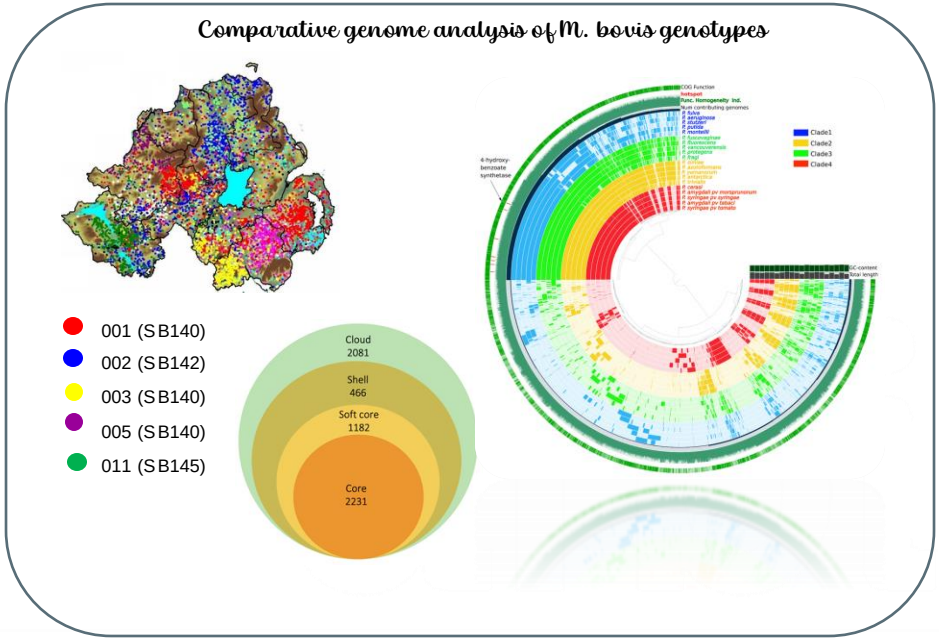
AGRI-FOOD
& BIOSCIENCES
INSTITUTE

Leading | Protecting | Enhancing

Pan-genomes of *Mycobacterium bovis* from infected livestock and wildlife.

AIMS2024 Meeting – ATU, Galway, Ireland



Purnika Damindi Ranasinghe

Bacteriology Branch, Veterinary Science Division, AFBI - Stormont

X: @purrd_mindi GitHub: @butterbee
LinkedIn: Purnika Ranasinghe

afbini.gov.uk

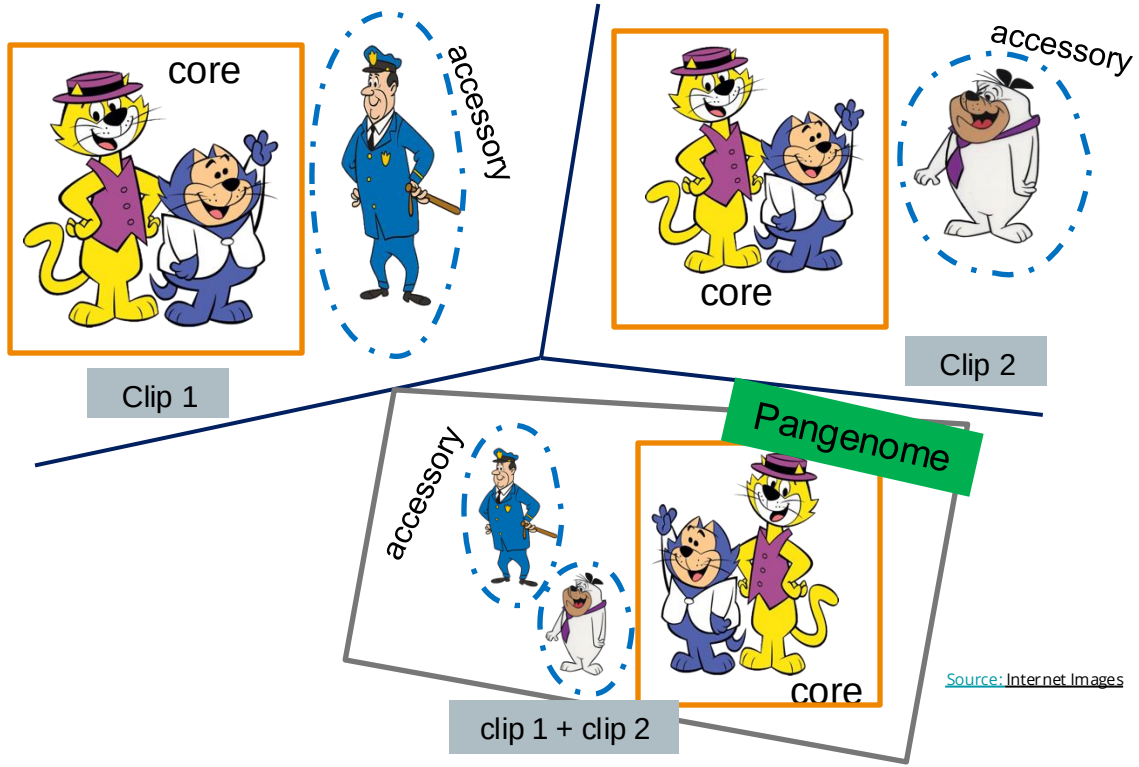
What is a Pan-genome ?

- **Pan-genome:**
Full complement (core & accessory) of genes within an organism.

- **Core-genes /genomes:**
the set of genes present in all members of a species.

- **Accessory genes/genomes:**
the set of genes absent in one or more genomes of a study group.

The *pan* means *everything* in ancient *Greek*.



Source: Internet Images

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

Pan-genome Analysis

The building block of pan-genomics is identifying **homologies among the genes (gene clusters)** that compose the genomes or are shared between organisms [1, 2].

- How pan-genome evolve?

A pan-genome evolves through gene gain, gene loss, and gene modification across different strains of a species.

➢ Such changes over time (gene gain, loss or modification) could lead to changes in the size of a pan-genome.

If the addition of new genomes increases the size of the pan-genome it is considered ***open (more to be discovered).*** if not ***closed.***

➢ Variation in gene content helps to understand the genetic diversity within a species [3].

➢ This knowledge is essential for investigating how pathogens evolve, spread and survive under different conditions.

Also, useful in molecular epi., particularly for genomes with low SNP diversity and in the absence of annotated reference genomes.

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

References:
1. Tettelin & Medini 2020. The pangenome: Diversity, dynamics and evolution of genomes
2. Tettelin et al., 2005 Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial pan-genome.
3. Bonnici and Chicco 2024 Seven quick tips for gene-focused computational pangenomic analysis

Mycobacterium bovis as a model organism

- *Mycobacterium bovis* is an animal-adapted member of the *Mycobacterium tuberculosis* complex (MTBC), which causes bovine tuberculosis in cattle and other mammalian species [1].
- *M. bovis* has a highly conserved and strictly clonal genome with little gene content variation and contributes to diversity through horizontal gene transfer (HGT) or recombination [1,2,3].
- In 2021, Reis and Cunha et al estimated an open pan-genome with a high percentage of accessory genes in European and African *M. bovis* genomes.
- A year Ceres et al. (2022) published their findings that *M. bovis* has a much smaller accessory genome (clonal evolution) than previously described and a closed pangenome with little gene content variation.

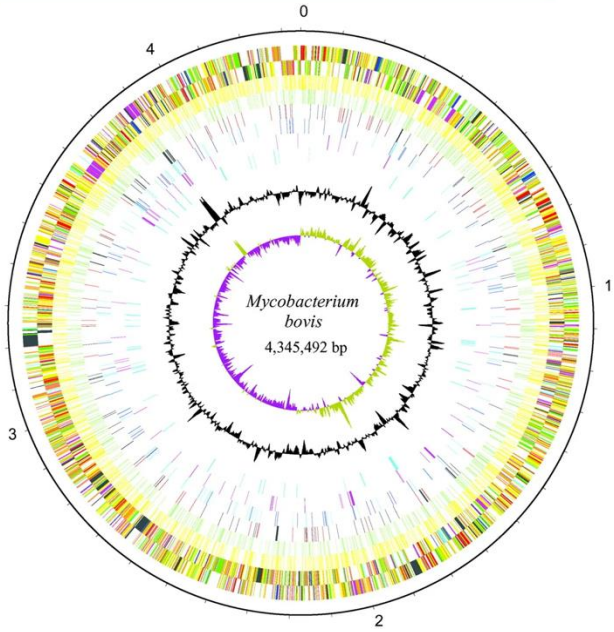


Fig 1: The complete genome sequence of *Mycobacterium bovis*.
Source: https://doi.org/10.1073/pnas.1130426100

References:
1. Ceres et al., 2022, A critical evaluation of *Mycobacterium bovis* pangenomics, with reference to its utility in outbreak investigation
2. Boritsch et al., 2016 Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria
3. Pantane et al., 2017 Patterns and processes of *Mycobacterium bovis* evolution revealed by phylogenomic analyses

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

Objectives

What is the pan-genome landscape of Northern/Irish *M. bovis* strains isolated from different geographical regions and hosts?

- to understand the strain-level genetic variation and phylogenetic relatedness through single nucleotide polymorphism (SNP) variants.
- to explore pan-genome architecture of clonally evolving *M. bovis* through comparative genome analysis.

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

Materials & Methods

Only high-quality *denovo* assemblies were selected.

- Study material: 1,001 *denovo* assembled genomes.

Origin: Northern Irish, Irish, South African and European (France)

Composition: Livestock and wildlife.

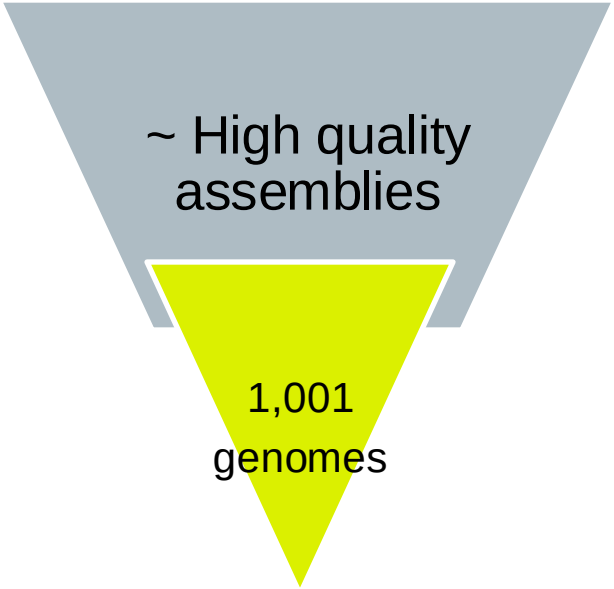
- NI & RoI: cattle (~850), badger (~100), deer (~20)
- African: buffalo (~7)
- European : badger, wild boar and cattle (~35)

Lineages: (most of the NI sub-lineages are spatially-localised)

- 11* different sub-lineages (NI and RoI) and SB0120 (others)

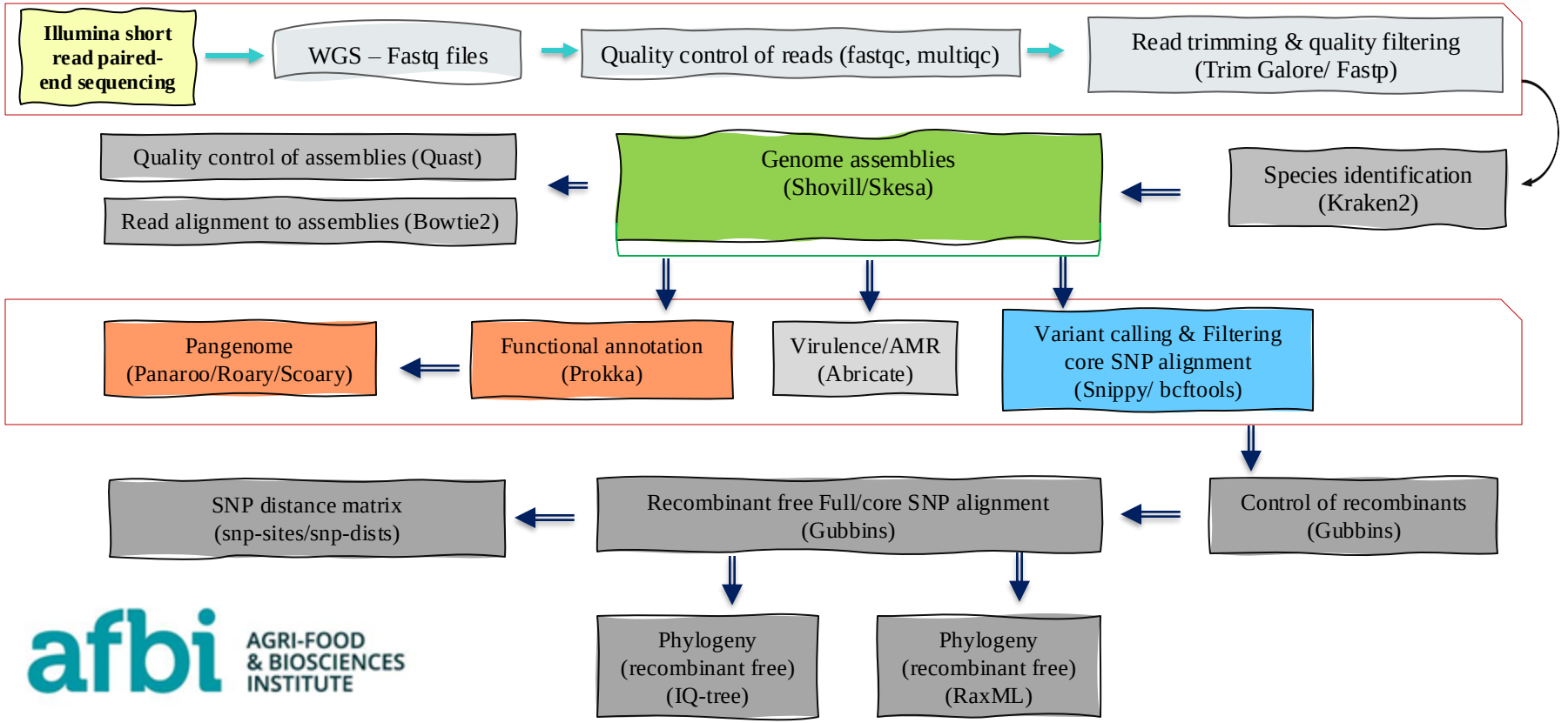
1.140, 2.142, 3.140, 4.140, 5.140, 6.623, 7.140, 9.273, 10.140 19.140 and 27.140

Reference genome: *M. bovis* AF2122/97 isolate



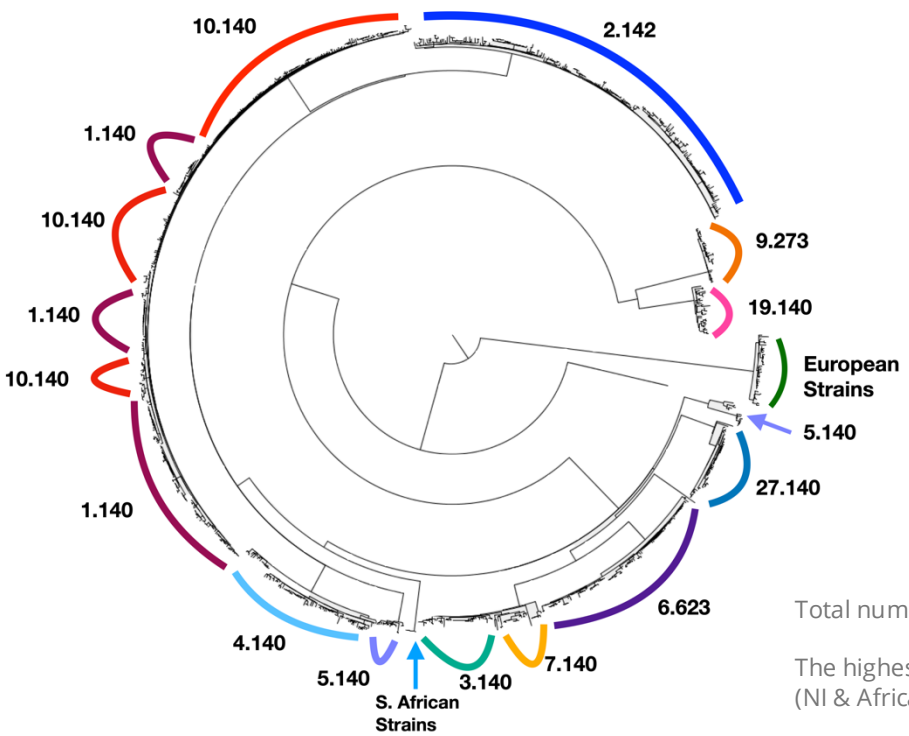
afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

AFBI – TB-RAVEN (Rapid Assembly and Variant Evaluation Network)



afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

Strain-level genetic variation explained by SNP-based phylogeny



- Core SNP phylogeny highlighted the strain-level genetic variation and phylogenetic relatedness of sub-lineages.
- Clades represented with corresponding sub-lineages. Multiple sub-clades have been identified within reflecting transmission events between hosts (SNP range: ~3 – 350).
- SB0140 is the most common type and it well-resolved into geographically localised sub-lineages e.g. 3.140, 4.140, 5,140 etc. (Skuce et. al 2010).
- However, the presence of low SNP diversity or similar SNP patterns could be a constraint in inferring transmission patterns.

Total number of SNP identified ~2,600 (in all) & 2,100 (in NI/RoI samples).

The highest SNP variants detected between European types and the other (NI & African) types.

Fig 2: SNP-based recombinant free maximum likelihood core phylogenetic tree for all 1,001 genomes.

Pangenome structure of *M. bovis*: what is the best method?

Tool	core_threshold/mode	core genes	soft core genes	shell genes	cloud genes	Total
	clustering%	100-99%	99-95%	95-15%	15-0%	
Roary	0.95	3,269	256	627	3,282	7,434
	1	2,395	1,171	513	2,752	6,831
Panaroo	moderate	3,922	89	224	338	4,573
	Strict	3,732	70	155	175	4,132
		100-95%		95-10%	10-0%	
Pirate	strict 0.95, 0.98	3,556		627	2,686	6,869
M. bovis 1,001 genomes analysed by three different pangenome analysis tools						

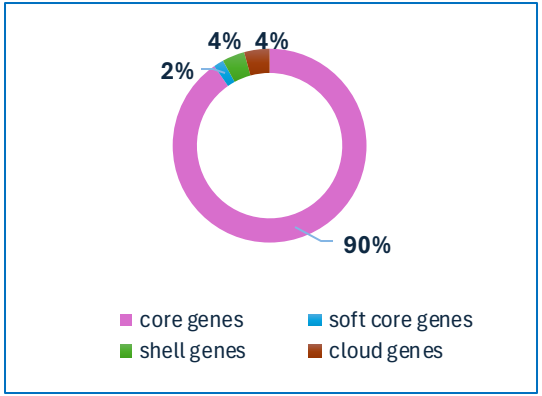
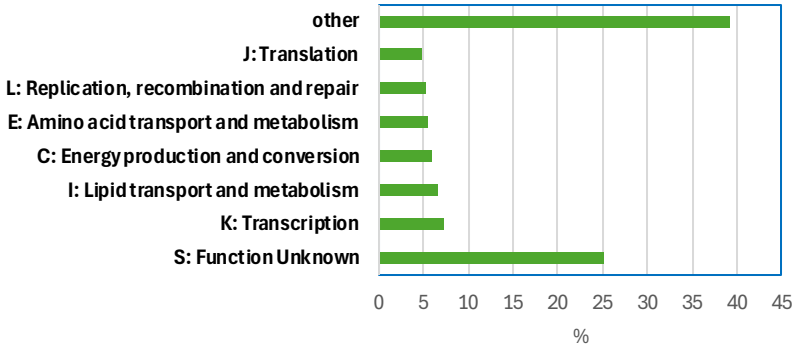


Fig 4: Pangenome analysis results produced by three different analysis tools (left) and summary of gene distribution by Panaroo (right) for all 1,001 genomes.

- Methodological Inconsistencies: characterization of gene fragments is challenging.
- **To consistent with latest literature, we chose Panaroo as our preferred method of analysis**.
- According to Panaroo: 90% of the total gene count accounts for core-genes.
- Smaller accessory genome evident of a closed pangenome (supporting clonal evolution of *M. bovis*).

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

Functional attributes and virulence landscape of *M. bovis* genomes



gene	count (out of 1,001)	function
espK	693	associated with ESX-1 virulence system. Involved in secretion of virulence factors that manipulate the host response. EspK gene is missing in BCG vaccine strain.
espB	832	EspK interacts with EspB to enable fucntion of secterion system.
esxM	874	esxM is a part of ESX-5 sectertion system associated with transport of virulence factors.
esxN	912	esxN is a part of ESX-5 sectertion system associated with transport of virulence factors.
PPE4	886	virulence factor - modulate host-pathgen inetraction.
eccB1	928	associted proteins in EsX-1 virulence system.
eccC3	984	associted proteins in EsX-1 virulence system.
eccD3	988	associted proteins in EsX-1 virulence system.
mbtA	988	gene code for protein mycobactin in mycobactin biosyntheis pathway invoves in iron acquisition.
mbtB	986	gene code for protein mycobactin in mycobactin biosyntheis pathway invoves in iron acquisition.
mbtE	978	gene code for protein mycobactin in mycobactin biosyntheis pathway invoves in iron acquisition.
irtB	981	ATP-binding casstte (ABC) transporters associated with Iron acquisition.

Fig 5: COG Functional annotation of core-genes (left) and virulence genes showing differential distribution across selected 1,001 genomes (right)

- Core genes are linked to biological processes and cellular structure/function.

- Within accessory genes, 12 of virulence-linked genes seems showing differential distribution across genomes.
- Majority of the genes are linked in modulating host-pathogen interactions. eg. * ESX-1 secretion system.

* Mycobacteria posses 5 known Type VII secretion systems, which enhances their transmissibility between mammalian hosts.

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

Focus group study:

In total, 266 genomes from sub-lineage 1.1.40 & 10.140 selected based on the distribution of virulence genes across genomes.

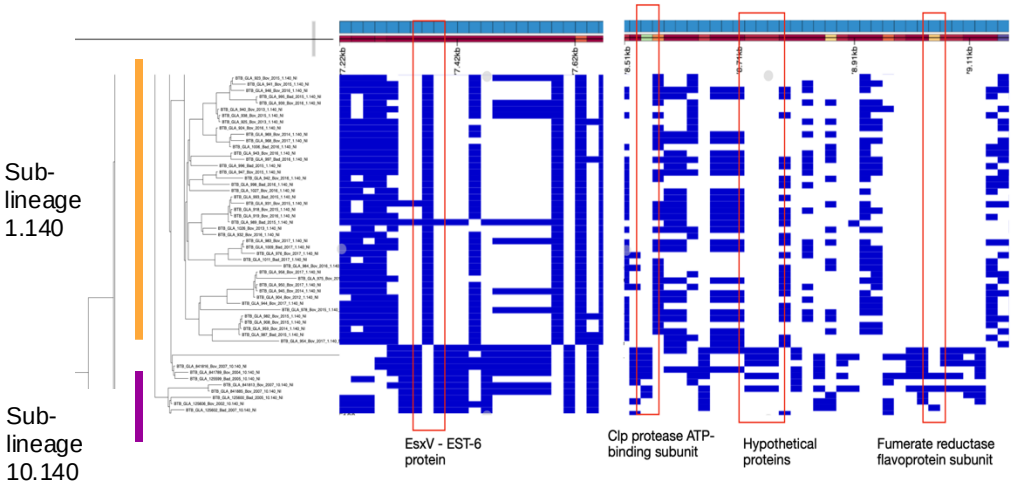


Fig 6: Gene presence and absence pattern of accessory genes highlighting gene deletions in multiple genomes.

- No correlation was observed between phylogeny by core SNPs variation and core-gene variation.

Assumption: If the pan-genome evolved clonally there should be a correlation between core phylogenetic groups and accessory gene content patterns.

- Highlight the need for critical validation on accessory gene classification.
- Yet, 1.140 and 10.140 sub-lineage show the deletion of multiple genes associated with functions like energy production.
- Such deletion could be important in shaping *M. bovis* pangenome and may lead to different functional variations within strains.

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

Take Home Messages:

- Core SNP phylogeny: major clades clustered based on sub-lineages reflecting strain-level genetic variation and phylogenetic relatedness of *M.bovis* sub-lineages.

- Our results revealed the limited acquisition of new genes (90% core genes) in *M. bovis* pangenome.

Even smaller accessory gene counts compared to Ceres et al study.

- Deletions observed in functional gene clusters could be a key mechanism in shaping *M. bovis* pangenome and may lead to different functional variations within strains.

- Pan-genome data are complex, different tools produce inconsistent results. There is no one-size fit for all solution.

- Future direction: **This is a study in progress**. Further analysis on accessory gene identification is needed to understand impact of pangenome on creating possible functional variations within a strain.

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

Acknowledgement:

Department of Agriculture, Environment and Rural Affairs

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE

E & I Projects and Funding

AFBI – VSD team in sampling, media, micro, molecular and genomics.

QUB-Kelvin 2 HPC facility.

AFBI Team : Visit our posters.

This presentation is available at

afbi AGRI-FOOD & BIOSCIENCES INSTITUTE