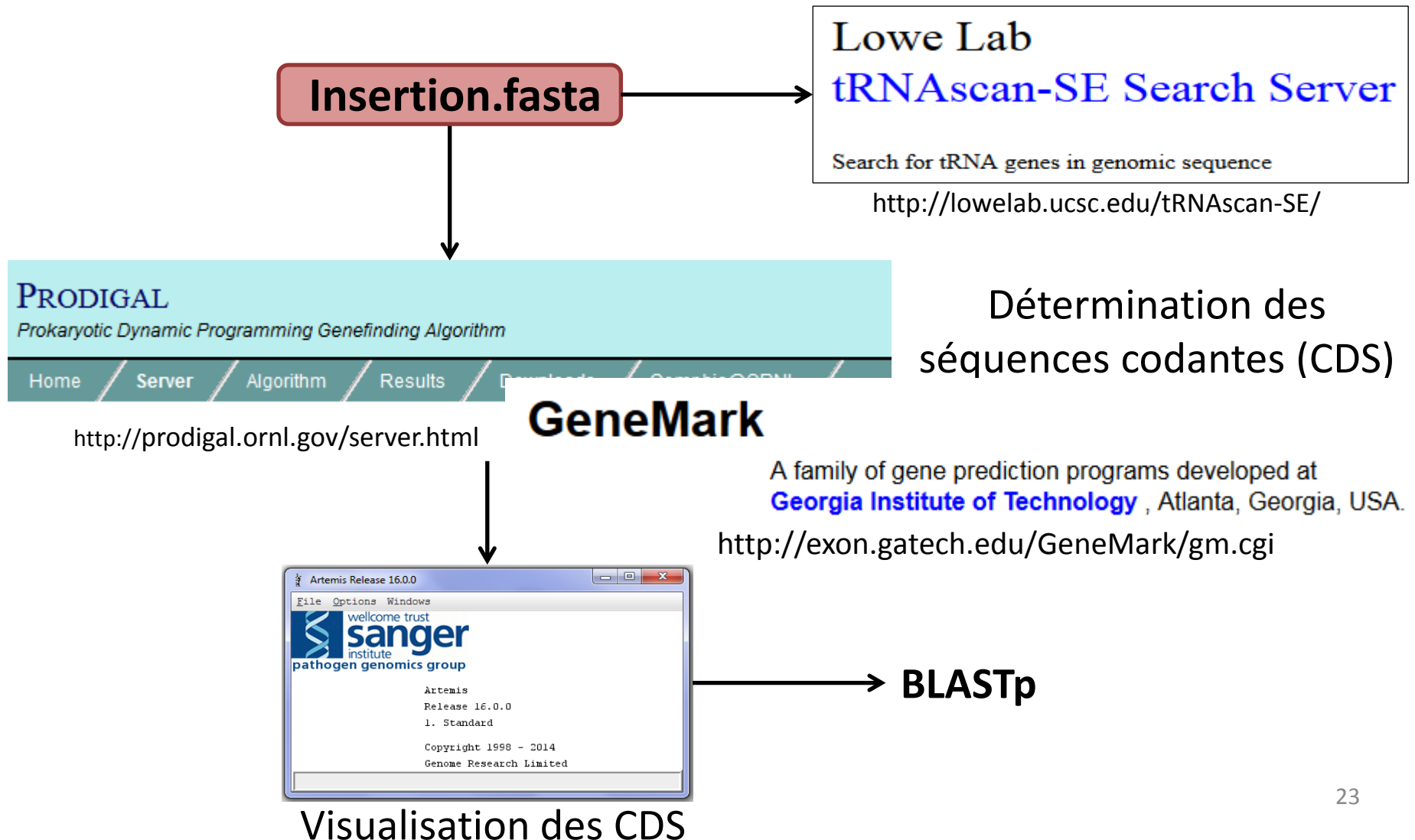
- Limité par les connaissances du moment présent
- Bases de données sont en constante évolution

Annotation d'une séquence inconnue



- Extraction seulement de l'insertion inconnue à partir de l'alignement dans le logiciel interactif ACT

Contre-vérification de l'annotation



Résumé du Pipeline d'analyse



Reads

Filtration et Assemblage en contigs

Alignement sur la référence avec CONTIGuator

Contigs mapped

Contigs unmapped

RAST

BLASTn des contigs
pour déterminer
leurs natures

PHAST

IslandViewer



Des Questions

