



MODULE 3. GENOMIC EPIDEMIOLOGY

Visualization & genomes

June 2024 – Francois Lebreton, PhD

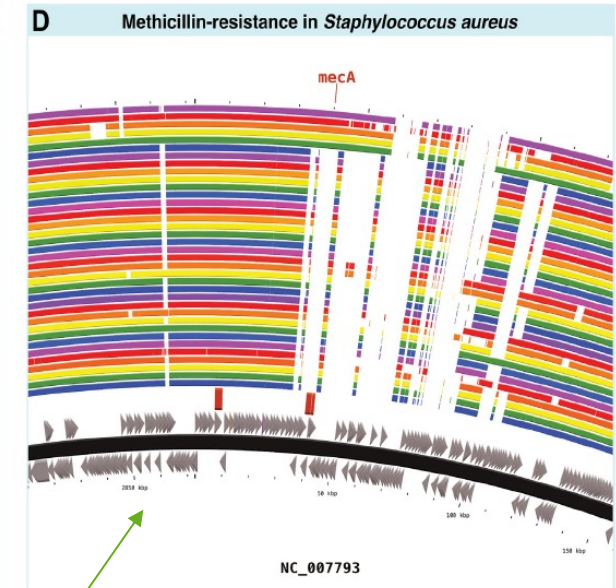
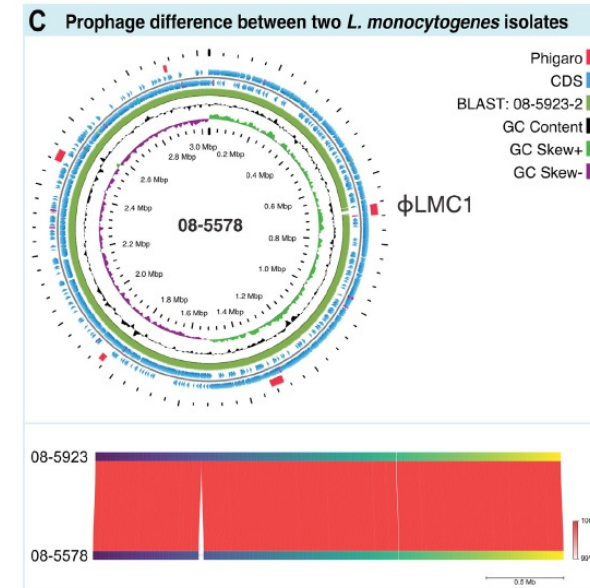
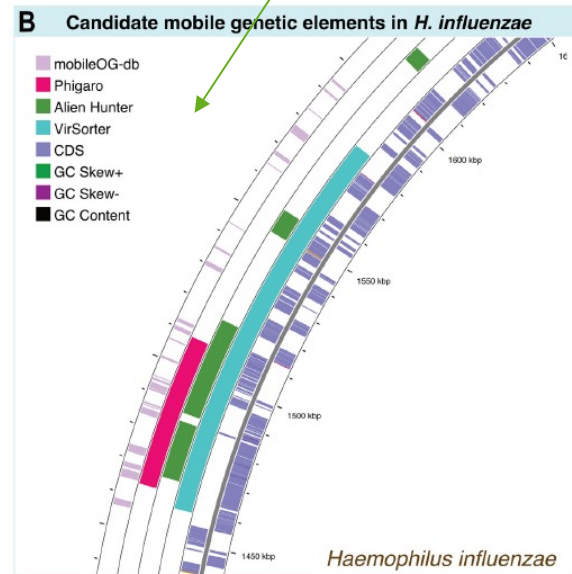
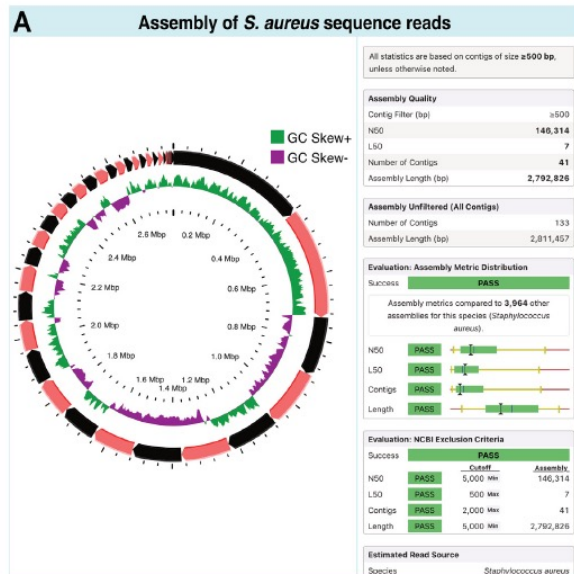
Intended Learning Objectives

Specific objectives of this session:

1. Learn about genomic comparison tools (such as Mauve, BRIG, Proksee..)
2. Limits and advantages
3. Develop ability to critically analyze publications using similar tools
4. Develop ability to use key functionalities of online tools (hands-on)

Intended Learning Objectives (in short)

Understanding this

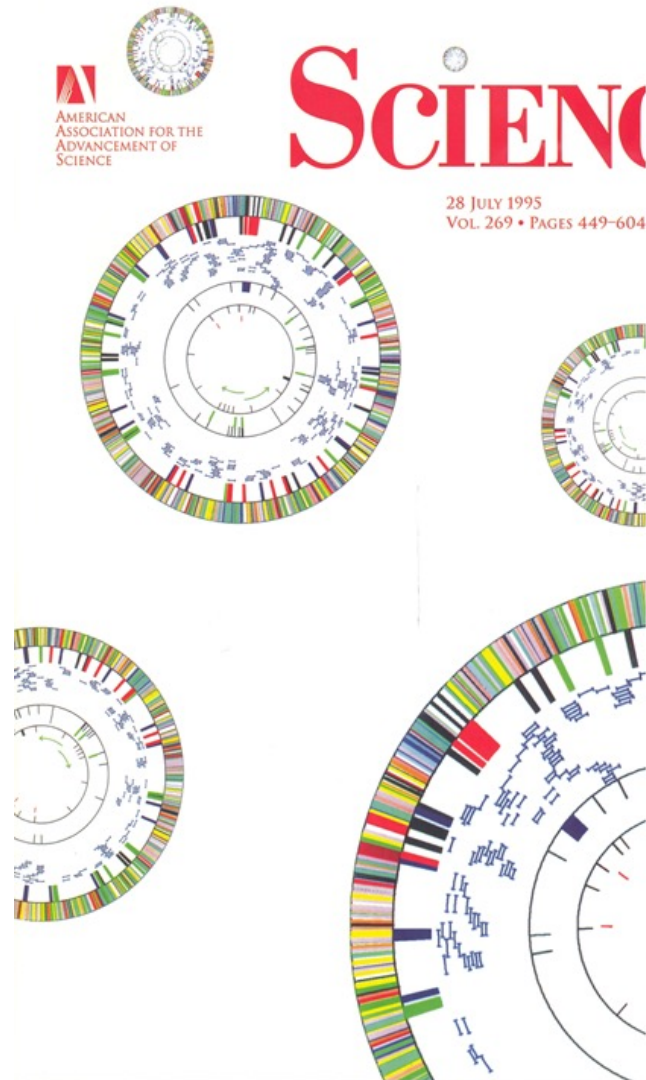


Being able to make that

Outline

- Introduction
- Comparison with Similar Tools
- Advantages/Limitations of Proksee
- Case study #1 – E. coli O104:H4 outbreak
- Application – Key Features & Functionalities
- Case study #2 – Convergent K. pneumoniae
- Application – Investigate plasmid content in outbreak (hands-on)
- Conclusion

Introduction



Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Saudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghagen, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Venter†

An approach for genome analysis based on sequencing and assembly of unselected pieces of DNA from the whole chromosome has been applied to obtain the complete nucleotide sequence (1,830,137 base pairs) of the genome from the bacterium *Haemophilus influenzae* Rd. This approach eliminates the need for initial mapping efforts and is therefore applicable to the vast array of microbial species for which genome maps are unavailable. The *H. influenzae* Rd genome sequence (Genome Sequence DataBase accession number L42023) represents the only complete genome sequence from a free-living organism.

- ## Genome Browsers:

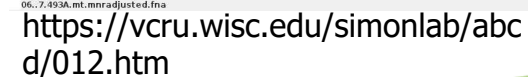
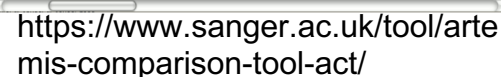
- ## Comparative Genomics Tools:

- ## Advances in Interactive Visualization:

- **2013:** **BRIG** (BLAST Ring Image Generator) introduced circular genome visualizations

Modern Tools:

- **2023: Proksee**, offering comprehensive online platforms for genome annotation, visualization, and comparative analysis with GUI



A catalogue of circular genome viz tools



Alikhan et al. BMC Genomics 2011, 12:402
<http://www.biomedcentral.com/1471-2164/12/402>



SOFTWARE

Open Access

BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons

Nabil-Fareed Alikhan, Nicola K Petty, Nouri L Ben Zakour and Scott A Beatson*



Circular genome visualization and exploration using CGView

Paul Stothard and David S. Wishart*

Department of Biological Sciences and Computing Science, University of Alberta, Edmonton, Alberta, Canada T6G 2E8

Received on June 7, 2004; revised on August 31, 2004; accepted on September 16, 2004
Advance Access publication October 12, 2004

Resource

Circos: An information aesthetic for comparative genomics

Martin Krzywinski,^{1,3} Jacqueline Schein,¹ Inanç Birol,¹ Joseph Connors,² Randy Gascoyne,² Doug Horsman,² Steven J. Jones,¹ and Marco A. Marra¹

¹Canada's Michael Smith Genome Sciences Center, Vancouver, British Columbia V5Z 4S6, Canada; ²British Columbia Cancer Research Center, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada

Nucleic Acids Research, 2023 1
<https://doi.org/10.1093/nar/gkad326>

Proksee: in-depth characterization and visualization of bacterial genomes

Jason R. Grant^{1,†}, Eric Enns^{2,†}, Eric Marinier², Arnab Mandal³, Emily K. Herman¹, Chih-yu Chen^{2,4}, Morag Graham^{2,3}, Gary Van Domselaar^{2,3,*} and Paul Stothard^{1,*}

¹Agriculture, Food & Nutritional Science, University of Alberta, Edmonton, Alberta T6G 2P5, Canada, ²National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba R3E 3R2, Canada, ³Medical Microbiology & Infectious Diseases, University of Manitoba, Winnipeg, Manitoba R3E 0J9, Canada and ⁴Biochemistry & Medical Genetics, University of Manitoba, Winnipeg, Manitoba R3E 0J9, Canada

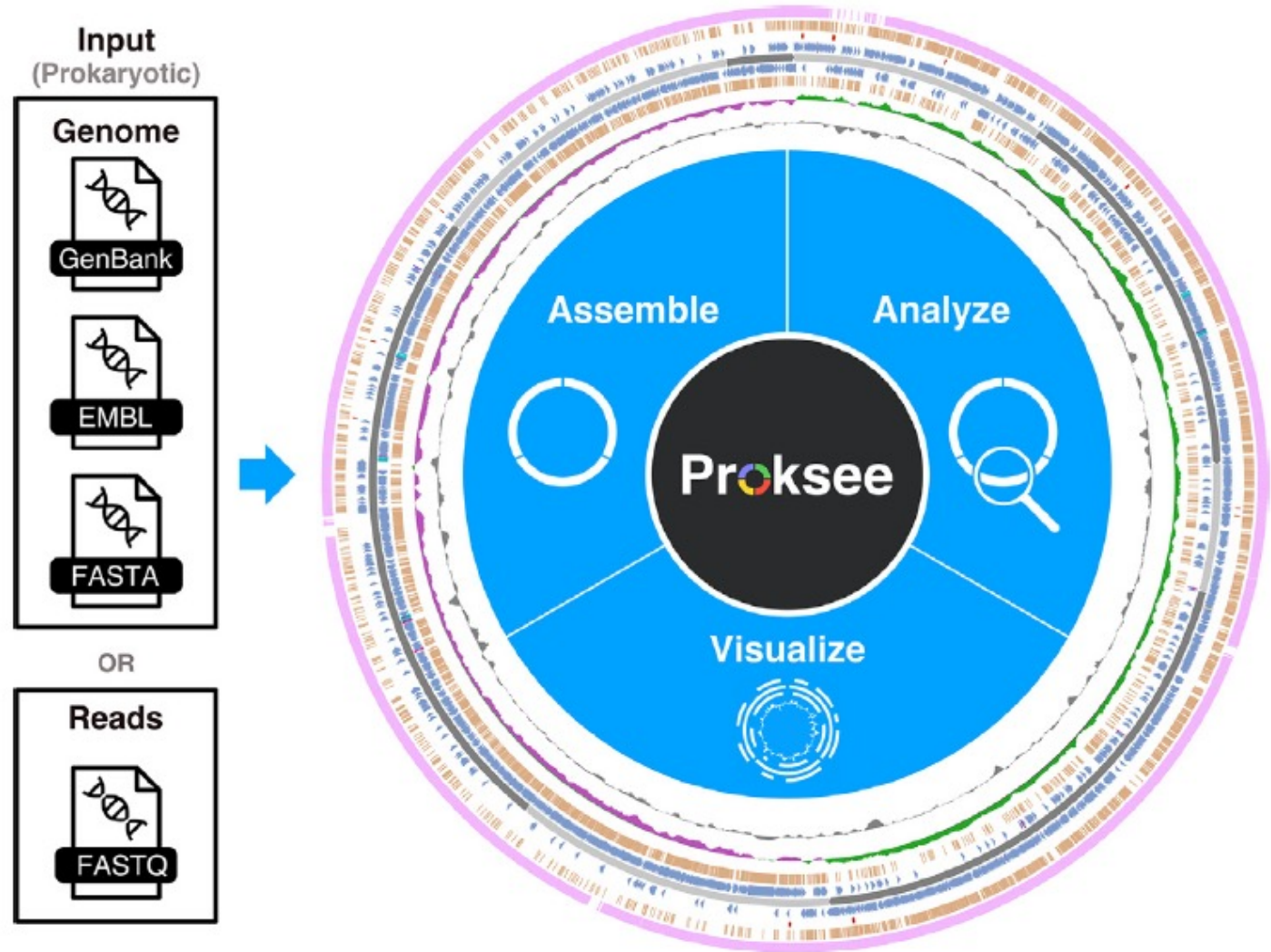
<https://proksee.ca/>

- **What is Proksee?**

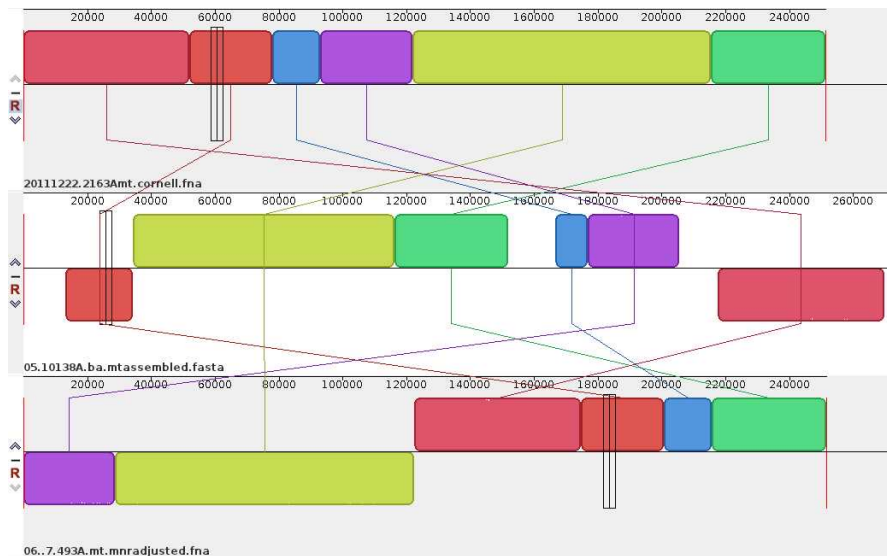
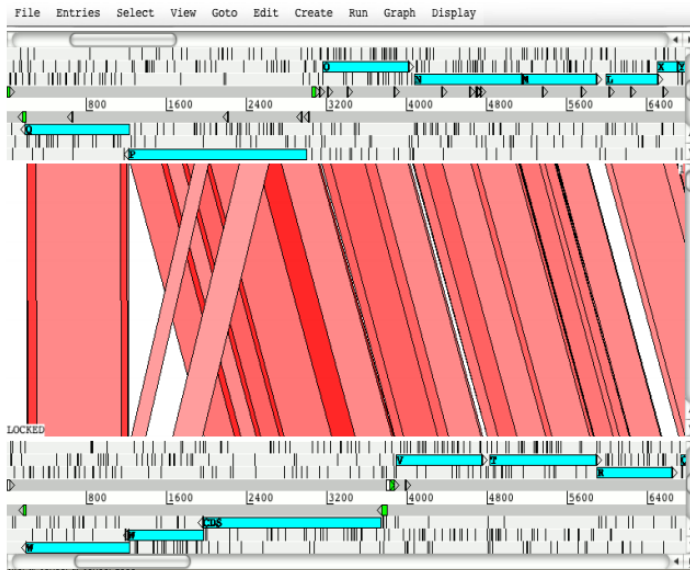
- Proksee is an online tool designed for the annotation, visualization, and comparative analysis of prokaryotic genomes.

- **Purpose and Target Audience**

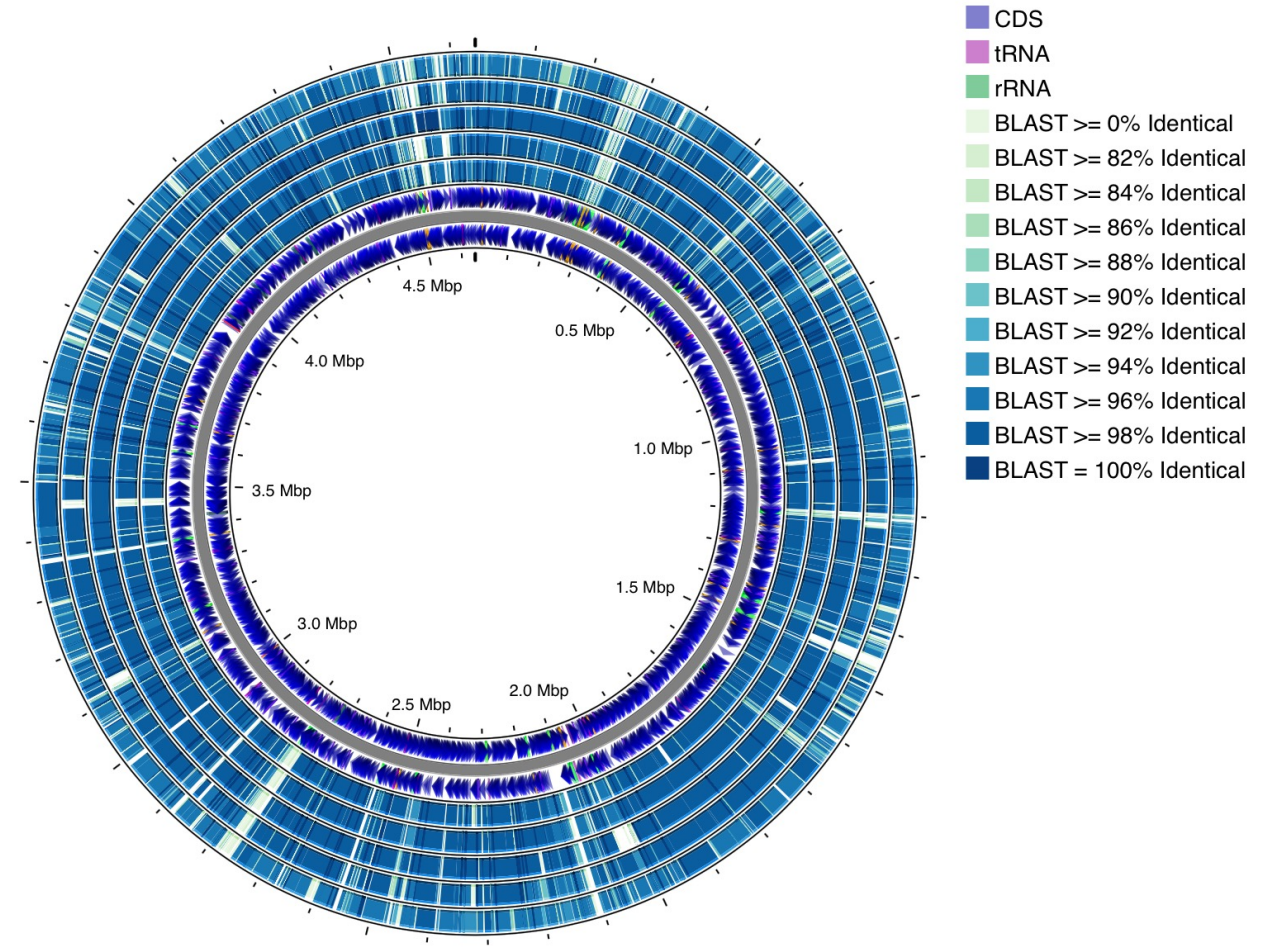
- Aimed at microbiologists, bioinformaticians, and researchers in genomic epidemiology.



What's the difference?



VS



Differences Between Mauve and Proksee

Functionalities:

- Mauve:** Aligns multiple genomes, identifies conserved regions, **deletions/insertions** and rearrangements, phylogenetic analysis.
- Proksee:** Automated genome annotation, comparative genomics, interactive visualization tools.

Type of Analyses:

- Mauve: Genomic rearrangements** (inversions, translocations), conserved sequence blocks.
- Proksee:** Functional annotation, **reference-based** comparative functional genomics, integrative analysis.

User Interface and Accessibility:

- Mauve:** Desktop application, requires local installation, steeper learning curve.
- Proksee:** Web-based platform, user-friendly interface, accessible without installation.

Case study

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

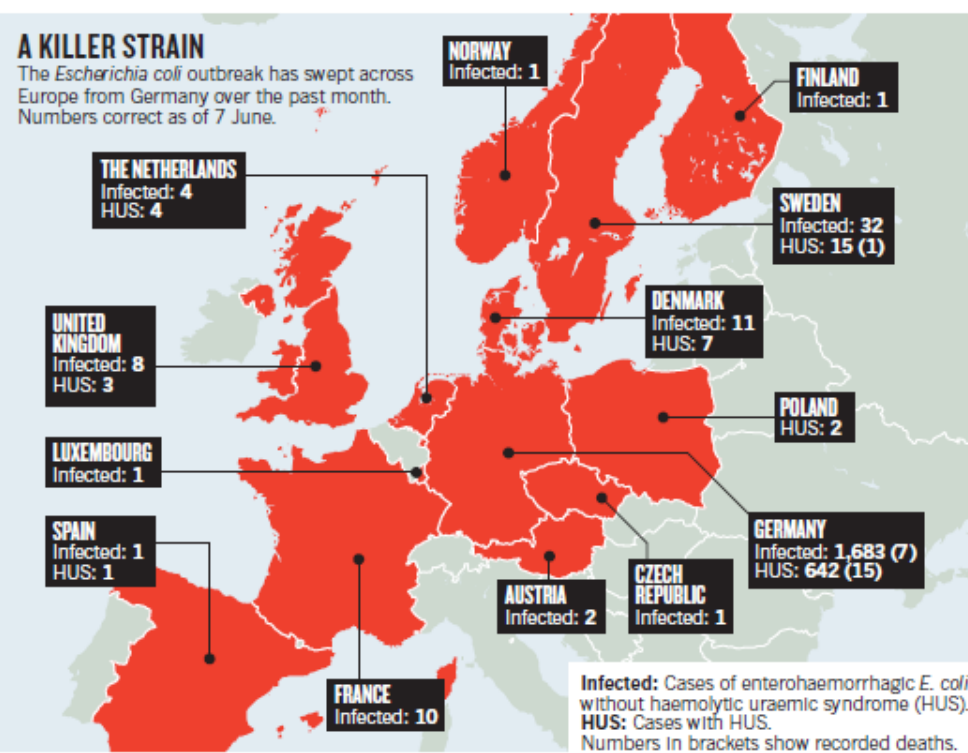
Origins of the *E. coli* Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany

David A. Rasko, Ph.D., Dale R. Webster, Ph.D., Jason W. Sahl, Ph.D.,
Ali Bashir, Ph.D., Nadia Boisen, Ph.D., Flemming Scheutz, Ph.D.,
Ellen E. Paxinos, Ph.D., Robert Sebra, Ph.D., Chen-Shan Chin, Ph.D.,
Dimitris Iliopoulos, Ph.D., Aaron Klammer, Ph.D., Paul Peluso, Ph.D.,
Lawrence Lee, Ph.D., Andrey O. Kislyuk, Ph.D., James Bullard, Ph.D.,
Andrew Kasarskis, Ph.D., Susanna Wang, B.S., John Eid, Ph.D.,
David Rank, Ph.D., Julia C. Redman, B.S., Susan R. Steyert, Ph.D.,
Jakob Frimodt-Møller, M.Sc.Eng., Carsten Struve, Ph.D., Andreas M. Petersen, Ph.D.,
Karen A. Krogfelt, Ph.D., James P. Nataro, M.D., Ph.D., M.B.A.,
Eric E. Schadt, Ph.D., and Matthew K. Waldor, M.D., Ph.D.

N ENGL J MED 365;8 NEJM.ORG AUGUST 25, 2011

A KILLER STRAIN

The *Escherichia coli* outbreak has swept across Europe from Germany over the past month. Numbers correct as of 7 June.



MICROBIOLOGY

Microbe outbreak panics Europe


Kampmeier et al. Current Topics in Microbiology and Immunology, vol 416. Springer

Began in Germany and resulted in >4,000 infections and >50 deaths across multiple countries.

Fenugreek sprouts

Case study

"Classical" Enteroaggregative *E. coli*



Unusual hemolytic-uremic syndrome

Browsing publicly available genomes in Proksee

Genome assembly ASM98676v1

Download

 datasets

curl

FTP

NCBI RefSeq assembly	GCF_000986765.1
Submitted GenBank assembly	GCA_000986765.1
Taxon	Escherichia coli O104:H4 str. C227-11
Strain	C227-11

Chromosome	GenBank	RefSeq	Size (bp)
chromosome	CP011331.1	NZ_CP011331.1	5,292,862
plasmid	CP011332.1	NZ_CP011332.1	75,079
plasmid	CP011333.1	NZ_CP011333.1	16,948
plasmid	CP011334.1	NZ_CP011334.1	2,954
plasmid	CP011335.1	NZ_CP011335.1	7,893
plasmid	CP011336.1	NZ_CP011336.1	18,052
plasmid	CP011337.1	NZ_CP011337.1	17,818
plasmid	CP011338.1	NZ_CP011338.1	14,400

<https://proksee.ca/>

Proksee

New Project

My Projects

About

Help



Welcome to Proksee

Proksee is an expert system for genome assembly, annotation and visualization. To begin using Proksee, provide a complete genome sequence, sequencing reads or a CGView/Proksee map JSON file.

Please Cite the Following

Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C, Graham M, Van Domselaar G, and Stothard P
[Proksee: in-depth characterization and visualization of bacterial genomes](#)
Nucleic Acids Research, 2023, gkad326, <https://doi.org/10.1093/nar/gkad326>

Create a New Map

Genome

Reads

Map JSON

Genome Sequence

Upload

NCBI

Browse

Upload file...

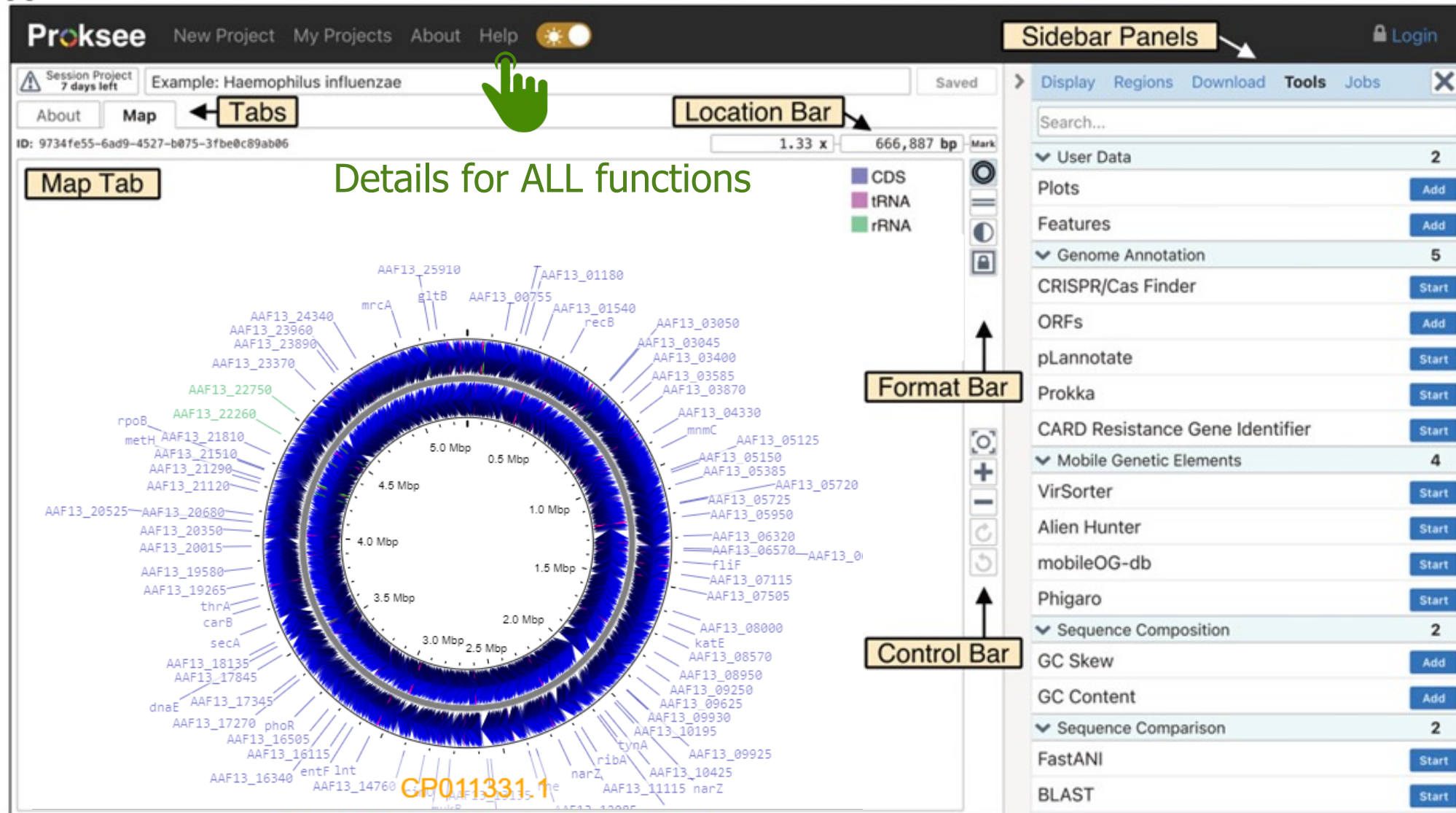
Proksee accepts GenBank, EMBL, FASTA, and raw formats. Sequences must be between 1,000 and 10 million bases long.

Create Map

OR

[Start an Example](#)

A



Annotate

Display Regions Download **Tools** Jobs

Search...

Genome Annotation 7

Bakta New Start

CARD Resistance Gene Identifier Start

CRISPR/Cas Finder Start

MITOS New Start

ORFs Add

pLannotate Start

Prokka Start

Map Refinements 2

BLAST Formatter Start

Track List Caption Add

Mobile Genetic Elements 4

Alien Hunter Start

mobileOG-db Start

Phigaro Start

VirSorter Start

Sequence Comparison 2

BLAST Start

FastANI Start



Compare

Display Regions Download **Tools** Jobs

Search...

Genome Annotation 7

Bakta New Start

CARD Resistance Gene Identifier Start

CRISPR/Cas Finder Start

MITOS New Start

ORFs Add

pLannotate Start

Prokka Start

Map Refinements 2

BLAST Formatter Add

Track List Caption Add

Mobile Genetic Elements 4

Alien Hunter Start

mobileOG-db Start

Phigaro Start

VirSorter Start

Sequence Comparison 2

BLAST Start

FastANI Start



Predict AMRg

Display Regions Download **Tools** Jobs

Search...

Genome Annotation 7

Bakta New Start

CARD Resistance Gene Identifier Start

CRISPR/Cas Finder Start

MITOS New Start

ORFs Add

pLannotate Start

Prokka Start

Map Refinements 2

BLAST Formatter Add

Track List Caption Add

Mobile Genetic Elements 4

Alien Hunter Start

mobileOG-db Start

Phigaro Start

VirSorter Start

Sequence Comparison 2

BLAST Start

FastANI Start



Download project

Display Regions **Download** Tools Jobs

Image - PNG

Download the map as a png image. The maximum size for the width/height is 8000px.

	Download Size	Current Size
Width:	1,079 px	1078.765625 px
Height:	573 px	573.203125 px
Percent:	100 %	

Download

Image - SVG

JSON (Reloadable Map Data)

Download the CGView JSON data. This file can be loaded later to view and edit the map.

Download

Sequences

Download image

Display Regions **Download** Tools Jobs

Image - PNG

Download the map as a png image. The maximum size for the width/height is 8000px.

	Download Size	Current Size
Width:	1,079 px	1078.765625 px
Height:	573 px	573.203125 px
Percent:	100 %	

Download

Image - SVG

JSON (Reloadable Map Data)

Download the CGView JSON data. This file can be loaded later to view and edit the map.

Download

Sequences

Case study



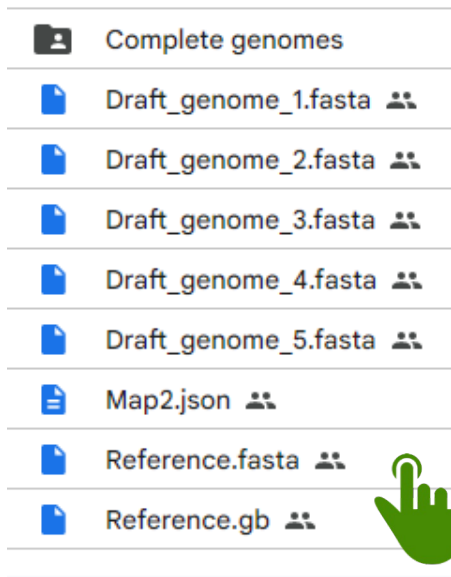
bioRxiv preprint doi: <https://doi.org/10.1101/2024.01.05.574329>; this version posted January 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Research Article: Convergence and global epidemiology of *Klebsiella pneumoniae* plasmids harbouring the *iuc3* virulence locus.

Marjorie J Gibbon^{1†}, Natacha Couto^{1,2,†}, Keira Cozens¹, Samia Habib¹, Lauren Cowley¹, David Aanensen², Jukka Corander^{3,4,5}, Harry Thorpe³, Marit AK Hetland^{6,7}, Davide Sassera⁸, Cristina Merla⁹, Marta Corbella⁹, Carolina Ferrari⁹, Katy ME Turner¹⁰, Kwanrawee Sirikancha¹¹, Punyawee Dulyayangkul¹⁰, Nour Alhusein¹², Nisanart Charoenlap¹¹, Visanu Thamlikitkul¹³, Matthew B Avison¹⁰, Edward J Feil^{1*}

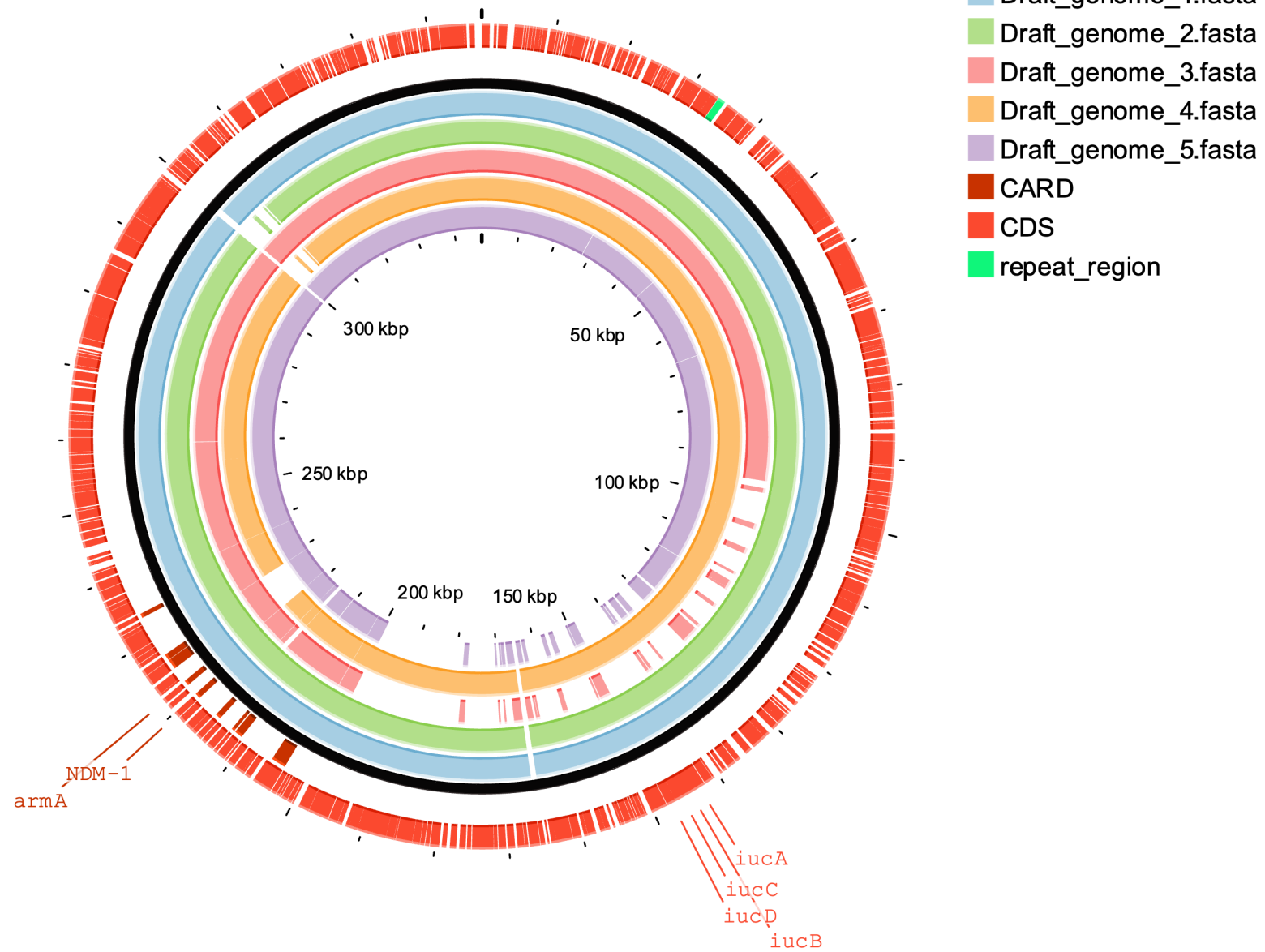
Application (hands on)

You are tracking an outbreak of convergent *K. pneumoniae*. You just received the draft assemblies for 5 genomes from patients in your hospital. Great luck, a reference plasmid sequence from an outbreak strain was just released by ECDC.

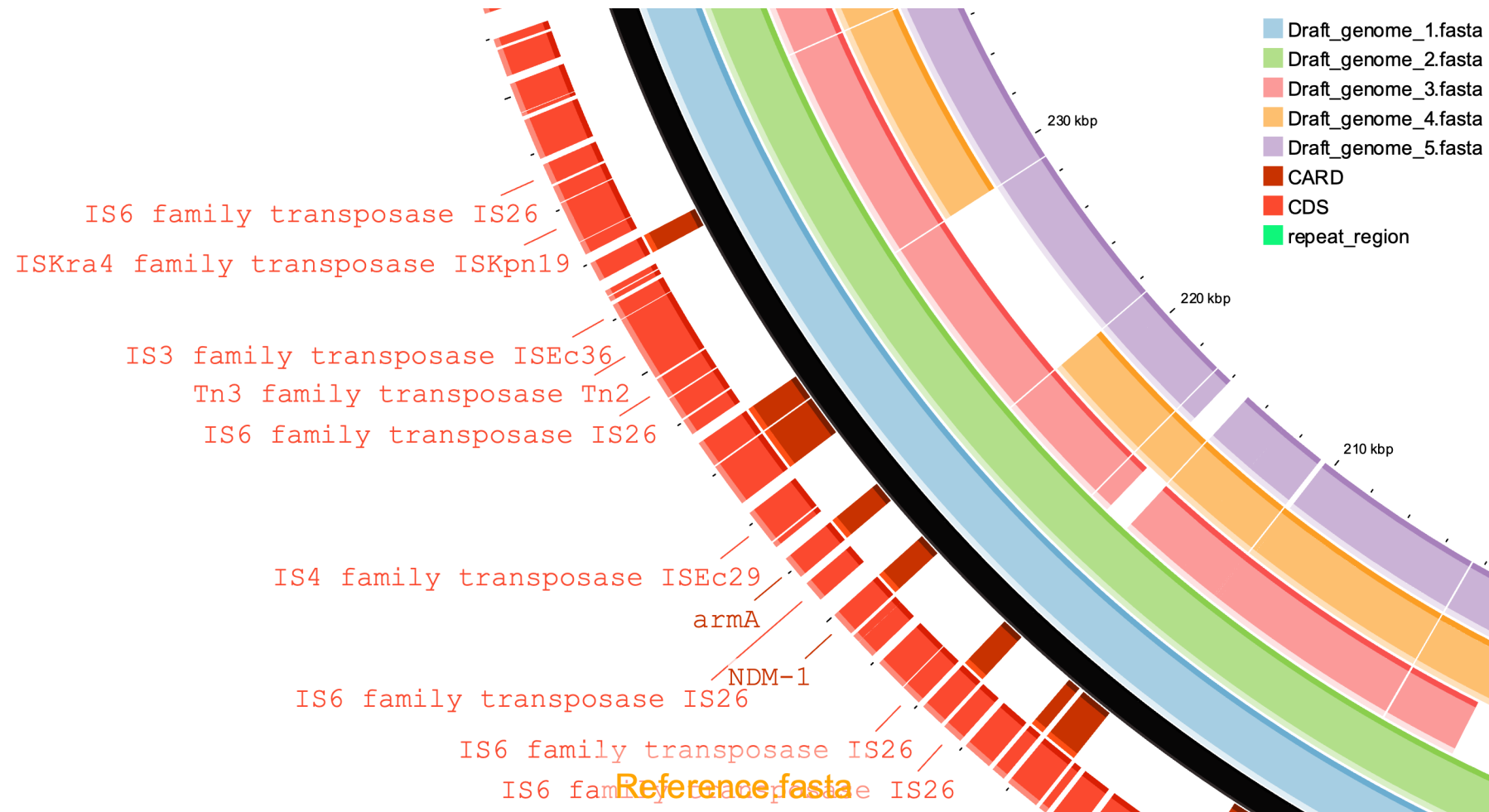


1. Upload reference plasmid and annotate
2. Identify AMRg and virulence genes
3. Compare draft assemblies for genomes 1-5
4. Analyze results

Application (hands on)















Application (hands on)



Application (hands on)

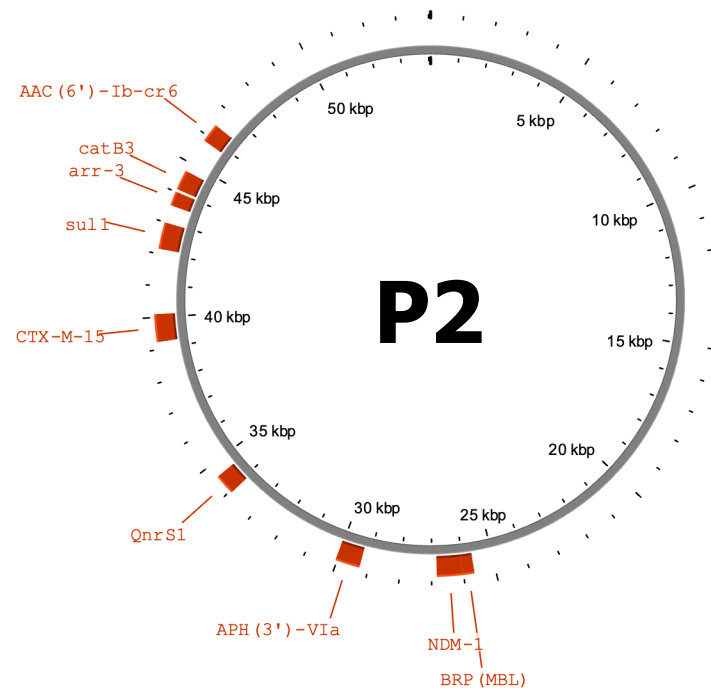
Your manuscript is ready, the competition is fierce, and you want to be first! Nevertheless, you decide to do long-read sequencing on two outbreak isolates “Just to be sure”.

 Complete genomes 		
 Complete_genome1_chr.fasta		5,2 Mo
 Complete_genome1_p1.fasta		366 Ko
 Complete_genome2_chr.fasta		5,1 Mo
 Complete_genome2_p1.fasta		328 Ko
 Complete_genome2_p2.fasta		53 Ko

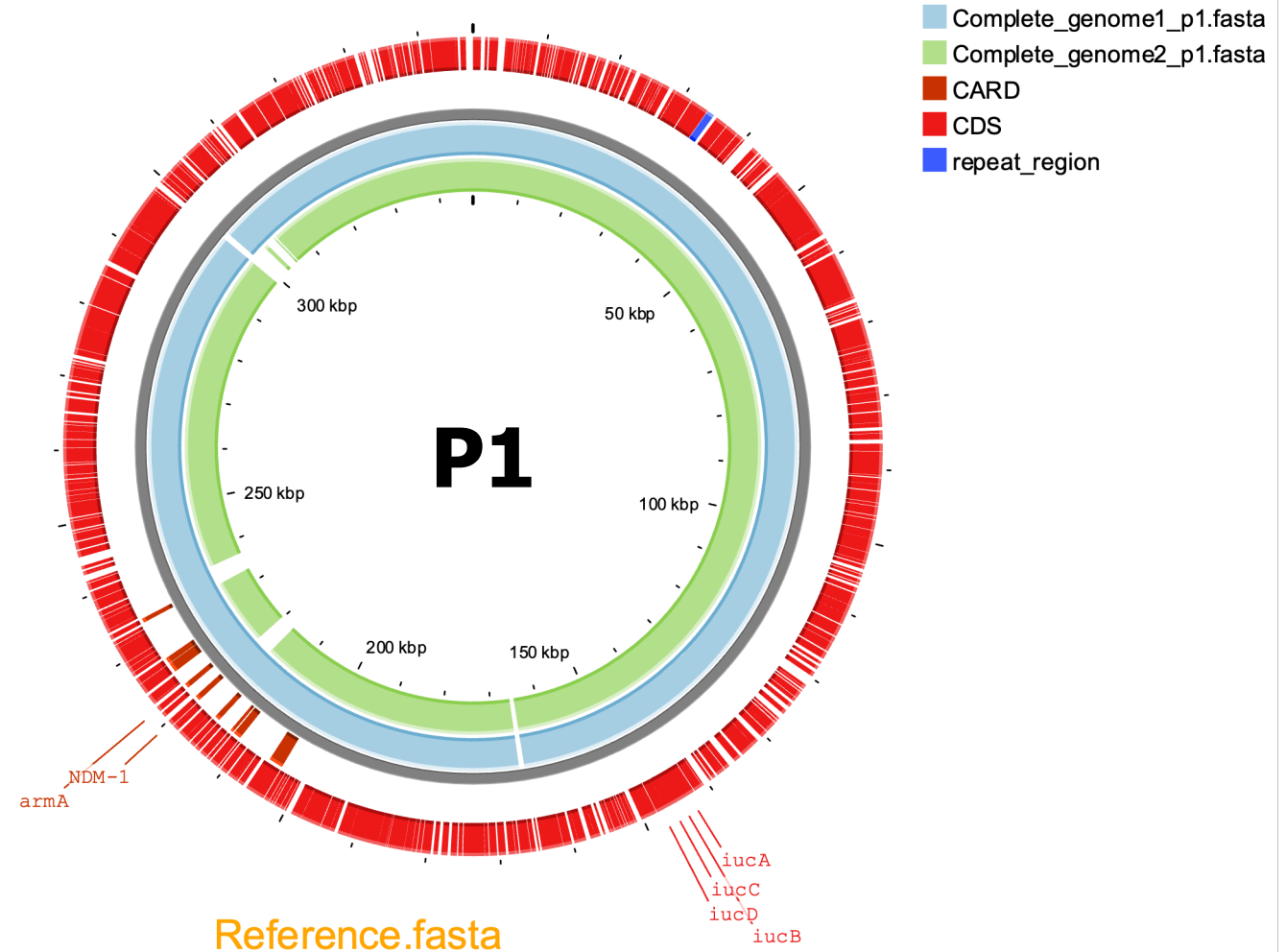
1. Dry you tears
2. What is going on?

Application (hands on)

 Complete_genome1_chr.fasta		5,2 Mo
 Complete_genome1_p1.fasta		366 Ko
 Complete_genome2_chr.fasta		5,1 Mo
 Complete_genome2_p1.fasta		328 Ko
 Complete_genome2_p2.fasta		53 Ko



Complete_genome2_p2.fasta



Reference.fasta

Summary

Advantages of Proksee

- User-friendly Interface
- Comprehensive Analysis (annotation and comparative genomics)
- Integrative Data Visualization

Limitations of Proksee

- Requires Internet Connection
- Limited to Prokaryotic Genomes
- Potential Scalability Issues
- Reference-based!
- Mapping draft genomes is risky vs closed genomes

References



Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen CY, Graham M, Van Domselaar G, Stothard P. Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res.* 2023 Jul 5;51(W1):W484-W492. doi:10.1093/nar/gkad326. PMID: 37140037; PMCID: PMC10320063.

Stothard P, Grant JR, Van Domselaar G. Visualizing and comparing circular genomes using the CGView family of tools. *Brief Bioinform.* 2019 Jul 19;20(4):1576-1582. doi: 10.1093/bib/bbx081. PMID: 28968859; PMCID: PMC6781573.

Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.* 2011 Aug 8;12:402. doi: 10.1186/1471-2164-12-402. PMID: 21824423; PMCID: PMC3163573.

Vallenet D, Calteau A, Dubois M, Amours P, Bazin A, Beuvin M, Burlot L, Bussell X, Fouteau S, Gautreau G, Lajus A, Langlois J, Planel R, Roche D, Rollin J, Rouy Z, Sabatet V, Médigue C. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res.* 2020 Jan 8;48(D1):D579-D589. doi: 10.1093/nar/gkz926. PMID: 31647104; PMCID: PMC7145621.

Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of Francois Lebreton, PhD

Francois Lebreton, PhD
francois.lebreton.ctr@health.mil
@Fr_Lebreton