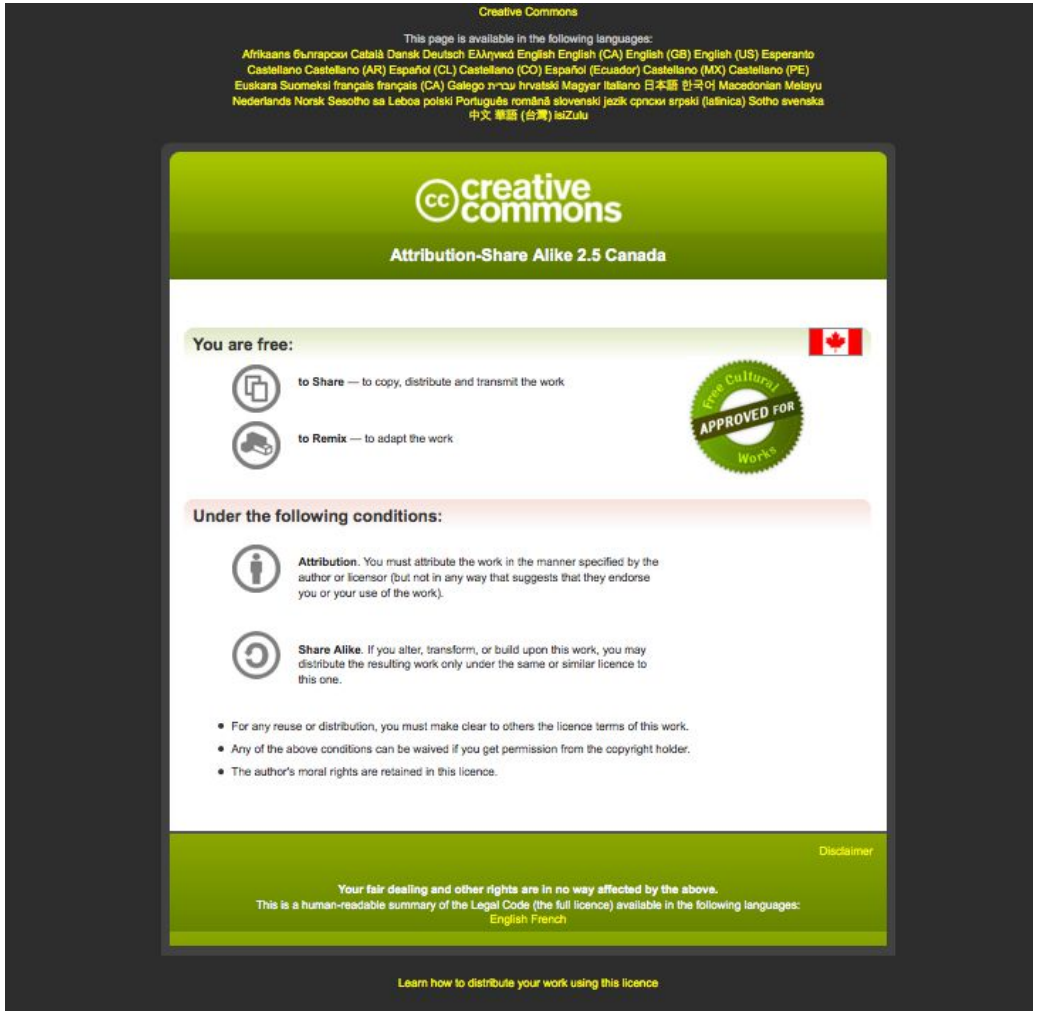
bioinformatics.ca

Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io


Creative Commons
This page is available in the following languages:
Afrikaans български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto
Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
Euskara Suomeksi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska
中文 華語 (台灣) isiZulu
creative commons
Attribution-Share Alike 2.5 Canada
You are free:
to Share — to copy, distribute and transmit the work
to Remix — to adapt the work
Free Cultural
APPROVED FOR
Works
Under the following conditions:
Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.
For any reuse or distribution, you must make clear to others the licence terms of this work.
Any of the above conditions can be waived if you get permission from the copyright holder.
The author's moral rights are retained in this licence.
Disclaimer
Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
English French
Learn how to distribute your work using this licence

Module 9: Mobile Genetic Elements and Environmental Microbiome

Gary Van Domselaar (for Rob Beiko)

Infectious Disease Genomic Epidemiology

April 18–21, 2023

bioinformatics.ca

Public Health Agency of Canada

Agence de la santé publique du Canada

DALHOUSIE UNIVERSITY

University of Manitoba

Learning Objectives

- By the end of this lecture, you will
  - Understand how environmental sequencing differs from clinical genomic sequencing;
  - Understand how environmental samples are used for pathogen surveillance;
  - Know the different types of mobile genetic elements that mediate horizontal gene transfer
  - Understand the impact of horizontal gene transfer on the analysis of the microbiome
  - Know the main bioinformatics tools that are used for pathogen detection and characterization.

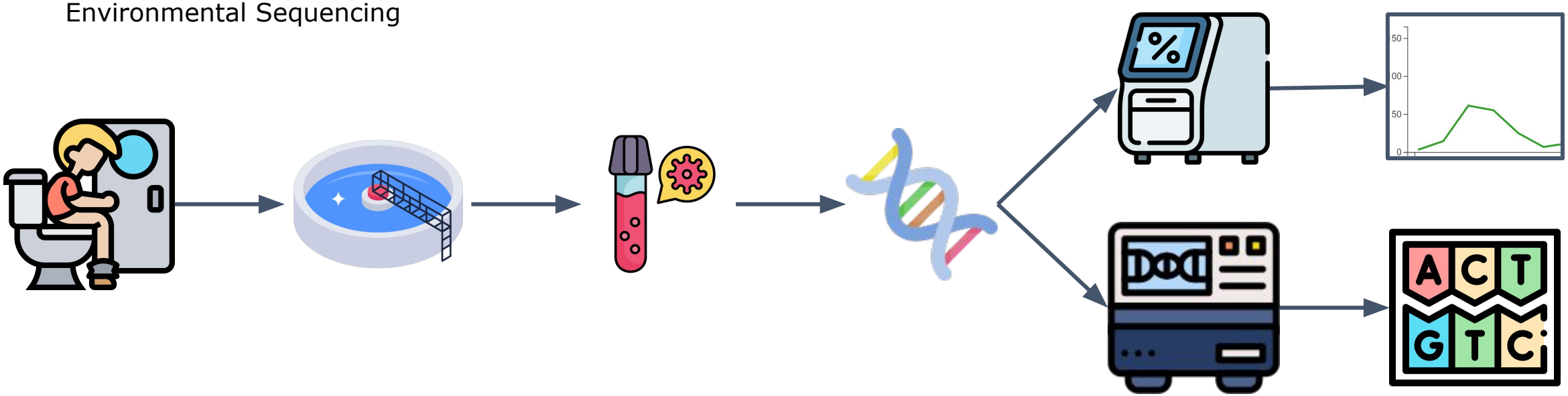
Part 1: Clinical sequencing versus environmental sequencing

Example: SARS-CoV-2

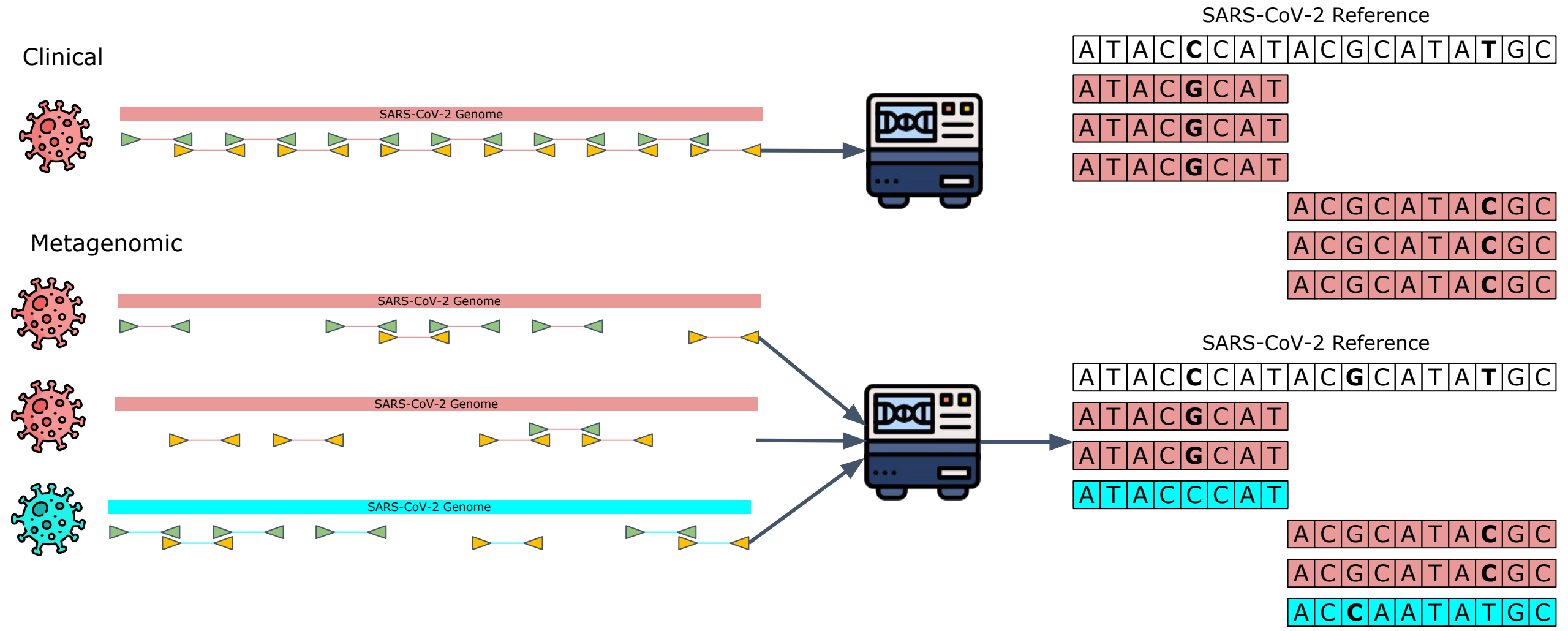
Clinical Sequencing

Example: SARS-CoV-2

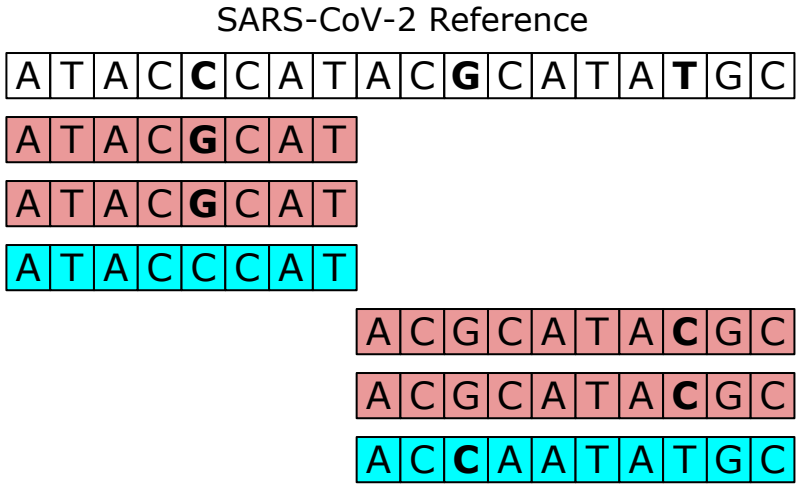
Environmental Sequencing

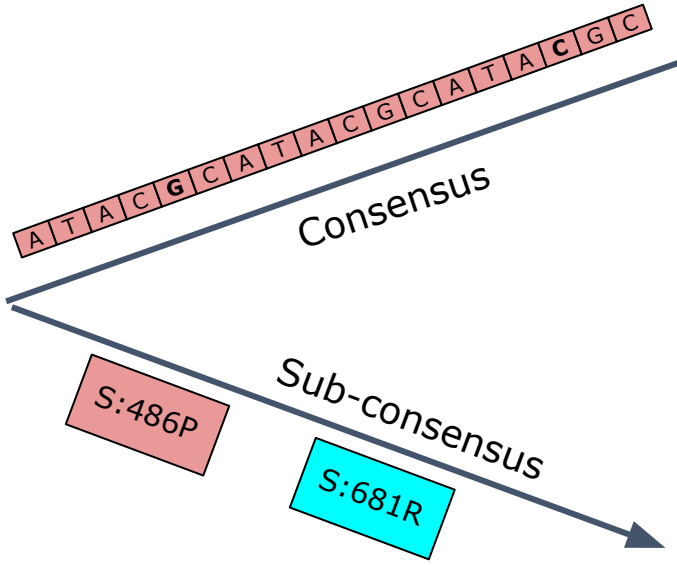


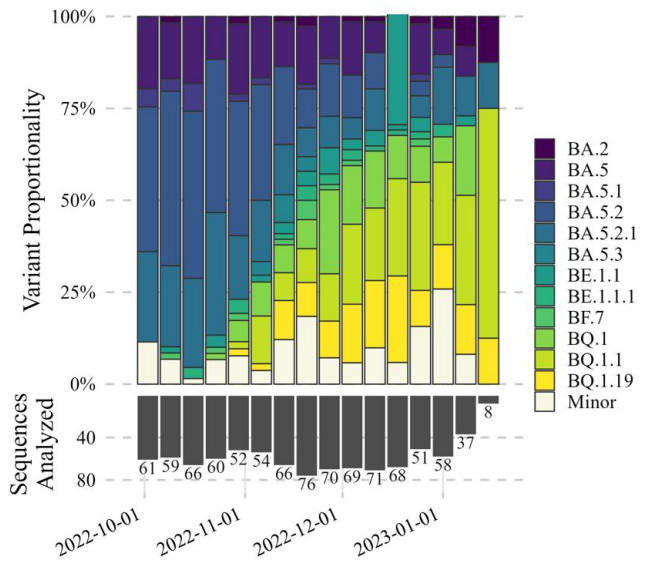
Clinical SARS-CoV-2 genome Sequencing vs environmental sequencing

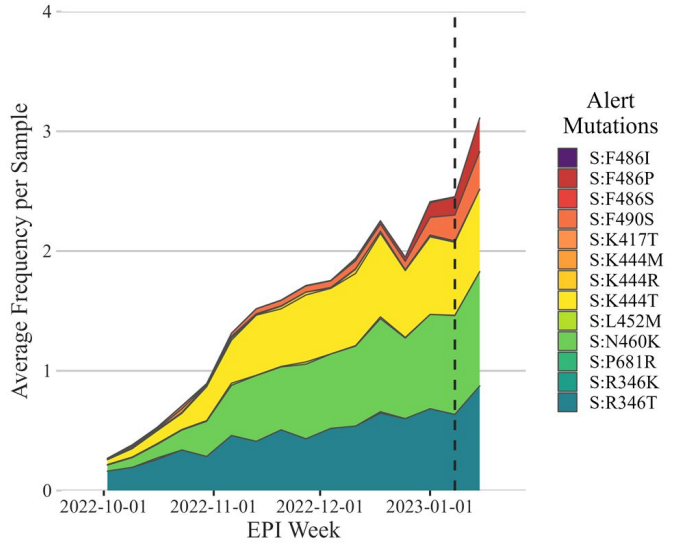


Consensus and sub-consensus





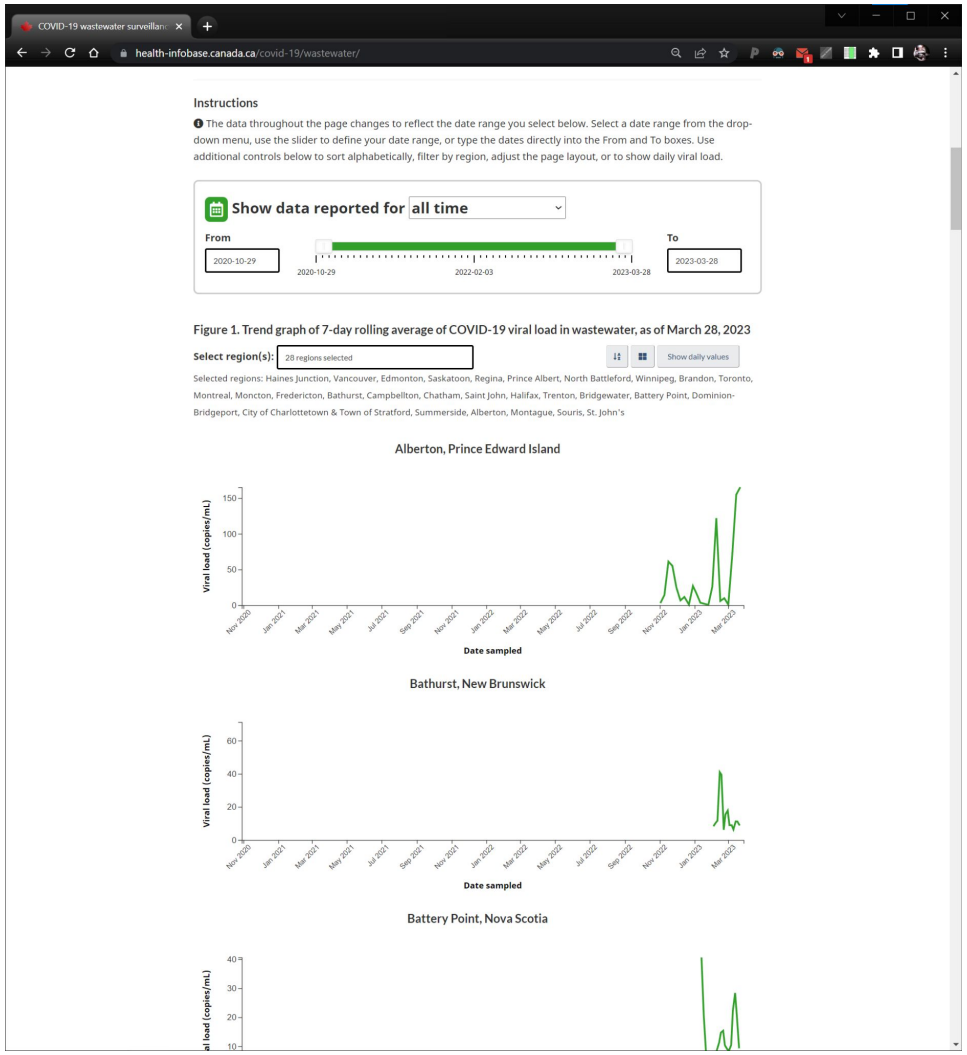




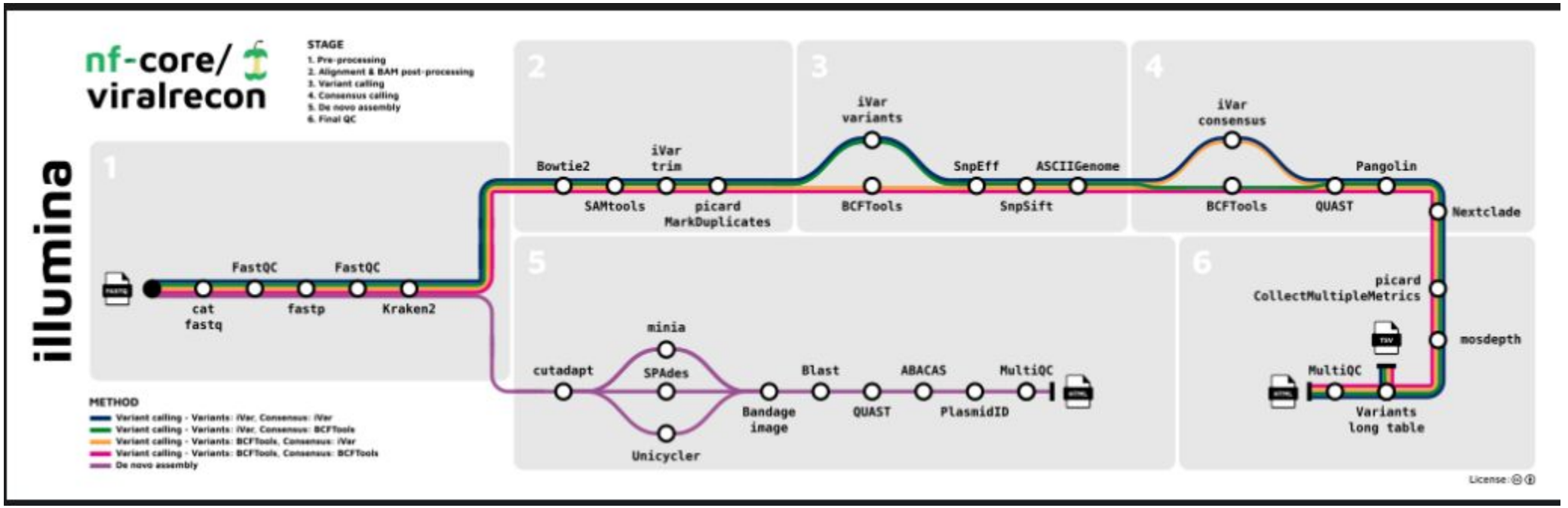
COVID-19 wastewater surveillance dashboard

RT-PCR results for sampled communities are published at the PHAC COVID-19 wastewater surveillance dashboard.

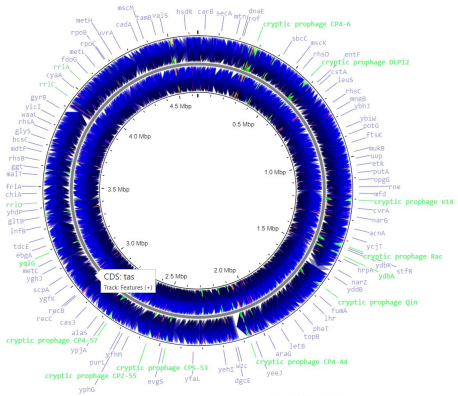
https://health-infobase.canada.ca/covid-19/wastewater/

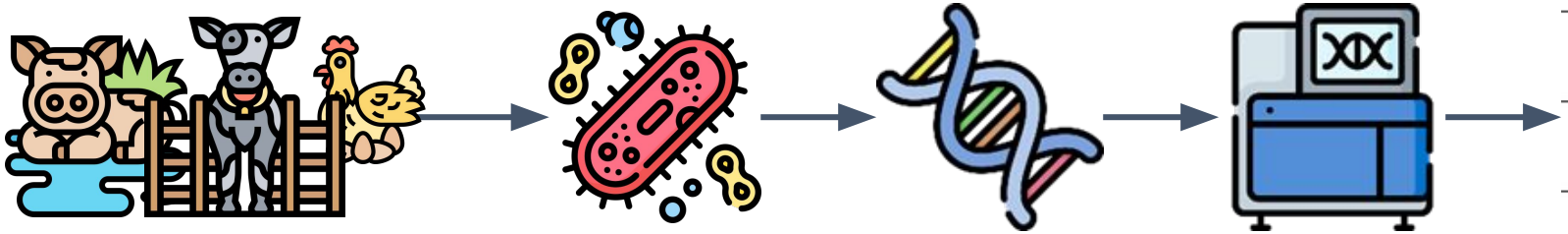


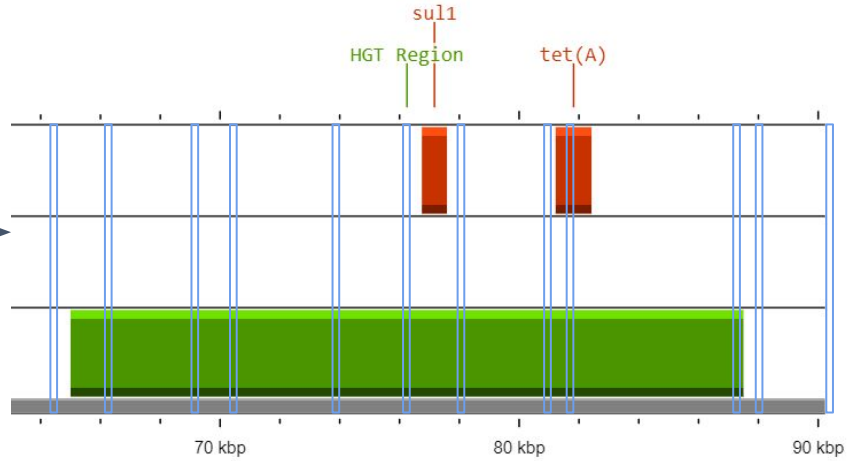
SARS-CoV-2 environmental sequencing pipeline



Clinical Sequencing versus Environmental Sequencing



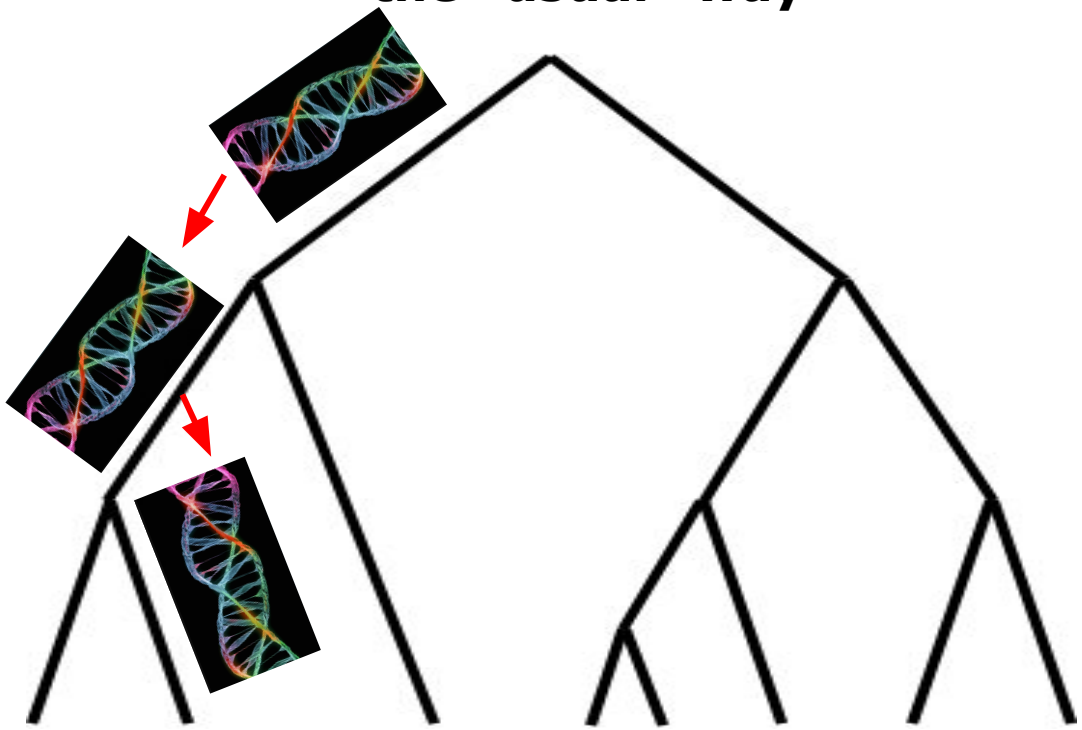




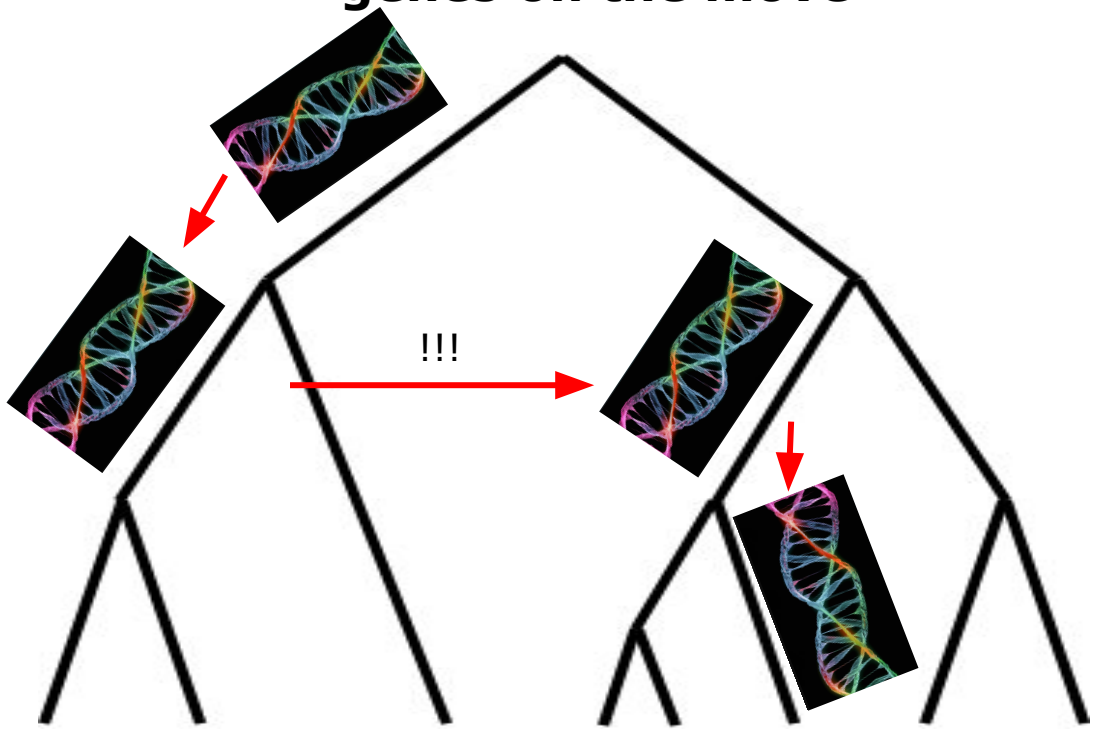
Part 2:genes that move

Recombination and Lateral (or horizontal) gene transfer

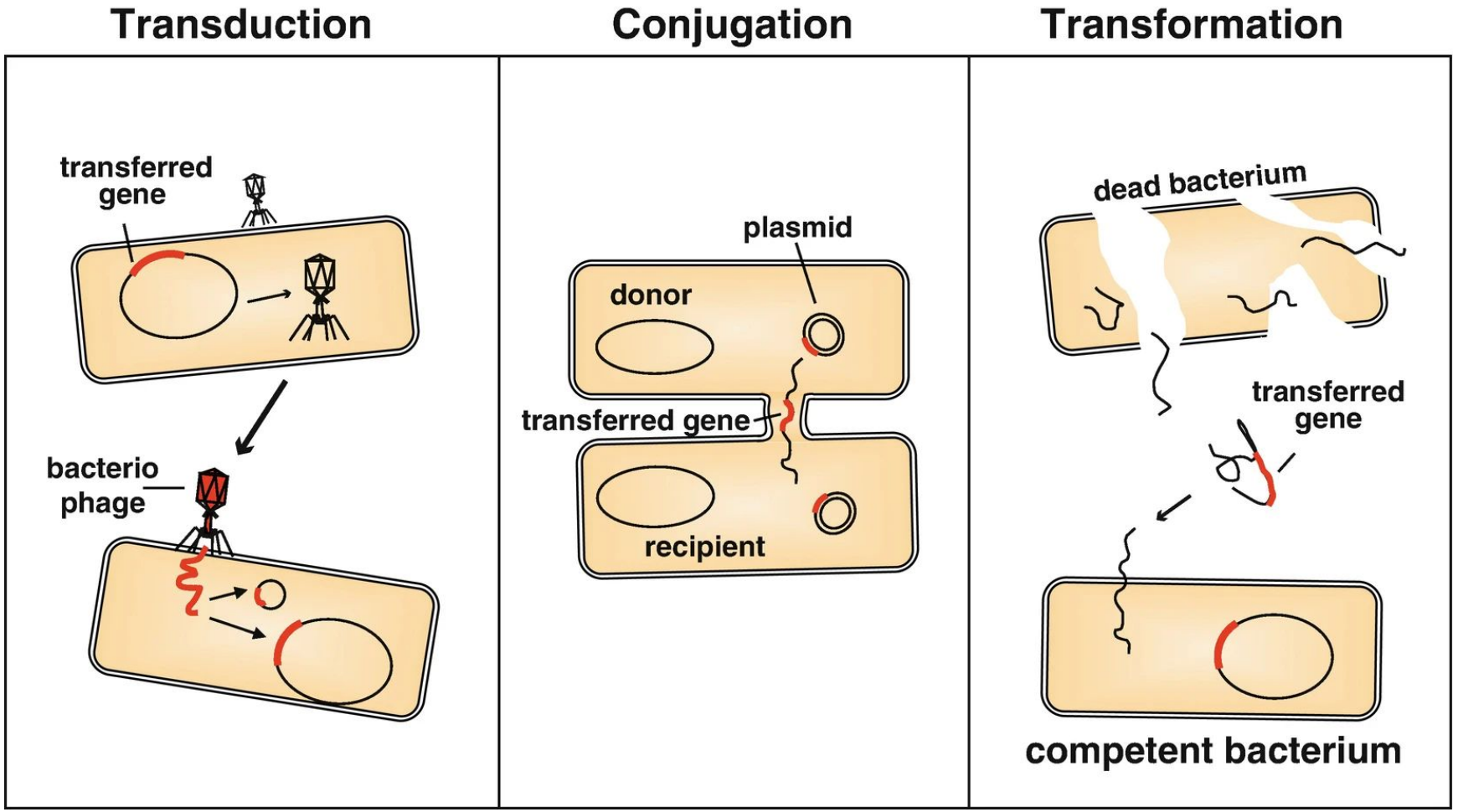
Vertical inheritance - the "usual" way



Recombination and LGT - genes on the move



Change of Address



Phages!

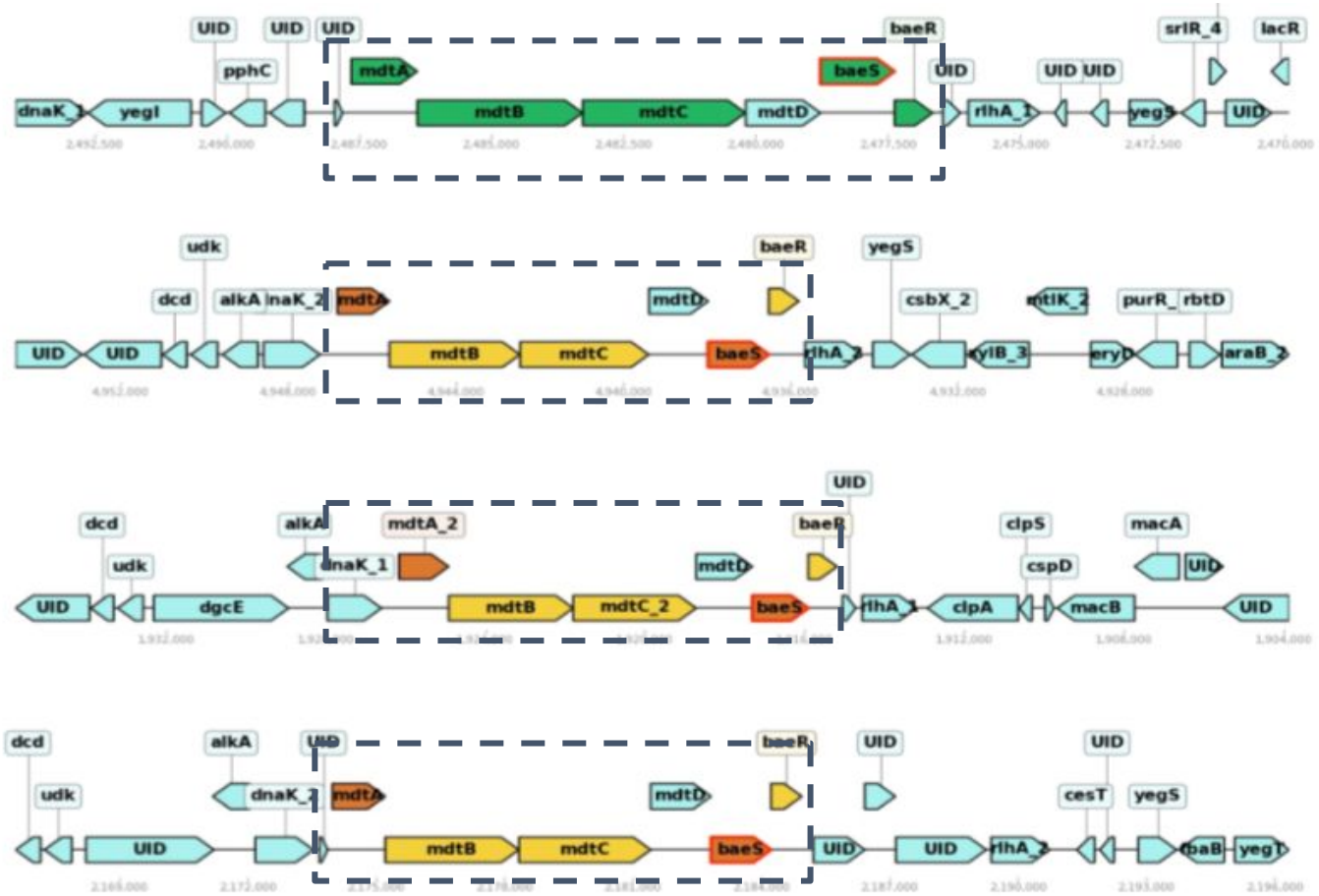
Plasmids!

Sort of just DNA!

Why is LGT important?

LGT gives microorganisms the capacity to acquire new functions from even potentially distantly related species

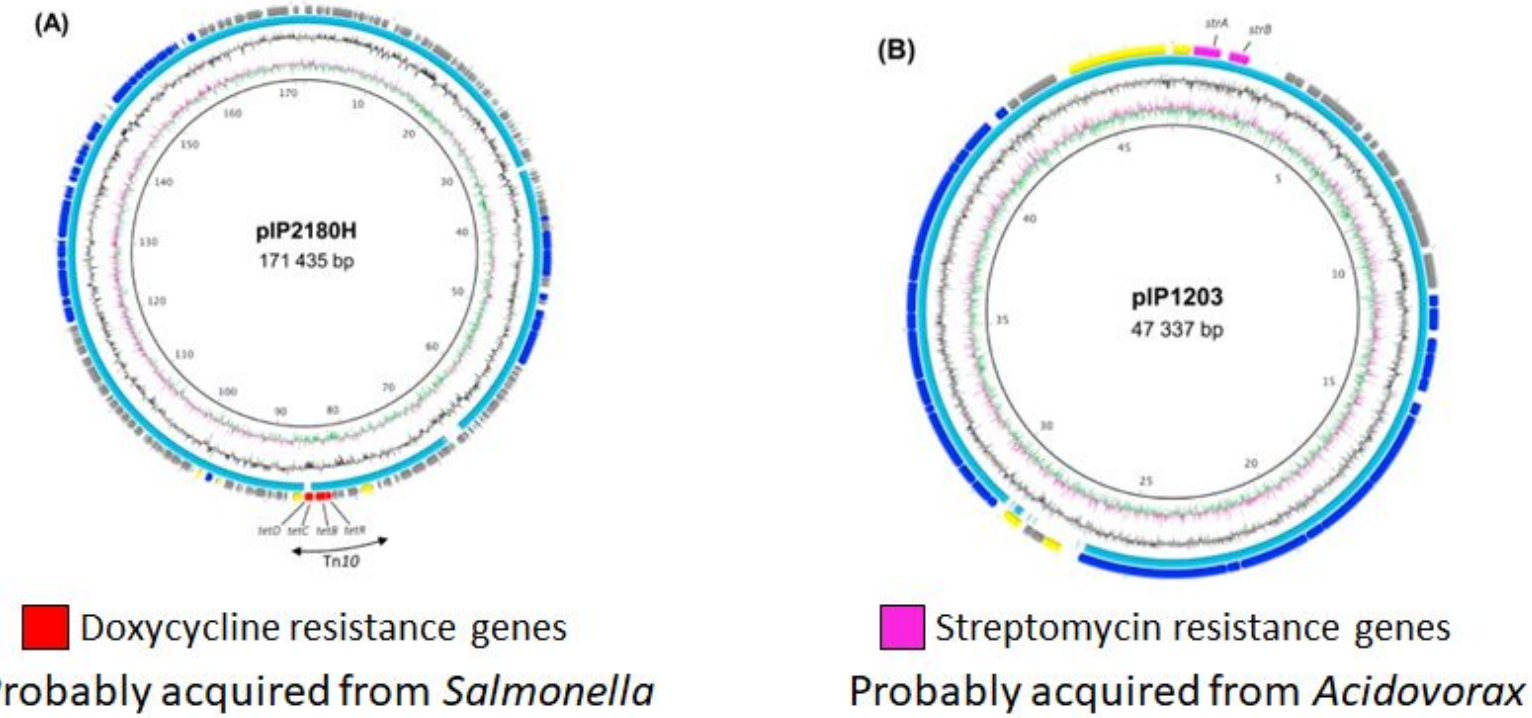
Example: *baeS* (aminoglycoside resistance) conserved neighbourhood between *Salmonella, Citrobacter, Escherichia,* and *Enterobacter*



Navanekere and Beiko (2021)

LGT and the Spread of Antimicrobial Resistance

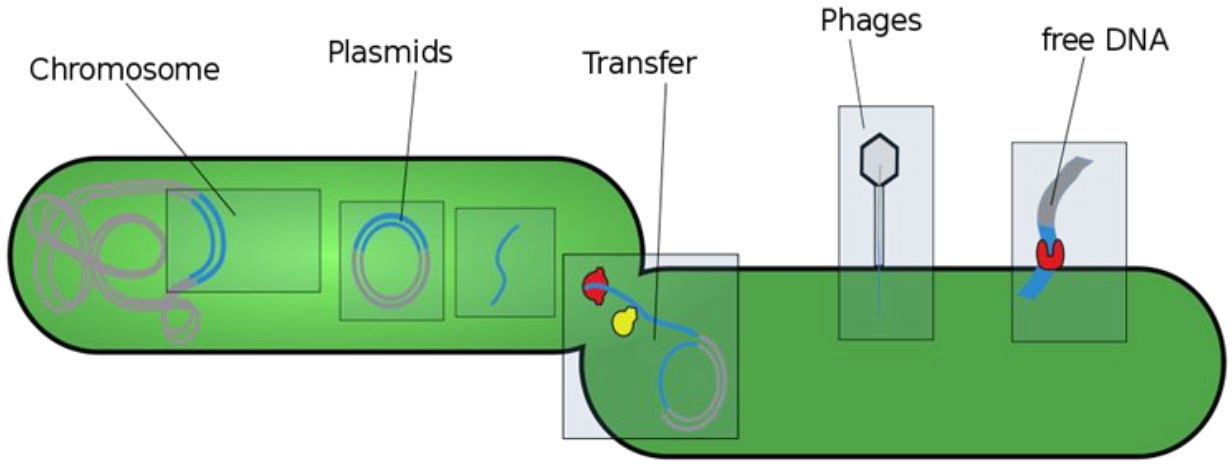
Yersinia pestis is the causative agent of plague. And as if that alone isn't bad enough…



Cabanel et al. (2018) *Int J Antimicrob Agents*

Mobile Genetic Elements

- Mobile Genetic Elements are a type of DNA that can move around within the genome. They include:
  - Plasmids
  - Transposons (also called transposable elements)
  - Bacteriophage elements, like Mu, which integrates randomly into the genome
  - Genomic islands
- The total of all mobile genetic elements in a genome are referred to as the mobilome.



Detecting recombination and transfer is important and difficult

- Detecting transferred genes can help us identify major modes and vectors of transmission
  - *Which genes are being shared?*
  - *What is the mechanism?*
  - *Who is sharing with whom?*
  - *Where is the sharing happening?* (e.g., in the hospital, in the community, etc)

- What can we do with this information?
  - Identify highest risk factors (which genes are mobilized? Which are not?) for stewardship purposes
  - Decide where to prioritize interventions

- But finding evidence for recombination and LGT isn't trivial!

The clues at our disposal

(1) Genes with unexpectedly weird phylogenetic trees

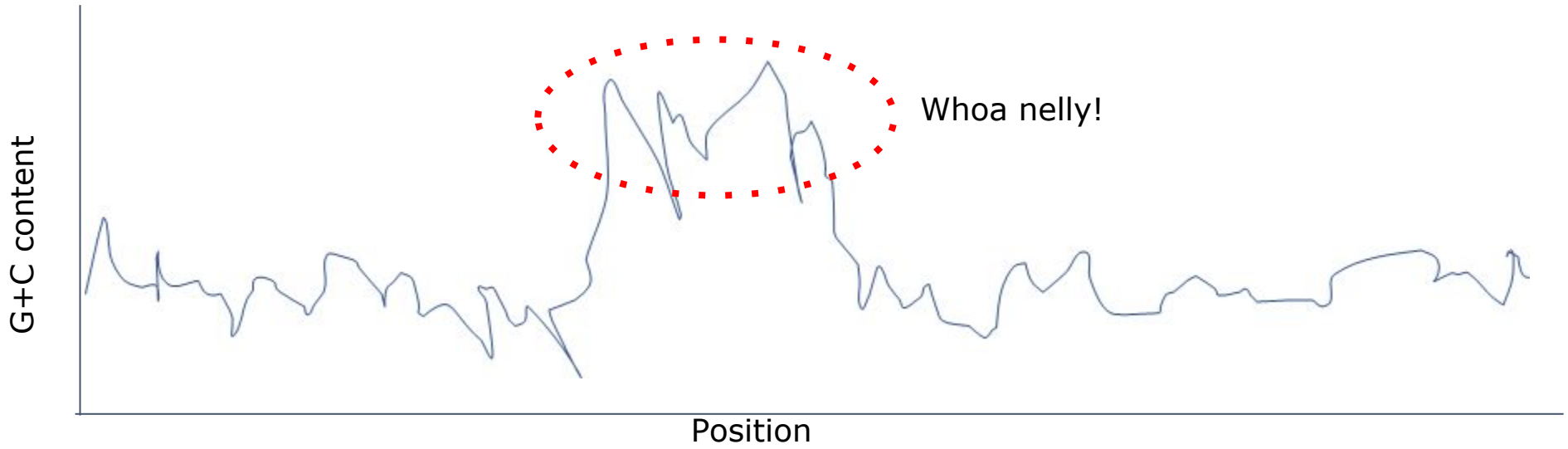
The SPECIES tree -
vertical evolution (probably)

The GENE tree -
Different! Why?

Whidden et al (2015) *Syst Biol*

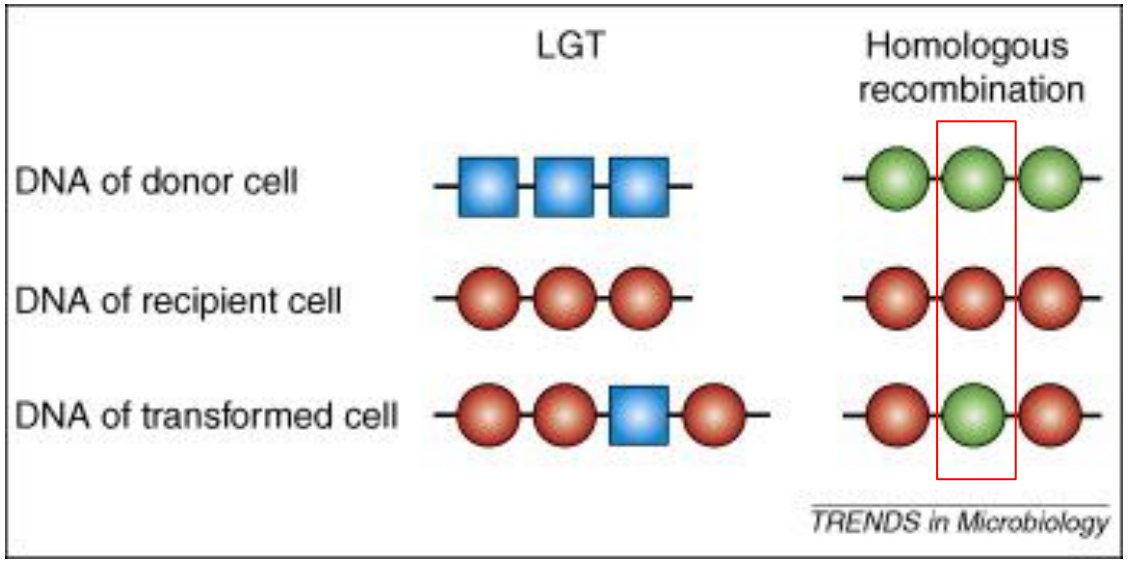
The clues at our disposal

Genes (or more generally stretches of DNA) with unexpectedly weird nucleotide composition



The clues at our disposal

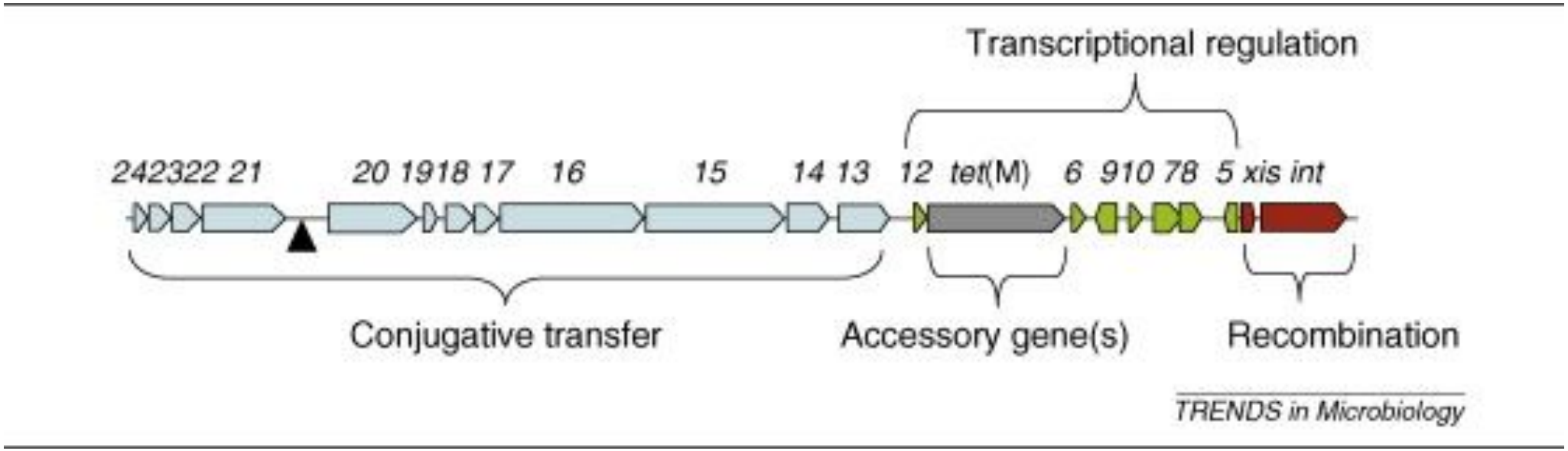
Unexpected sequence similarity between genomes



Vos (2009) *Trends Microbiol*

The clues at our disposal

Localization to MGEs (and other upwardly mobile genes in the neighbourhood)



Roberts and Mullany (2009) *Trends Microbiol*

Part 3: Detecting things!

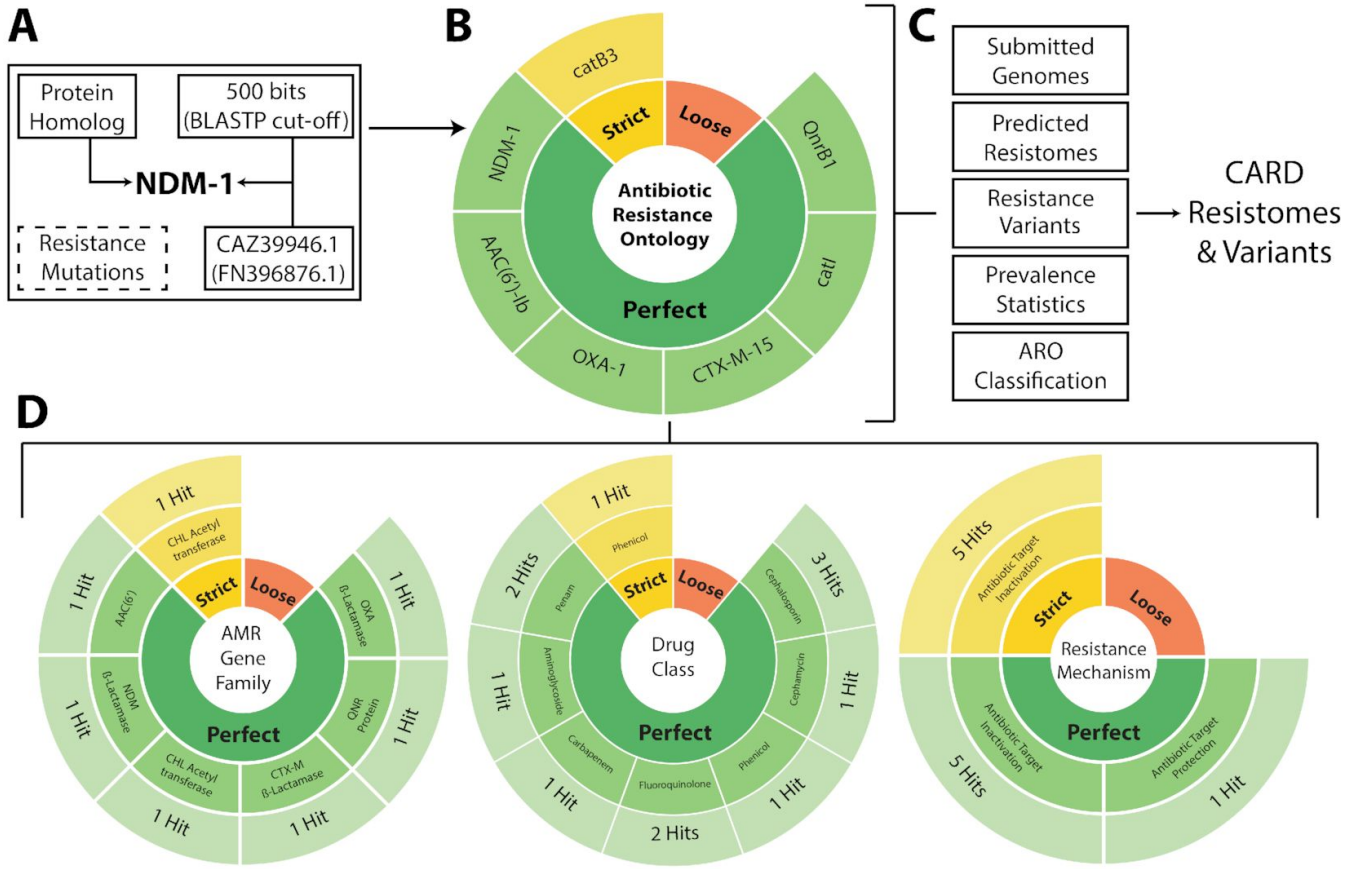
Finding mobile genetic elements

General clues:

- Because they bounce around a lot, MGEs can look “foreign” relative to the rest of the genome (e.g., different nucleotide compositions)
- Some genes can be (imperfect) signatures for MGEs:
  - Replicases (plasmids), Transposases, Integrases, Excisionases
  - Certain types of virulence factors and AMR genes
  - Secretion-system genes
- Search against reference databases can be very helpful
  - Homology search: BLAST/DIAMOND
  - Heuristic distances: MASH

AMR Genes - the Resistance Gene Identifier

Homology search vs. manually curated database



Alcock et al. (2020) *Nucleic Acids Res*

Plasmids: MOB-Suite

- MOB-cluster: Cluster a high-quality reference plasmid database (focused on Gram-negative enterics!)
- MOB-recon: find and classify plasmids in my dataset.
  - Circular contigs? It's a plasmid.
  - Look for relaxase or replicon genes -> candidate plasmid.
  - Compare against MOB-clustered database
- MOB-typer: look at relaxase and other information to classify as "conjugative", "mobilizable", or "non-mobilizable"

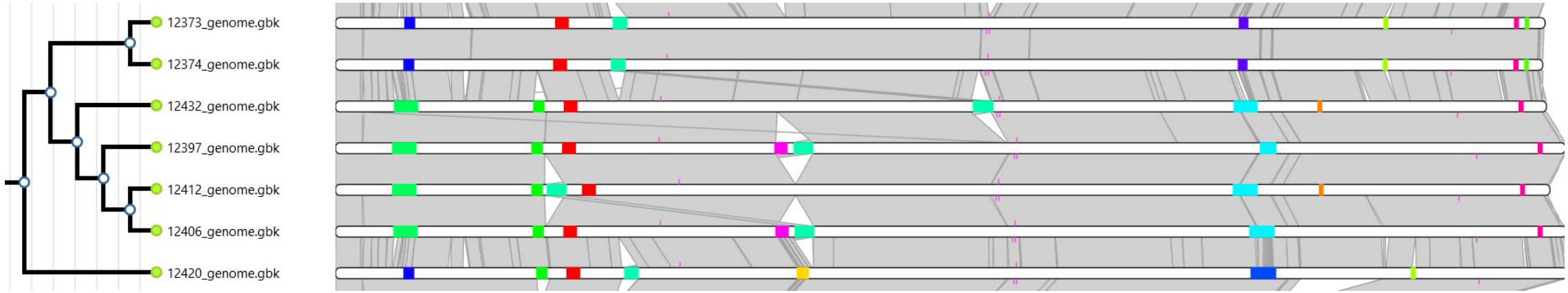
Robertson et al. (2020) *Microbial Genomics*

Genomic Islands: Island*

IslandPath: Look for unusual patterns of nucleotide composition, AND the presence of “smoking gun” genes such as mobility genes (using Pfam)

IslandView: View predicted islands with probable AMR genes / virulence factors

IslandCompare: Compare and cluster genomic islands across genomes



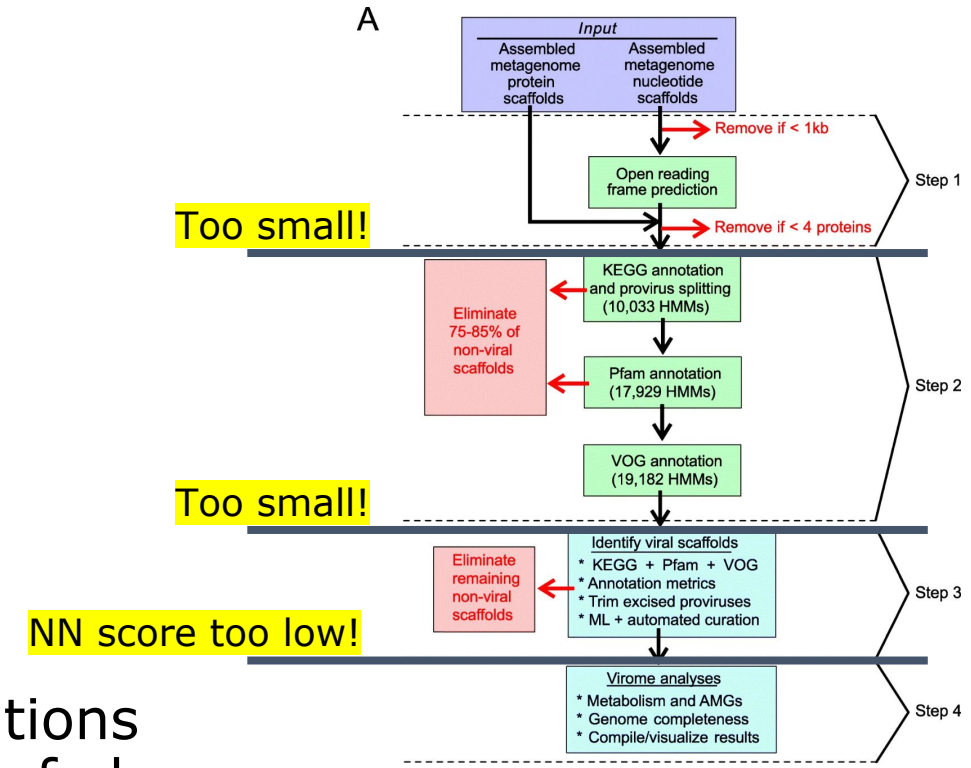
Bertelli, Gray et al. (2022) *Microbial Genomics*
https://islandcompare.ca/

Prophages: VIBRANT (and many, many, many others)

Uses reference databases:

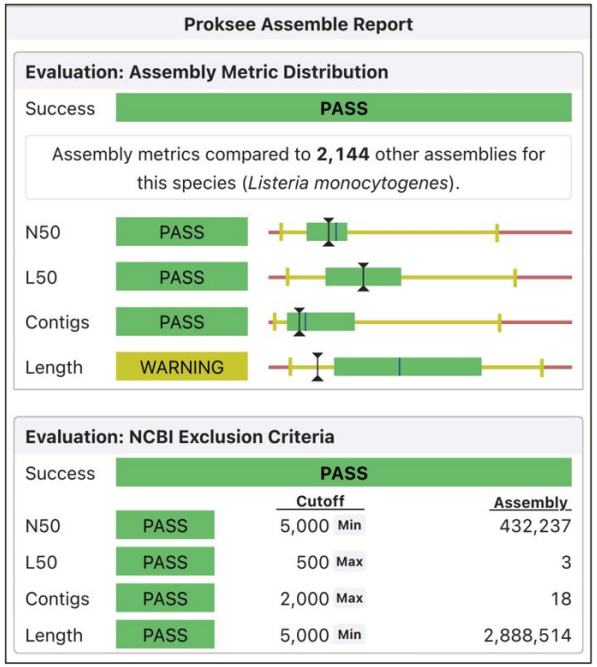
- KEGG KoFam
- Pfam
- Virus Orthologous Groups

Key concept is *v-scores* based on annotations from different databases; 27 of these are fed into a neural network to generate predictions

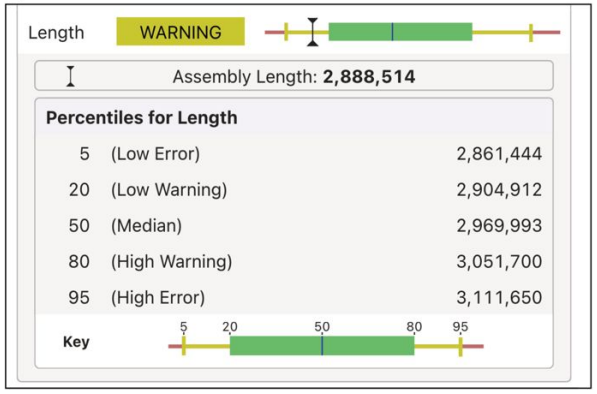


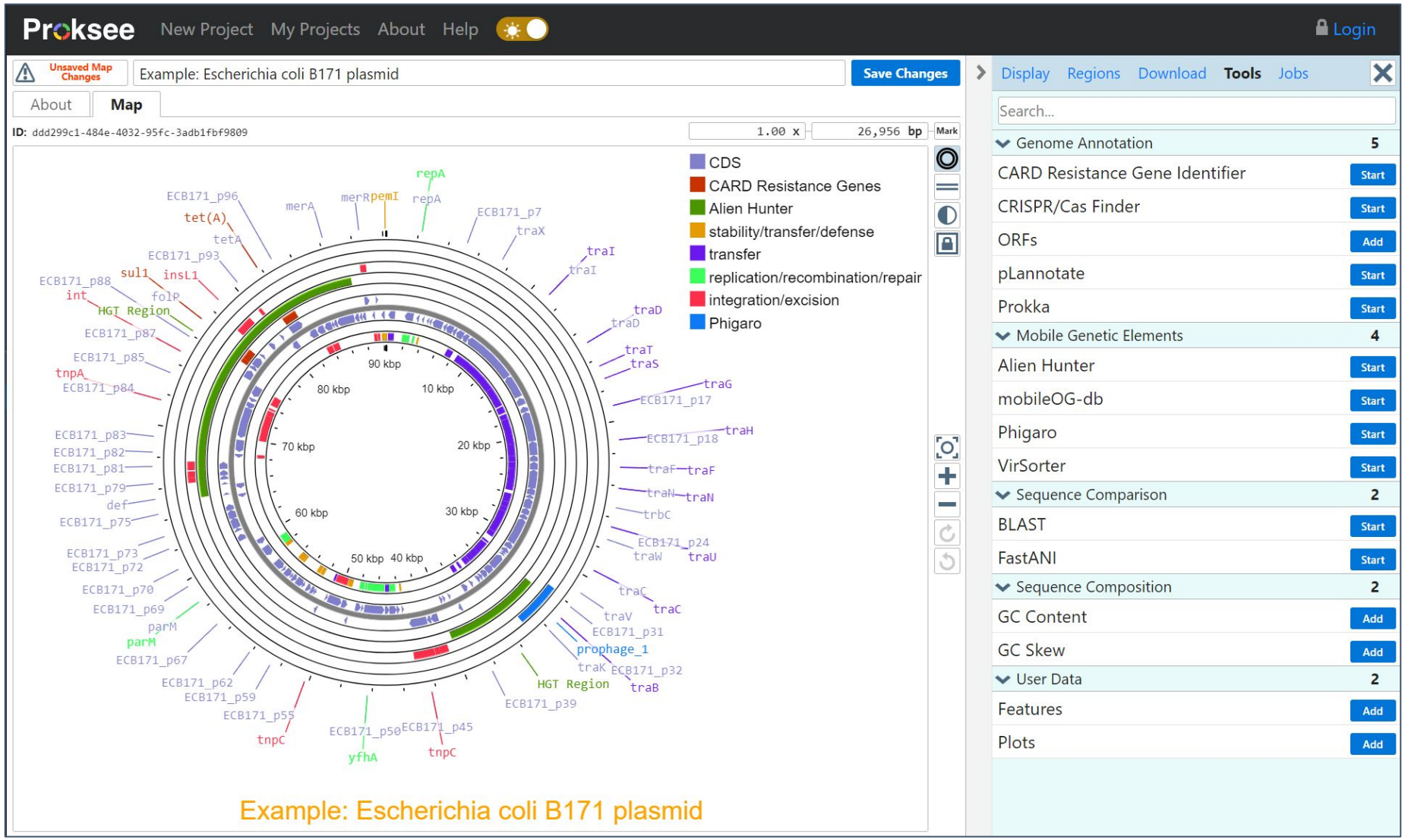
Comparison framework: https://github.com/linsalrob/ProphagePredictionComparisons

Proksee



B





https://proksee.ca

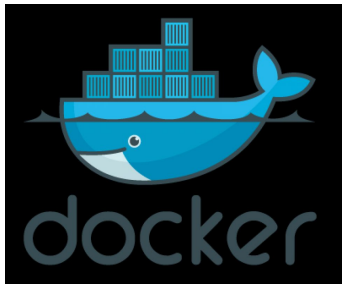
BUT

- MGEs are by definition highly mutable, so gene content / order / composition can change rapidly
- Databases are far better for some organisms (e.g., Gram-negative enterics) than others, so comparative methods may fall over
- Short-read assembly tends to fail spectacularly for MGEs - good luck getting beautiful circular plasmids with HiSeq data!
- Predicting the BOUNDARIES of things like genomic islands and prophages can be quite hard

I don’t feel like installing a billion software packages to analyze my genomes

So let’s create environments where software packages are (in theory) easier to deploy

Manage software dependencies without having to do all sorts of manual version control / etc



Virtual environment: basically a “fake OS” in which everything is already installed

(Singularity) Compatible with Docker images, but better integrated with host OS and less obnoxious re: access permissions

All have their advantages and disadvantages, and can interact with systems in odd (and sometimes frustrating) ways

But often better than the alternative!

Antimicrobial Resistance: Emergence, Transmission, and Ecology (ARETE)

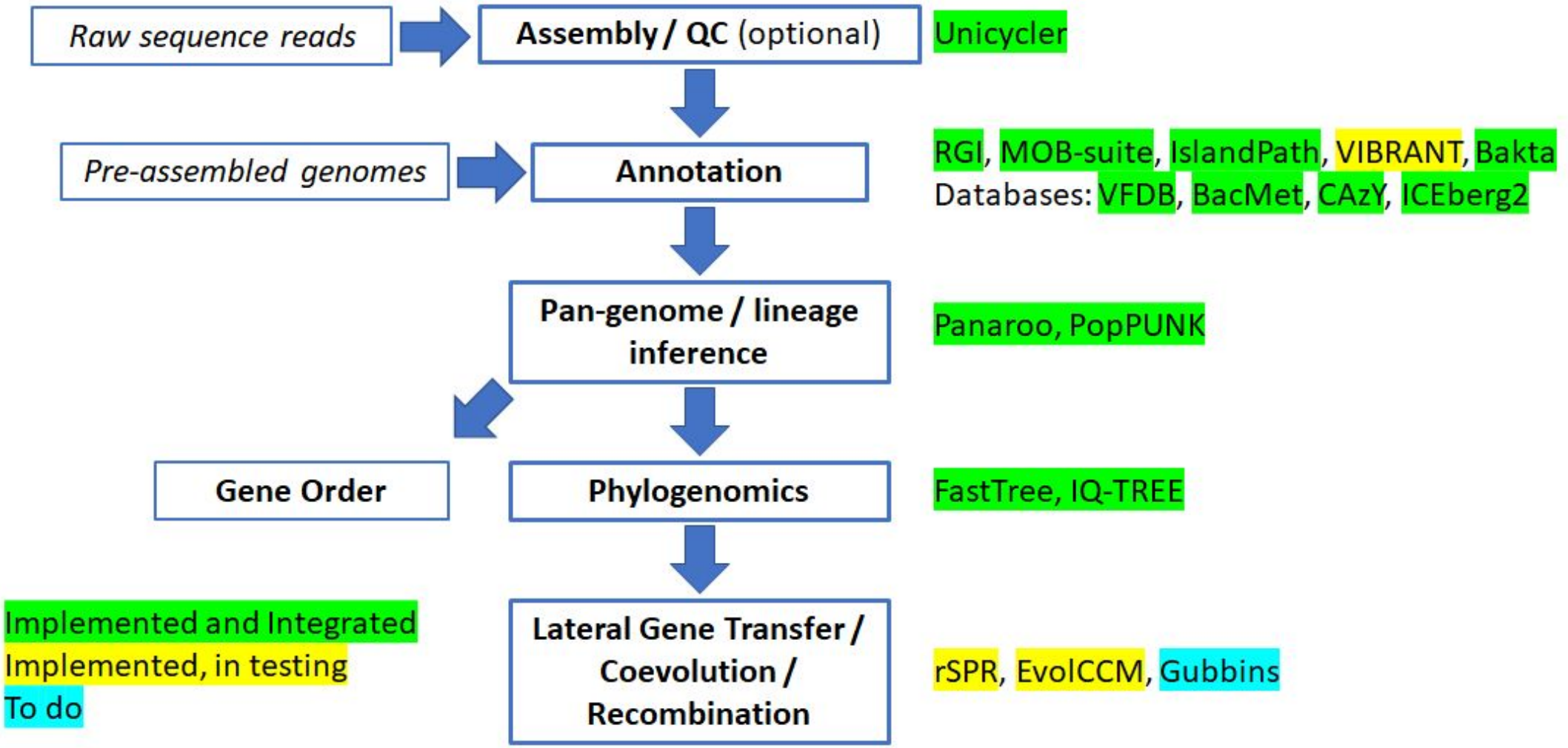
nextflow

A workflow manager that provides effective chaining of containers. Can handle job submission, resumption, …

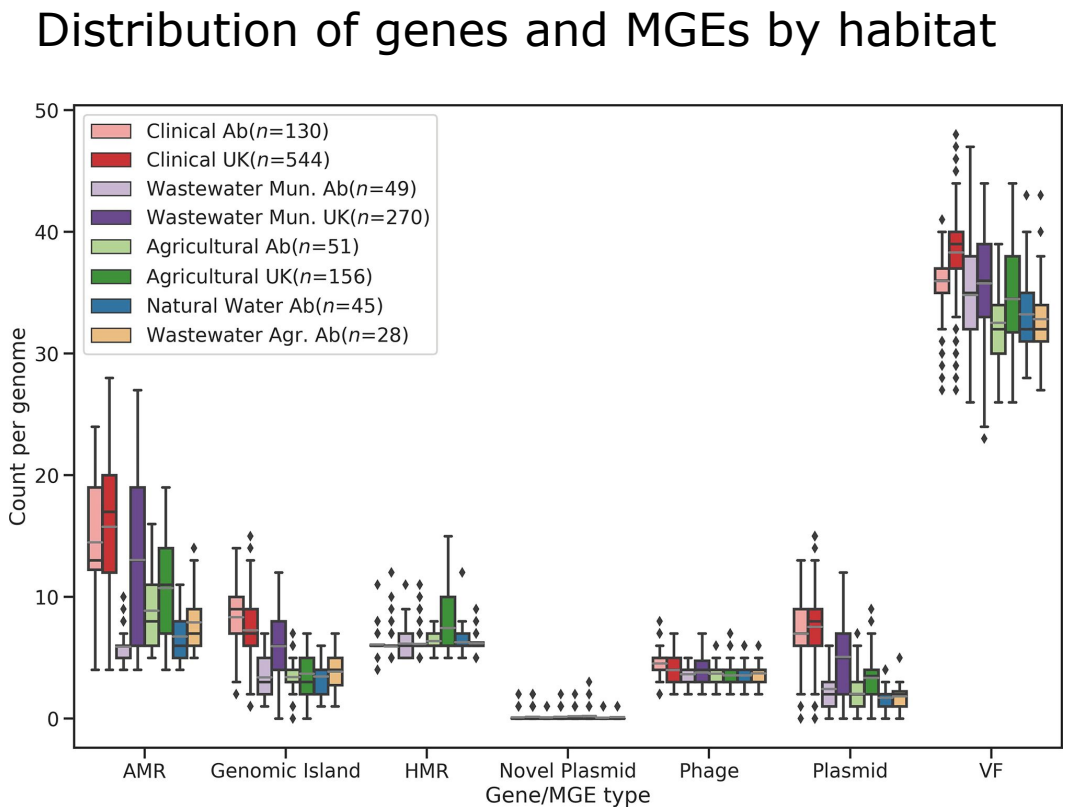
The big questions:

- What are the major distributional patterns of “things” (genes, MGEs) in my set of genomes?
- Where is there significant evidence of LGT / recombination?
- What do these transmission events associate with?
  - Geographic proximity?
  - Habitat?
  - Phylogenetic relatedness?
  - Antimicrobial usage?

ARETE Overview

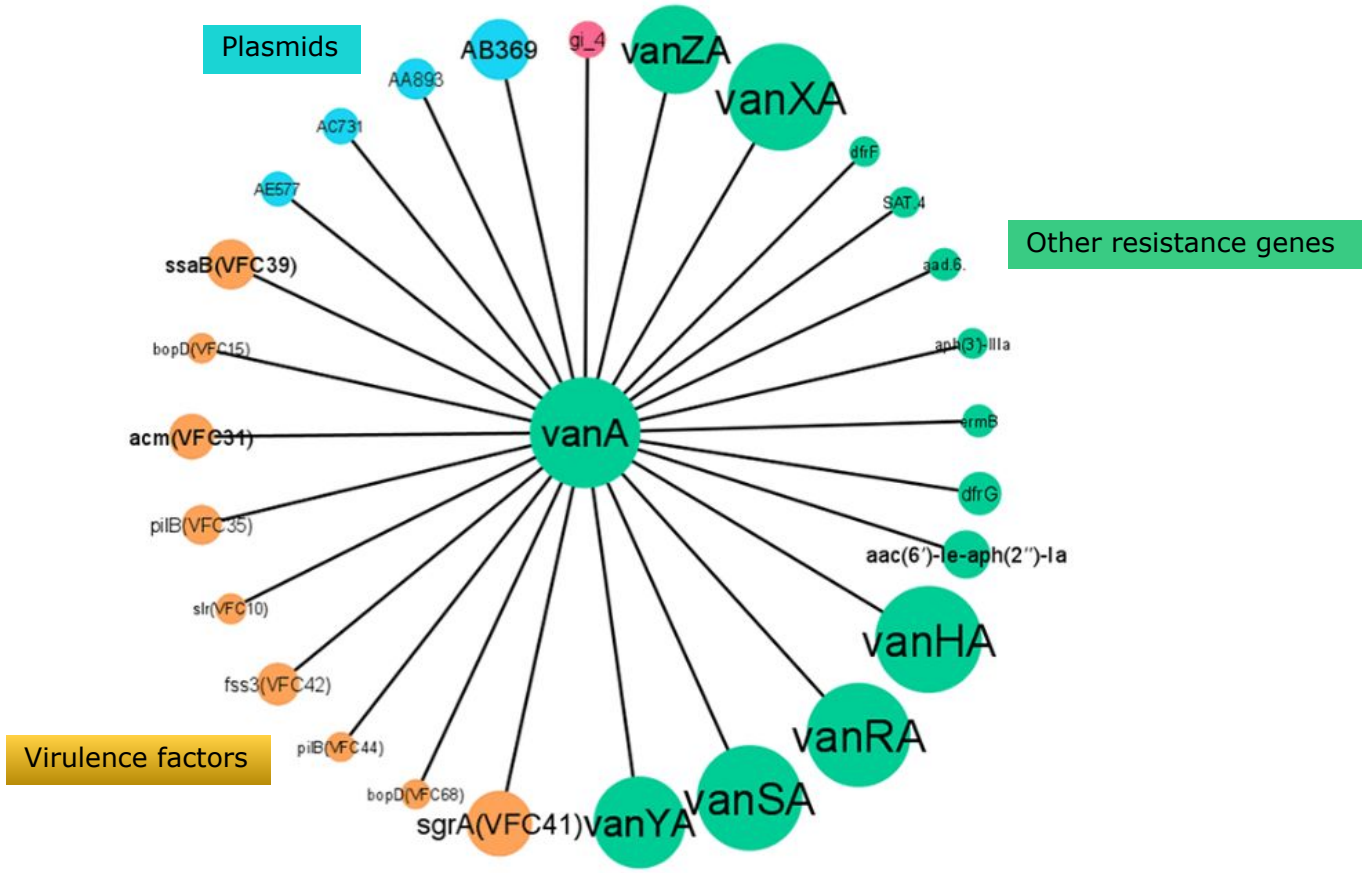


Example: an investigation of AMR in *Enterococcus faecium*

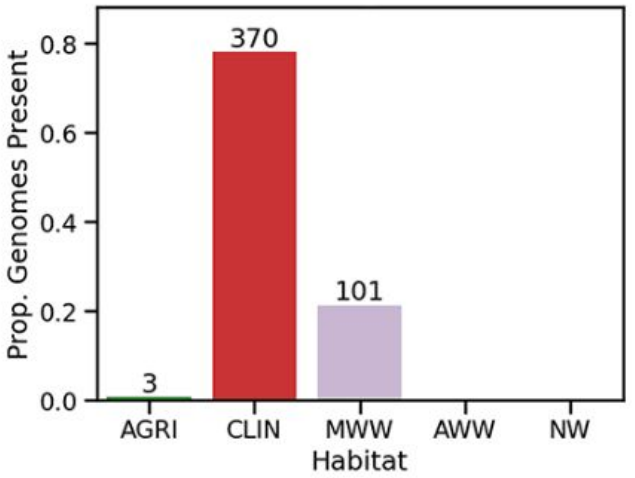


Gray, Sanderson et al. (2022) *Microbial Genomics*

Vancomycin-resistance genes are associated with…



And almost completely restricted to clinical and wastewater samples



Gray, Sanderson et al. (2022) *Microbial Genomics*

We are on a Coffee Break & Networking Session

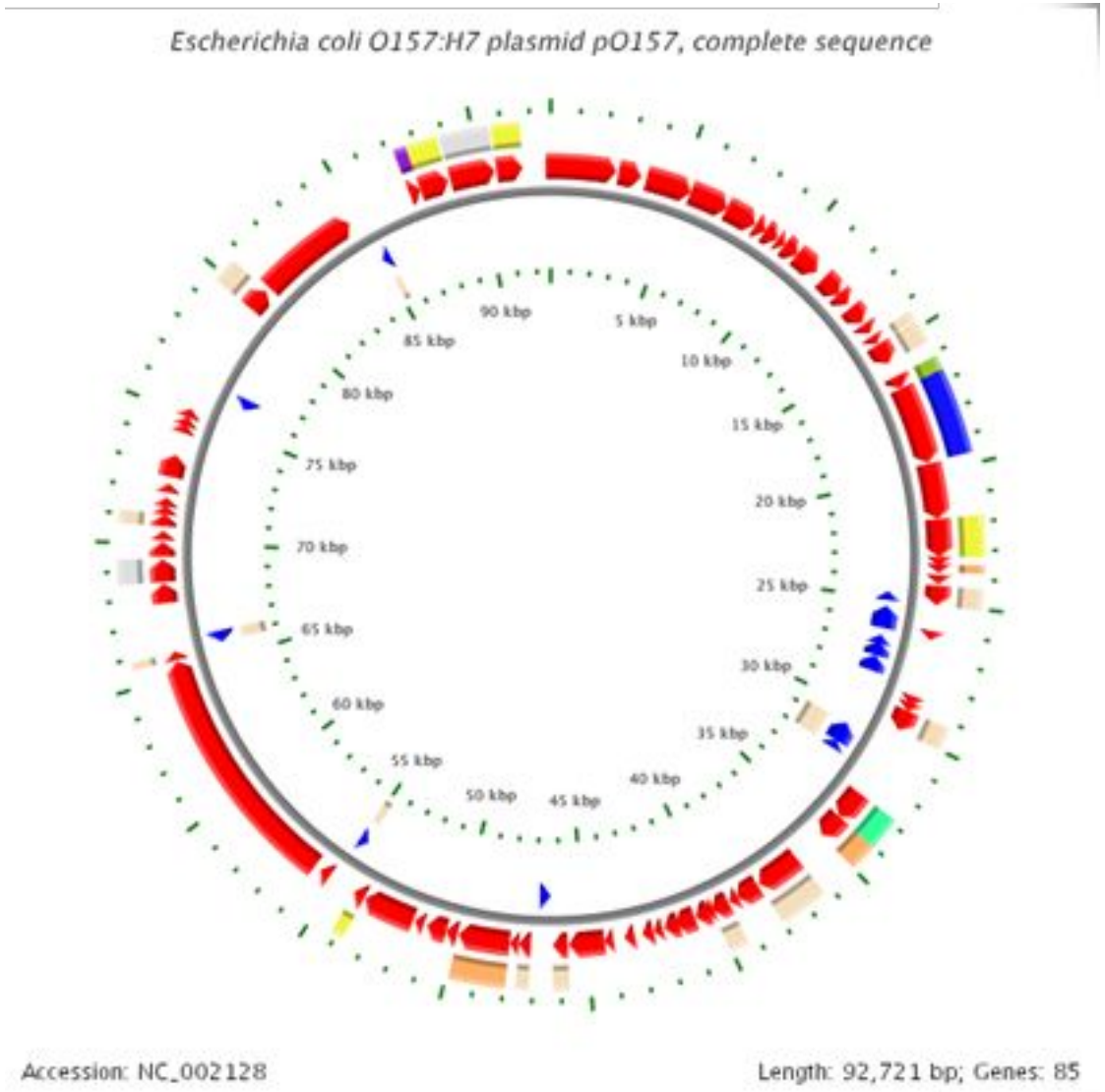
Workshop Sponsors:

Canadian Centre for Computational Genomics

Additional slides

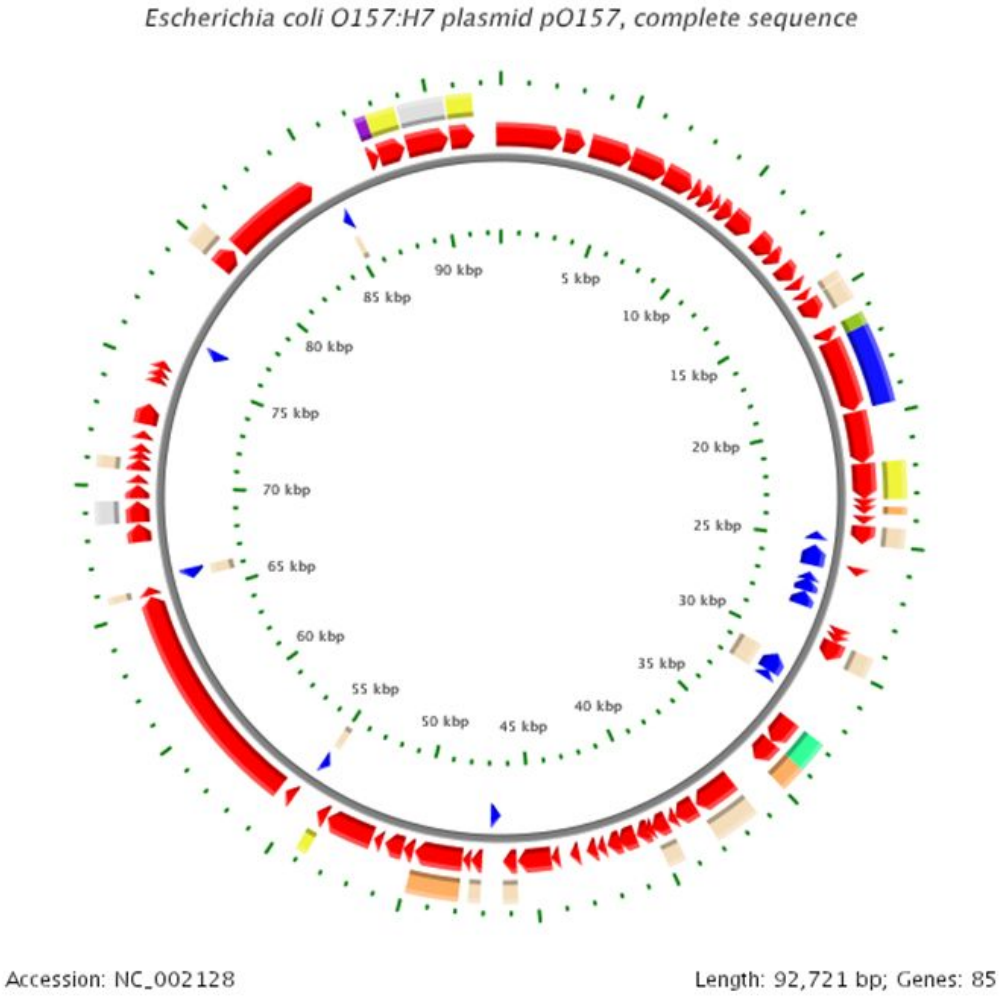
Plasmids

- Plasmids are smaller independent circular or linear DNA molecules
- Can vary in size from ~1 kpb to > 1 Mbp
- Prokaryotes can have multiple plasmid compositions.
- Plasmids can exist in multiple copy numbers.
- Plasmids can be exchanged through horizontal gene transfer (mobilome).
- Genes carried by plasmids provide selective advantage in a given environmental state, e.g. antibiotic resistance, pathogenicity, disinfectant resistance.



Plasmid Characteristics

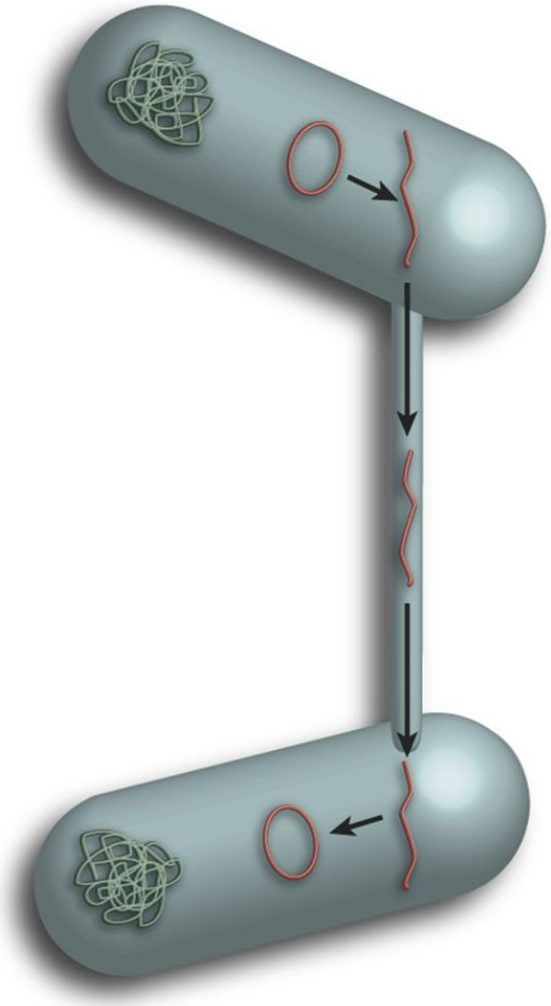
- One characteristic of plasmids is the ability to replicate independently from the chromosome, thus they contain their own origin of replication (*oriV*) and replication (*rep*) genes typically close by.
- Another characteristic of plasmids is regulation of copy number. This is determined mostly by the ori region and copy regulation (*cop*) genes.
- Plasmids avoid being lost from dividing cells by carrying partitioning functions which ensure at least one copy of the plasmid segregates into each daughter cell. The functions involved in partitioning are called *par* functions.
-



Plasmid Characteristics

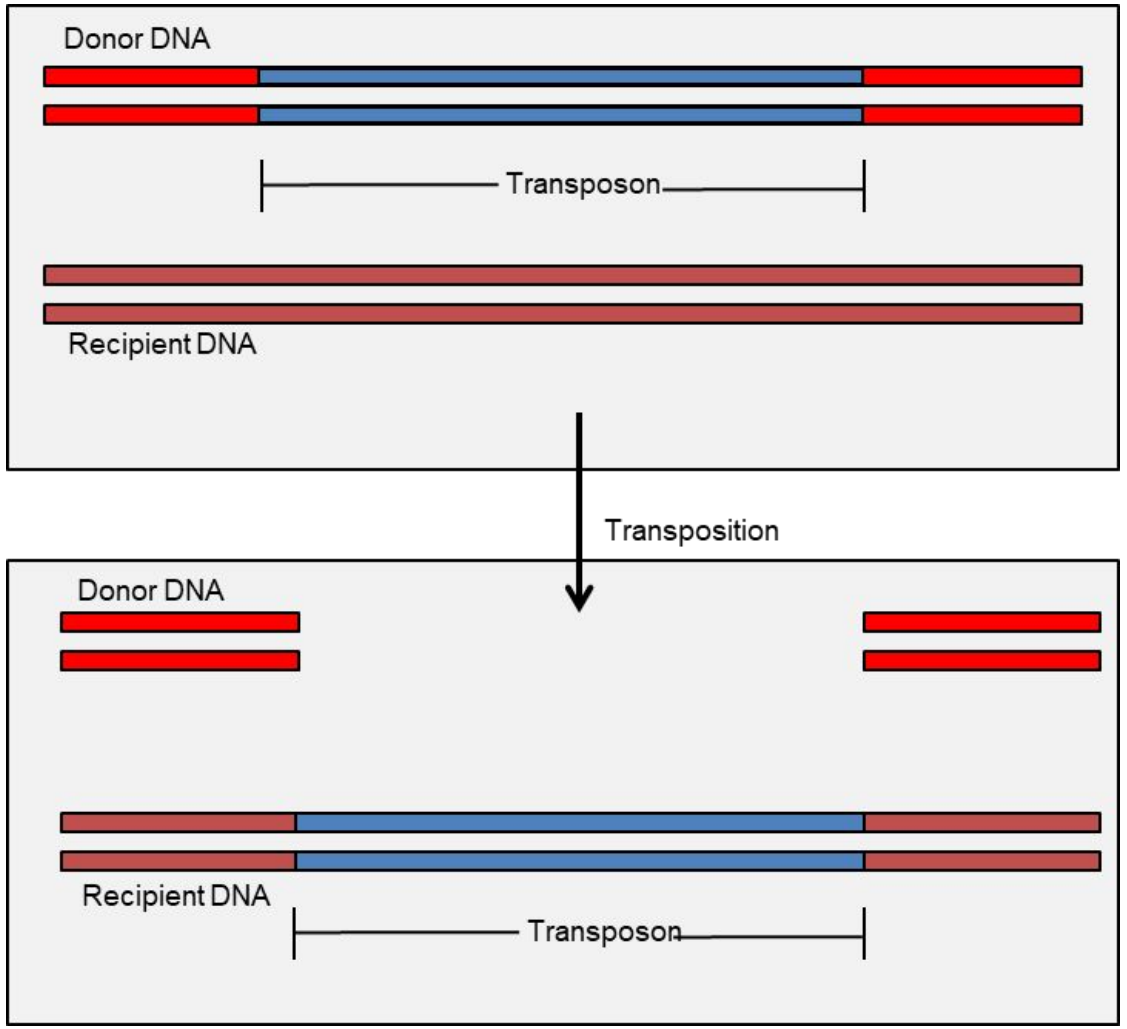
Conjugation

- Many plasmids have the ability to transfer themselves from one cell to another in a process called conjugation.
- Most naturally occurring plasmids are either self-transmissible or mobilizable.
- Self transmissible plasmids encode all the functions they need to transfer from donor to recipient. The genes required for transfer are called *tra* genes. In addition to the *tra* genes, a cis-acting *oriT* site is required.
- *tra* genes can be divided into two components: the *Dtr* component ("DNA transfer and replication") and the *Mpf* component ("mating pair formation").
- The *Dtr* component includes the relaxase and the primase. Many plasmids also encode a helicase
- The *Mpf* component includes the pilus, the channel and the coupling proteins.



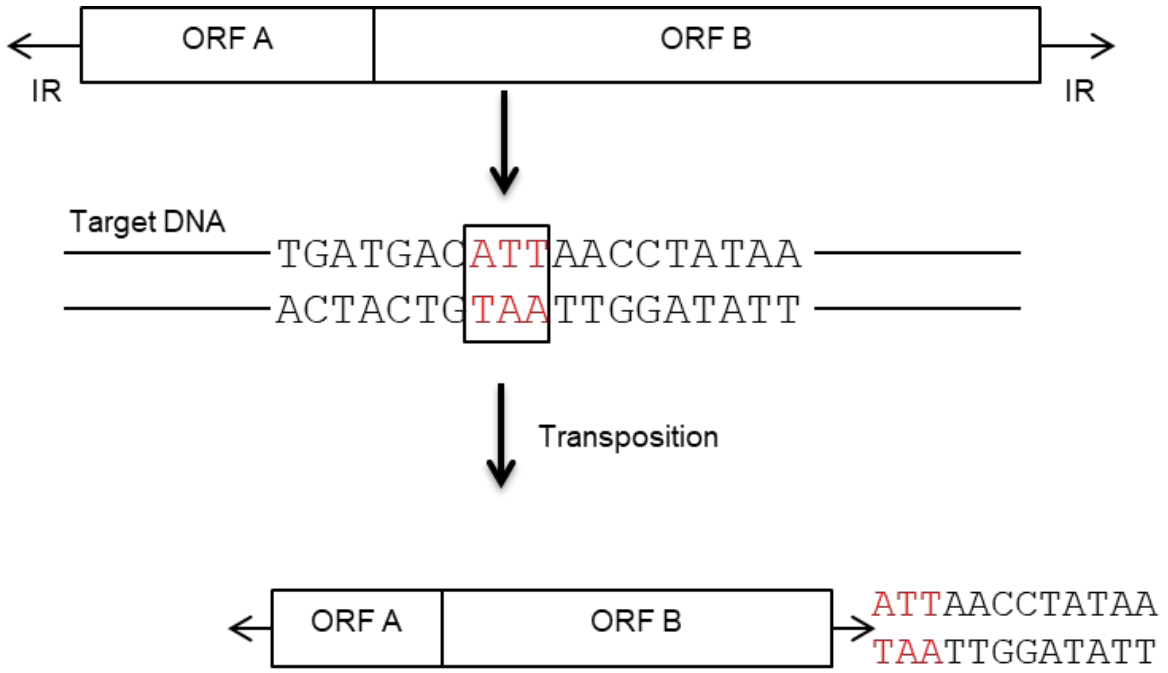
Transposons

- Transposons are DNA elements that can hop, or transpose from one place in DNA to another.
- Mediated by transposases.
- Transposons typically harbour their own transposases.
- Some transposons are ‘cut’ from one place (the donor DNA) and placed in another (the target DNA); this is called conservative or “cut and paste” transposition.
- Other transposons are copied from one place and then inserted elsewhere; this is referred to as replicative or “copy and paste” transposition.
- Can insert between genes (thus inactivating the gene)
- Transposition is tightly regulated
- Occurrence is once every 10^3 to 10^8 cell divisions



Transposition

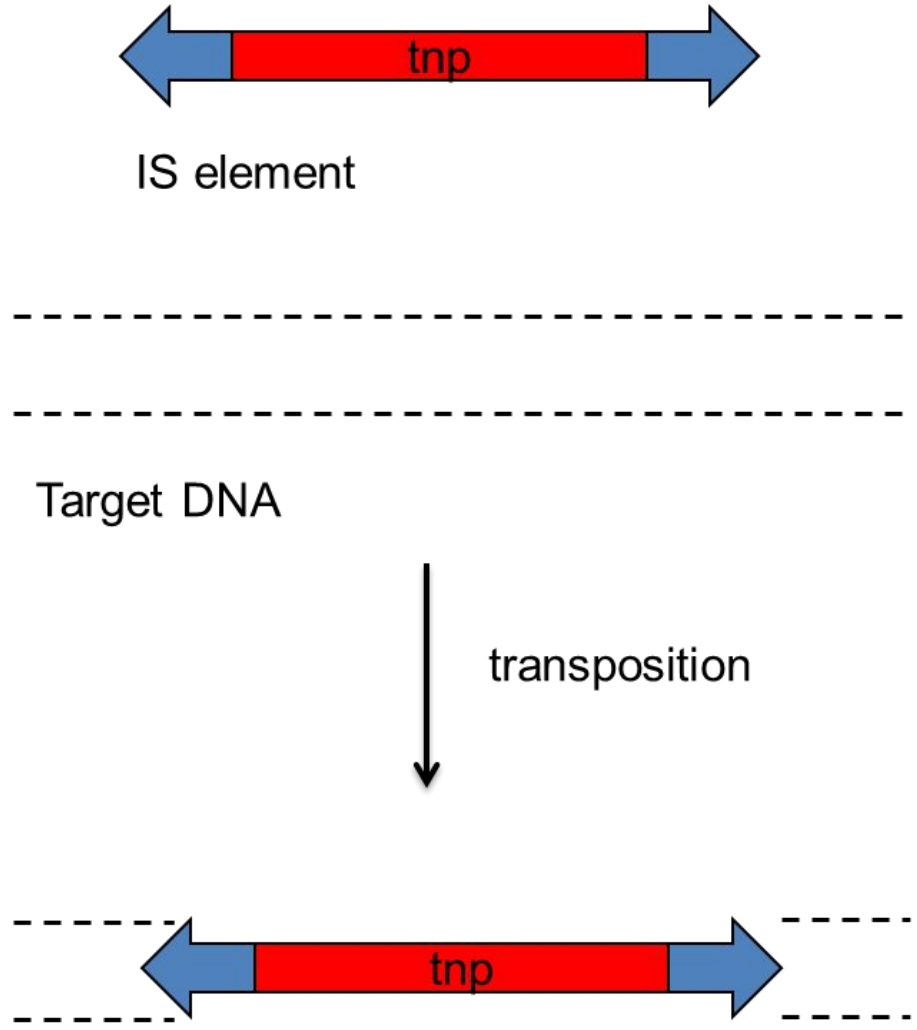
- Almost all bacterial transposons are DNA transposons.
- All known bacterial transposons contain inverted repeats (IRs).
- Another common feature is the presence of short direct repeats in the target DNA.
- Direct repeats have the same or nearly the same 5’-3’ sequence of nucleotides on the same strand (the sequence is duplicated during transposition).



Types of Transposons

- Insertion sequence elements are transposons that encode little more than the transposases that promote their transposition.
- They carry no selectable genes, their only detectable phenotype is that they inactivate the genes that they insert.
- The ISFinder website can be used to detect IS elements:

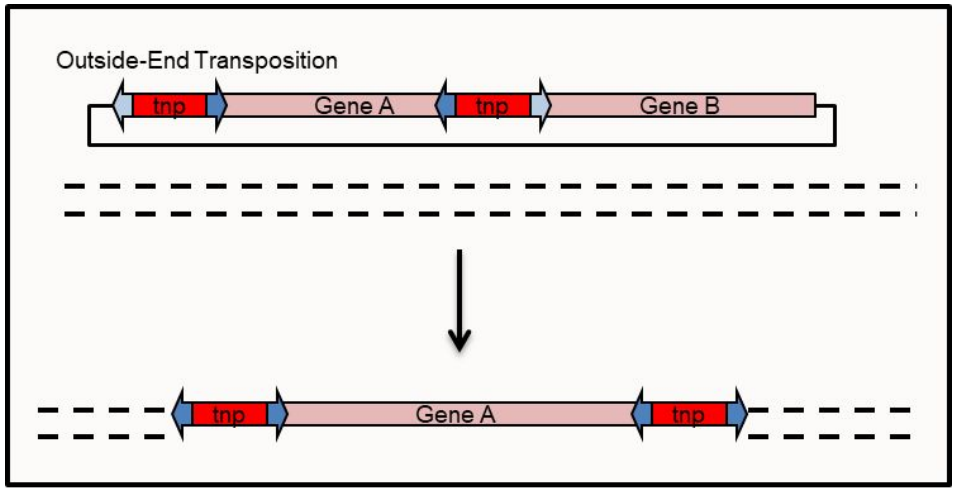
  http://www-is.biotoul.fr/

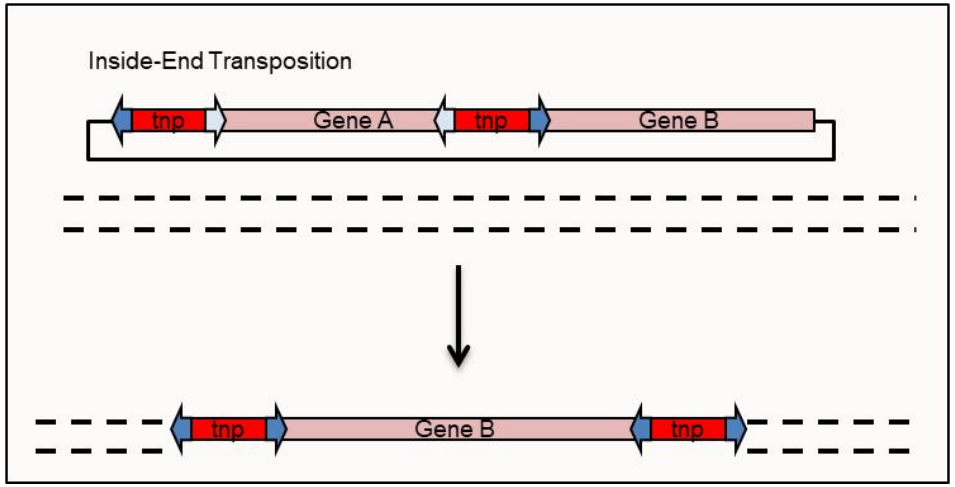


Types of Transposons

Composite Transposons

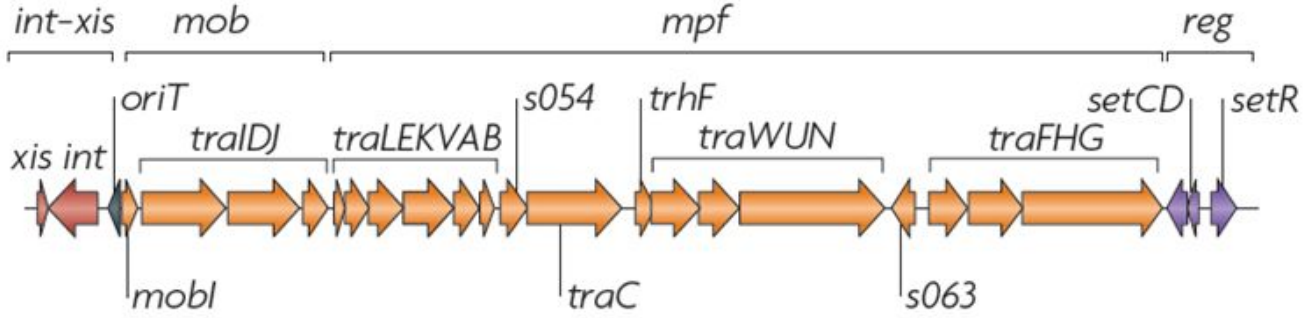
- Sometimes two IS elements form a larger transposon.
- Two IS elements can transpose any DNA between them.
- If the Donor DNA is circular, composite transposition can occur at the outside-end or inside-end, generating two different transposons.

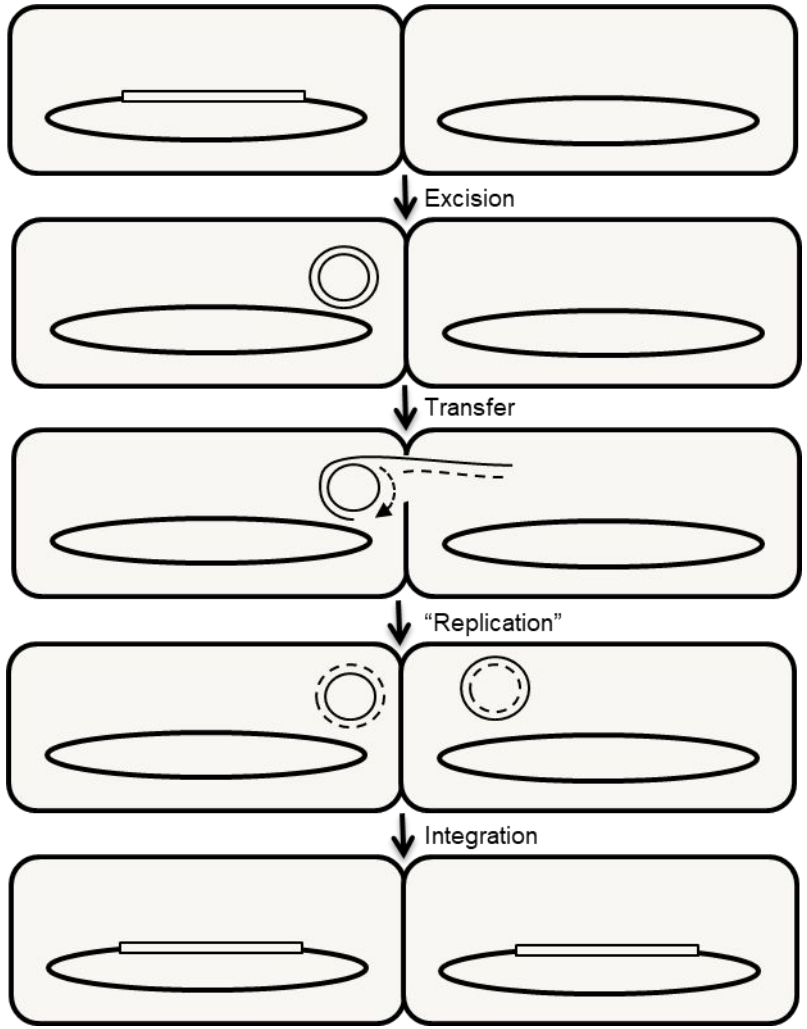




Conjugative Transposons

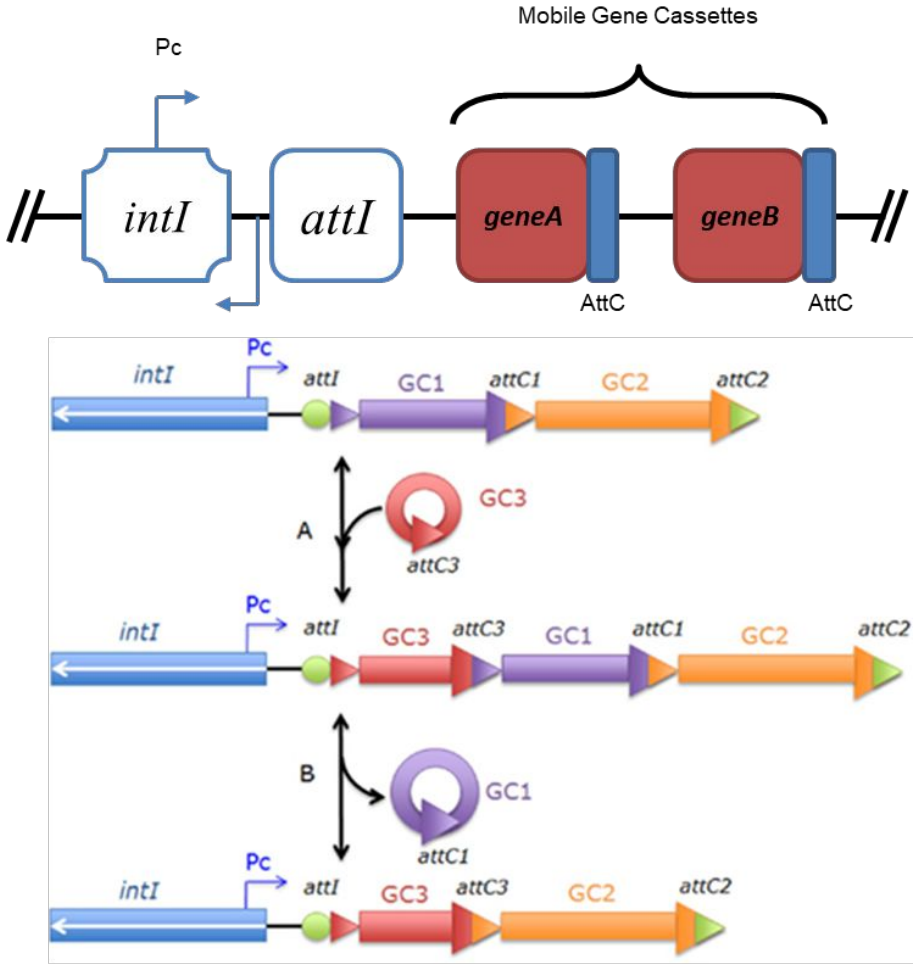
- Conjugative transposons are transposons combined with transfer functions such as those of a self-transmissible plasmid. Also called integrative and conjugative elements (ICEs).
- In order to transpose from the DNA of one cell to the DNA of another the DNA must first excise itself from the donor DNA, then be replicated and transferred into a target cell, then integrated. Because they have to encode the genes for all these functions ICEs are quite large.





Integrons

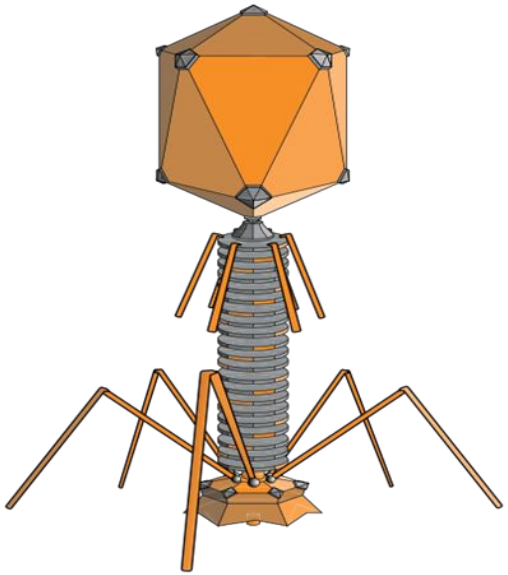
- Integrons contain an attI site into which multiple 600-1500 bp segments of DNA, called gene cassettes can be inserted by an integrase enzyme whose intI gene is adjacent to attI.
- Gene cassettes typically have a single ORF and a attC site beyond the 3’ terminal site.
- The integron cannot replicate autonomously; it must be harboured by a larger replicon (plasmid, chromosome).
- Pi is the promoter for integrase expression
- Pc is the promoter for cassette genes expression
- The integrase can excise gene cassettes and insert them at another attI site. It can also reposition a cassette in the integron.
- ACID can be used to annotate integrons: PubMed PMID: 19383137



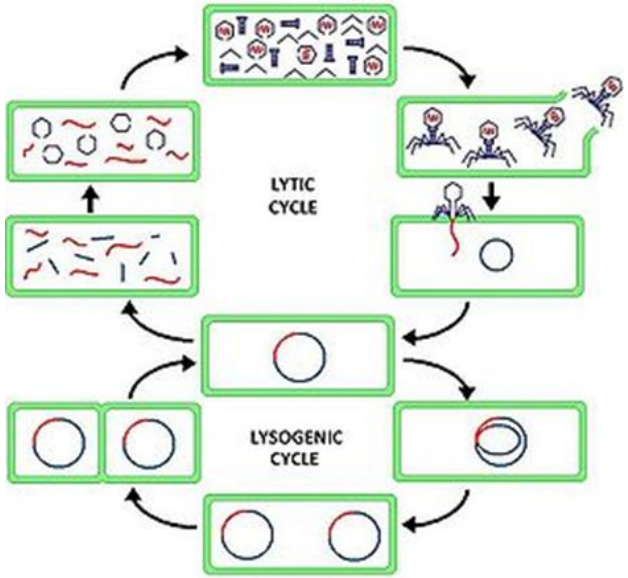
Front. Microbio., 09 April 2012 | doi: 10.3389/fmicb.2012.00119

Bacteriophage

- A bacteriophage, or phage is a virus that infects bacteria.
- Phage can have DNA or RNA genomes.
- Bacteriophages may have a lytic cycle or a lysogenic cycle.
- In lytic phage, the viral DNA exists as a separate molecule within the bacterial cell, and replicates separately from the host bacterial DNA.
- With lytic phages, bacterial cells are broken open (lysed) and destroyed after immediate replication of the virion.
- In contrast, the lysogenic cycle does not result in immediate lysing of the host cell. Those phages are able to undergo lysogeny are known as temperate phages. Their viral genome will integrate with host DNA (called a prophage) and replicate along with it fairly harmlessly, or may even become established as a plasmid.



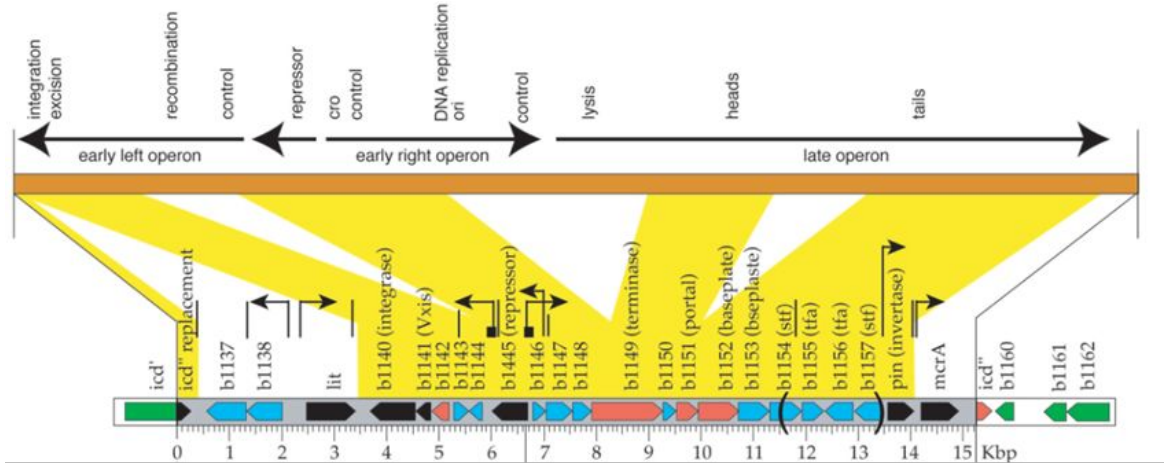
http://en.wikipedia.org/wiki/File:PhageExterior.svg

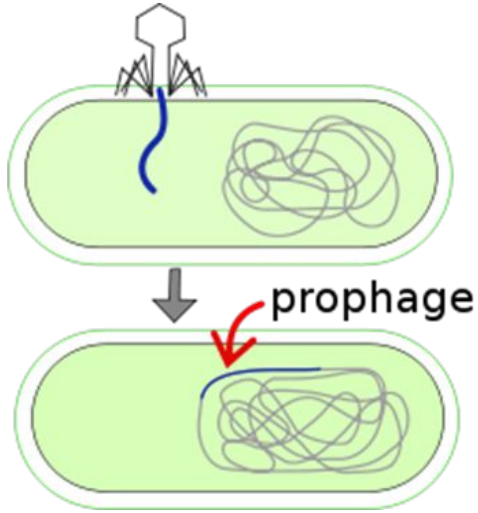


http://en.wikipedia.org/wiki/File:Phage2.JPG

Prophage

- A prophage is a phage (viral) genome inserted and integrated into the circular bacterial DNA chromosome or as an extrachromosomal plasmid.
- Prophages can constitute as much as 10-20% of a bacterium's genome.
- Prophages often confer pathogenic phenotypes, e.g. prophage-encoded toxins, bacterial cell surface alterations, or resistance to a host's immune system.
- Prophage integration into the host genome can inactivate or alter the expression of host genes
- Prophages can also alter the phenotype of bacteria at the population level by facilitating the spread of favorable genes.

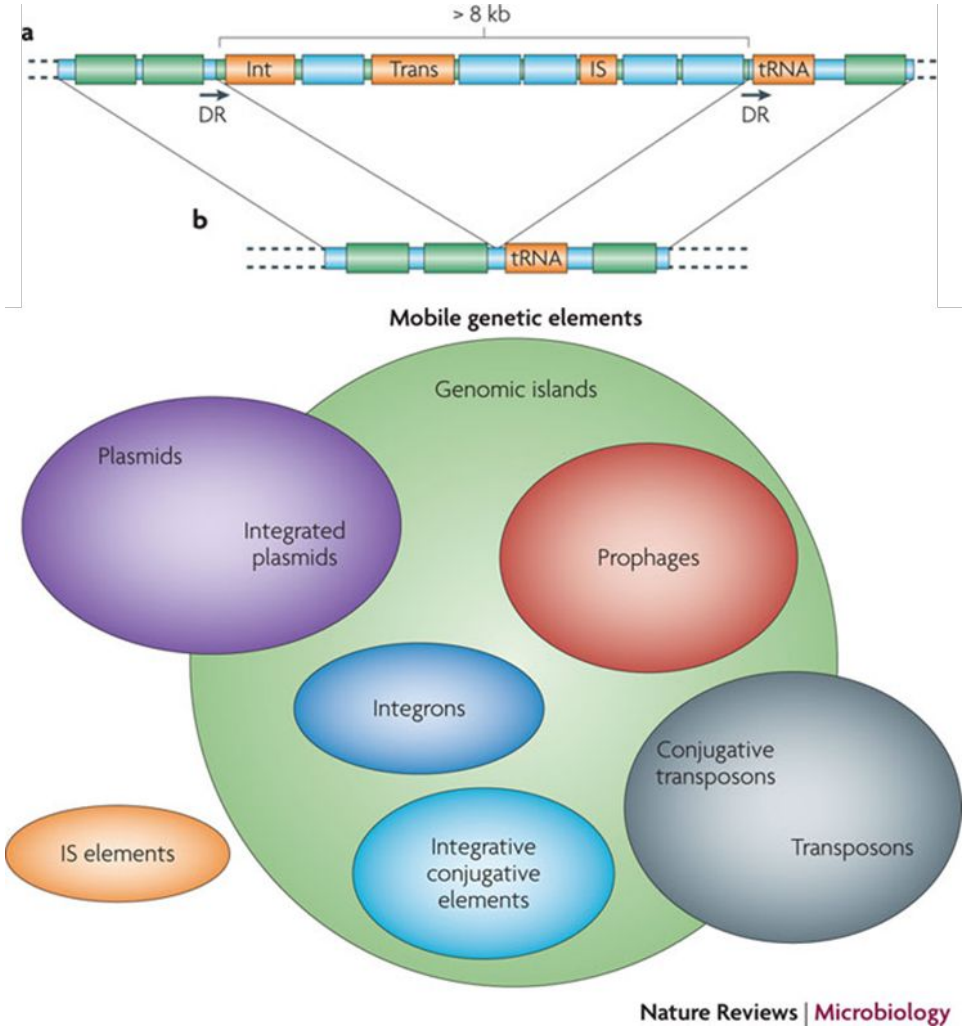




http://en.wikipedia.org/wiki/Prophage

Genomic Islands

- Genomic Islands (GIs) are clusters of genes in genomes that show evidence of horizontal origins.
- Transferred through horizontal gene transfer events such as transfer by large mobile elements such as prophages, integrons, conjugative transposons and integrative conjugative elements.



Morgan G. I. Langille, William W. L. Hsiao & Fiona S. L. Brinkman *Nature Reviews Microbiology* **8**, 373-382 (May 2010)

Genomic Islands

- Genomic islands are large regions of a genome that have evidence of horizontal origin.
- Often confer special properties on the bacteria that carry them
- Often precisely excise from the chromosome
- Not capable of autonomous replication, but are mobile
- Often carry genes for pathogenicity (pathogenicity islands).