Phylogenetic Analysis and Epidemiological Investigation of *Brucella* spp. in a One Health Context

Taxonomic analysis and typing

Ana Rita Rebelo, DTU Food anrire@food.dtu.dk 19 Jan 2026

Intended Learning Objectives

Specific objectives of this session:

1. Learn about taxonomic analysis and typing of *Brucella* spp.
2. Learn how to perform those analyses using WGS-based methods

Outline

This session consists of the following elements:

1. Introduction to taxonomic analysis and typing
2. Explanation of taxonomy of *Brucella* spp.
3. General overview of methods for identification of *Brucella* spp.
4. Detailed explanation of WGS-based methods for identification and typing of *Brucella* spp.

Taxonomy of *Brucella* spp. [1-3]

Gram-negative, aerobic, non-motile and non-spore-forming coccobacilli

Scientific classification of *Brucella* spp. (taxonomic ranks):

Domain Bacteria → Kingdom Pseudomonadati → Phylum Pseudomonadota/Proteobacteria → Class Alphaproteobacteria → Order Hyphomicrobiales/Rhizobiales → Family Brucellaceae → Genus *Brucella*

Originally defined according to the host / manifestation:

- *B. melitensis* (David Bruce, 1887) → in the British military in Malta
- *B. abortus* (Bang, 1906) → causing abortions in cattle
- *B. suis* (Traum, 1914 / Huddleson, 1929) → causing abortion in pigs
- *B. ovis* (Buddle, 1956), *B. canis* (Carmichael, 1968) → sheep and dogs, respectively

Taxonomy of *Brucella* spp. [1-3]

Late 1900's → **Attempts to systematize** *Brucella* taxonomy by early molecular methods, such as:

- Restriction fragment length polymorphism
- Pulsed-field gel electrophoresis
- Insertion sequence based typing
- DNA-DNA hybridization

Since the 2000's → Progress in taxonomic differentiation using **advanced molecular methods**:

- 16S rRNA sequencing
- Sequencing of *omp2* locus / IS*711* / other loci
- Multilocus sequencing analysis (MLSA)
- Multilocus Variable Number of Tandem Repeat Analysis (MLVA)
- **Whole-genome sequencing (WGS)**

Taxonomy of *Brucella* spp. [1-3]

Table 1. Currently recognised and potential *Brucella* species.

Species	Biovars	Major natural host(s)	Zoonotic
'Classical' Species			
B. abortus	1–6, 9[a]	Bovine *Cows*	++
B. melitensis	1–3	Ovine, Caprine *Sheep, goats*	++
B. suis	1–5	Porcine, Leporine *Pigs, hares* Rangerifine (bv 4) *Reindeer* Rodents (bv 5)	++ (not bv 2)
B. canis	–	Canine *Dogs*	+
B. ovis	–	Ovine *Sheep*	–
B. neotomae	–	Desert Wood Rat	–[b]

Whatmore 2021 (adapted)

++ 6 more recently described species
++ 6 awaiting taxonomic placement

Taxonomy of *Brucella* spp. [1-3]

Biotyping *Brucella* spp.

→ **Colony morphology** (dependent on outer membrane LPS structure)

Most species: 'smooth' phenotype
- Presence of a complete LPS (Lipid A + core oligosaccharide + O-side chain polysaccharide).
- 'Rough' mutants can occur, especially following repeated laboratory subculture.

B. ovis and *B canis*: 'rough' phenotype
- Do not contain the O-side chains.
- Not highly pathogenic for humans.

→ **Biovars**

Reflect variations in metabolism within the same species.
Examples: agglutination with monospecific antisera (A/M), growth on dye-containing media, CO2 requirement, hydrogen sulphide production, …

Identification and typing of *Brucella* spp.

Examples of methods for identification and typing of *Brucella* spp. (excluding WGS)

→ Phenotypic methods

Selective media - Examples: Brucella ovis medium; Ewalt's; Farrell's; Kuzdas and Morse; BruSIC

API Strip 20E from Biomérieux - Commercial system with small-scale biochemical tests

Biotyping methods mentioned previously

→ MALDI-TOF MS

Matrix-assisted laser desorption/ionization–time of flight mass spectrometry

→ PCR methods

Multiplex PCR Bruceladder

Single-nucleotide polymorphisms (SNP) based identification (real time PCRs, multiplex PCRs)

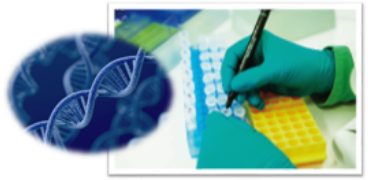
PCR for other targets

Whole-genome sequence-based tools

WGS workflow – single isolates

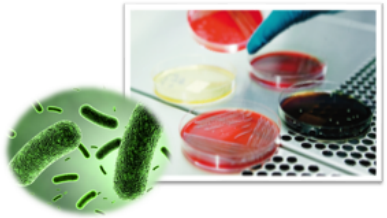
Represent the real sequence of nucleotides (A-T-C-G) in the genome

Fasta files (or fastq)



```
>sample_1
ATGATGCAGCATACTTCTGTGTGGTACCG
GTCAGTCCGTTTGTTCTTGTGGCGAGTGT
>sample_2
TTTCTTGACCGCGACCGCCAATCTTACCT
AAAATCAGCCAAACCTATCCCATCGCGGA
```

A sample is collected → The bacteria are isolated → The DNA is extracted → The DNA is sequenced → The data are processed → We investigate the genomes



Whole-genome sequence-based tools

Identification by alignment against references

The whole genome will be aligned against the reference genomes in the database.

The tool will show which reference genome in the database is best match to our query genome.

Example: KmerFinder

Whole-genome sequence-based tools [4]

KmerFinder for *Brucella* spp. → https://cge.food.dtu.dk/services/KmerFinder/

KmerFinder-3.2 Server - Results

KmerFinder 3.2 results:

Template	Num	Score	Expected	Template_length	Query_Coverage	Template_Coverage	Depth	tot_query_Coverage	tot_template_Coverage	tot_depth	q_value	p_value
NZ_CP026337.1 Brucella melitensis strain CIT43 chromosome, complete genome	20560	124591	0	124931	99.75	99.31	0.99	99.75	99.31	0.99	124590.72	1.0e-26

EXTENDED OUTPUT

Assembly	Num	Score	Expected	Template_length	Query_Coverage	Template_Coverage	Depth	tot_query_Coverage	tot_template_Coverage	tot_depth	q_value	p_value	Accession Number	Description	TAXID	Taxonomy	TAXID Species	Species
GCF_003516085.1	20560	124591	0	124931	99.75	99.31	0.99	99.75	99.31	0.99	124590.72	1.0e-26	NZ_CP026337.1	Brucella melitensis strain CIT43 chromosome, complete genome	29459	cellular organisms; Bacteria; Proteobacteria; Alphaproteobacteria; Hyphomicrobiales; Brucellaceae; Brucella/Ochrobactrum group; Brucella; Brucella melitensis	29459	Brucella melitensis

Input Files: *Animal_217.contigs.ge500.fasta*

RESULTS as text (tab separated) | Results as spa

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	# Assembly	Num	Score	Expected	Template_length	Query_Coverage	Template_Coverage	Depth	tot_query_Coverage	tot_template_Coverage	tot_depth	q_value	p_value	Accession Number	Description	TAXID	Taxonomy	TAXID Species	Species	
2	GCF_003516085.1	20560	124591	0	124931	99.75	99.31	0.99	99.75	99.31	0.99	124590.7	1.0e-26	NZ_CP026337.1	Brucella melitensis strain CIT43 chromosome, complete genome	29459	cellular organisms; Bacteria; Proteobacteria; Alphaproteobacteria; Hyphomicrobiales; Brucellaceae; Brucella/Ochrobactrum group; Brucella; Brucella melitensis	29459	Brucella melitensis	
3																				

Identification and typing of *Brucella* spp.

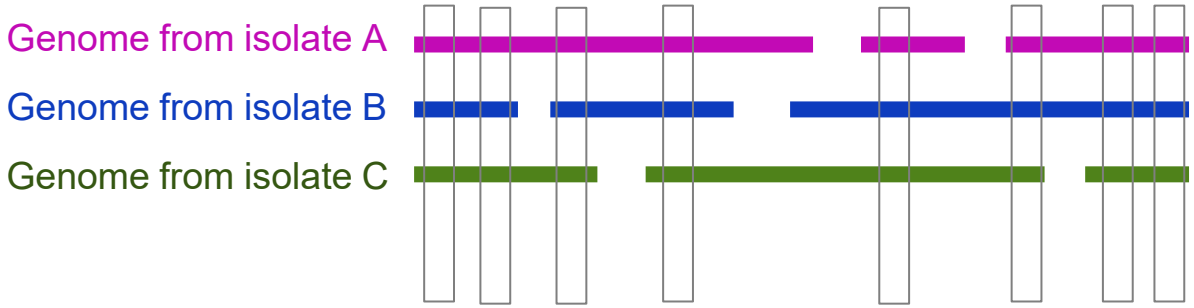
Typing with MLST

Multi-locus sequence typing

It is an **allelic profile scheme**.

It uses **conserved housekeeping genes** that are present in all isolates of the species.

For each bacterial species, the MLST scheme(s) contain specific and well defined genes (= loci).



Conserved loci that are present in all isolates from the species

Identification and typing of *Brucella* spp.

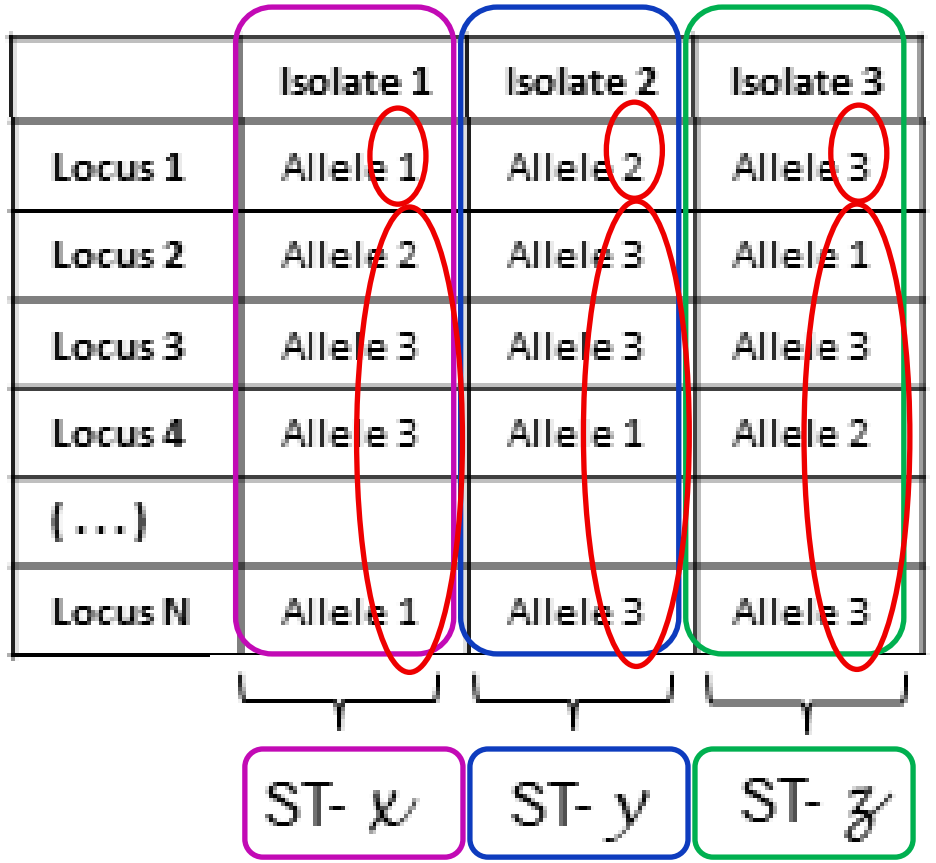
Typing with MLST

For each locus the specific genetic sequences can be different, because they accumulated mutations. Each specific sequence is called an allele.

Each bacterial isolate will have a **specific allele in each loci**.

The **exact combination** of alleles corresponds to a sequence type (ST).

If two isolates posses the same alleles they will belong to the same sequence type (ST). They are more **closely related** than isolates belonging to other STs.



Identification and typing of *Brucella* spp. [5-7]

MLST for *Brucella* spp. → https://pubmlst.org/organisms/brucella-spp

9 loci scheme: BruMLSA9 (Whatmore 2007)

gap, aroA, glk, dnaK, gyrB, trpE, cobQ, omp25, int-*hyp*
Total 4.396 bp
At development 27 STs, currently 124 profiles defined (as of 13 Jan 2026)

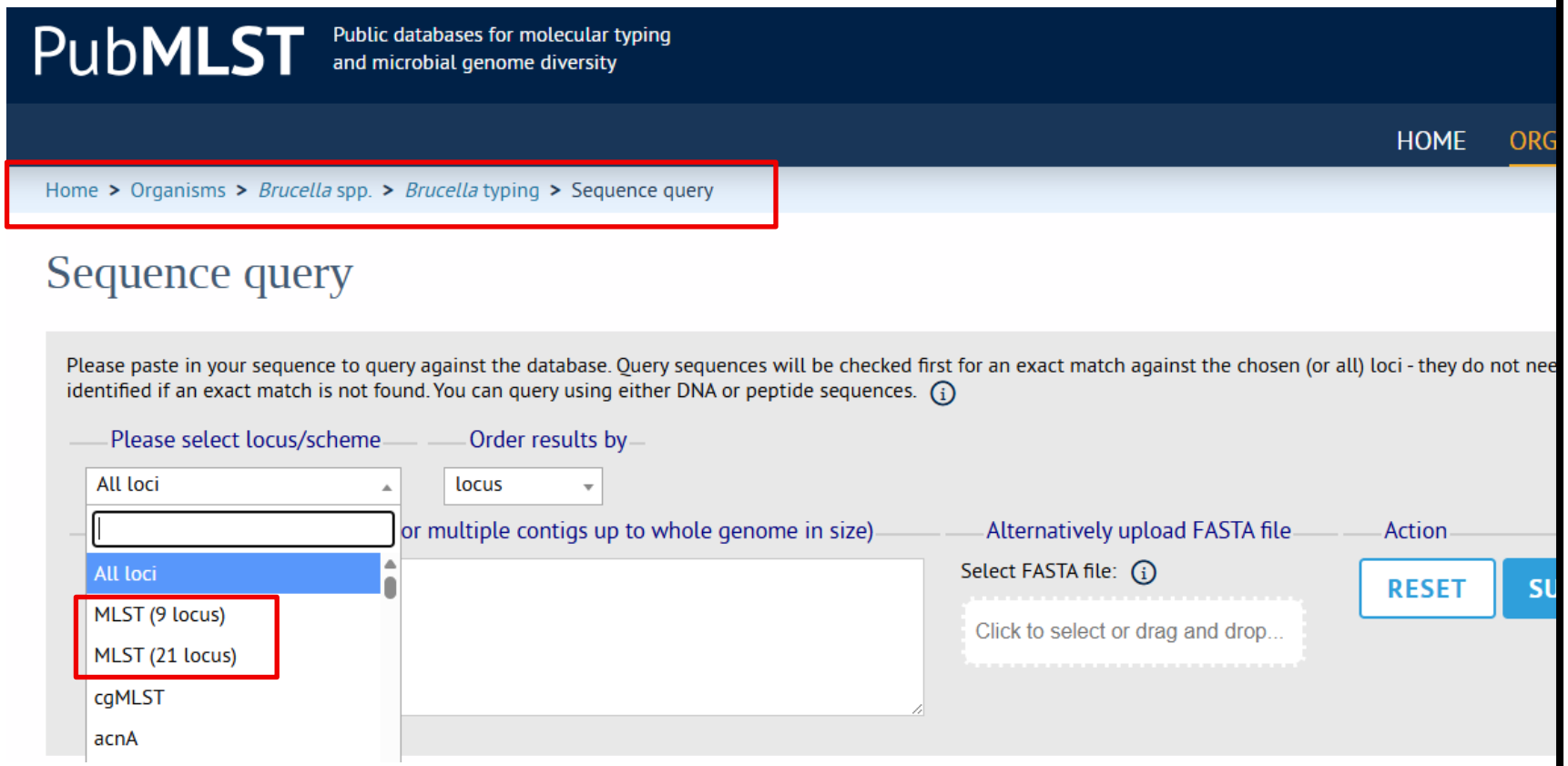
21 loci scheme: BruMLSA21 (Whatmore 2016)

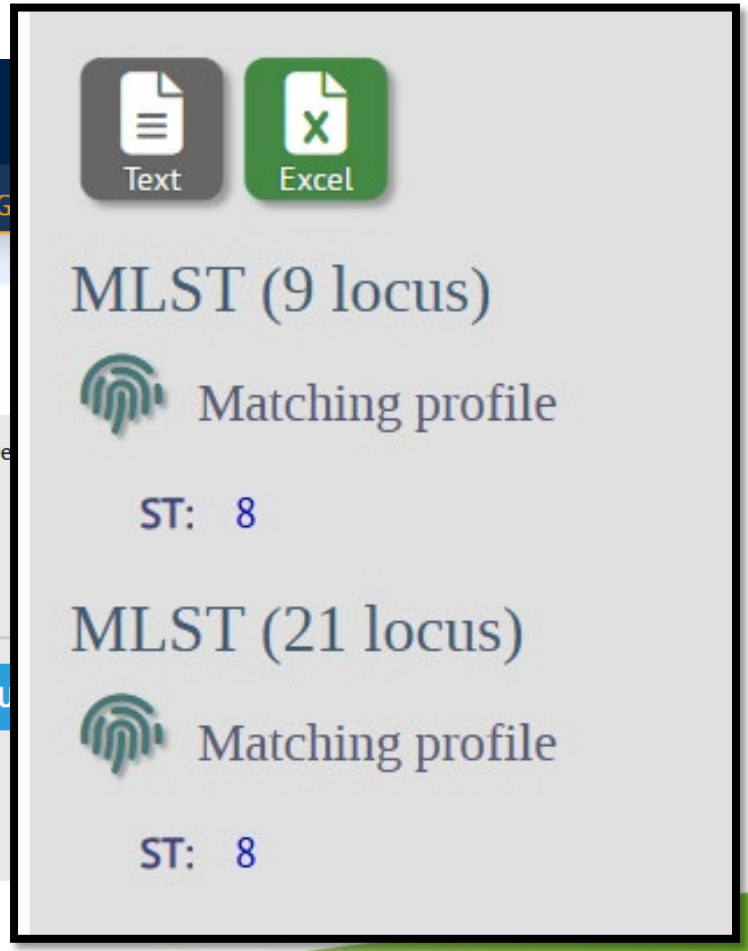
Same 9 loci, plus 12 others
Total 10.257 bp → more discriminatory
At development 101 STs, currently 179 profiles defined (as of 13 Jan 2026)

Both schemes: *No strong correlation between STs and biovars.*

Identification and typing of *Brucella* spp. [5-7]

MLST for *Brucella* spp. → https://pubmlst.org/organisms/brucella-spp





Identification and typing of *Brucella* spp. [5-7]

MLST for *Brucella* spp. → https://pubmlst.org/organisms/brucella-spp

NOTE

STs that seem close to each other (e.g. ST-6 and ST-7) are **not** more similar than STs that are far away (e.g. ST-1 and ST-28).

BruMLSA9

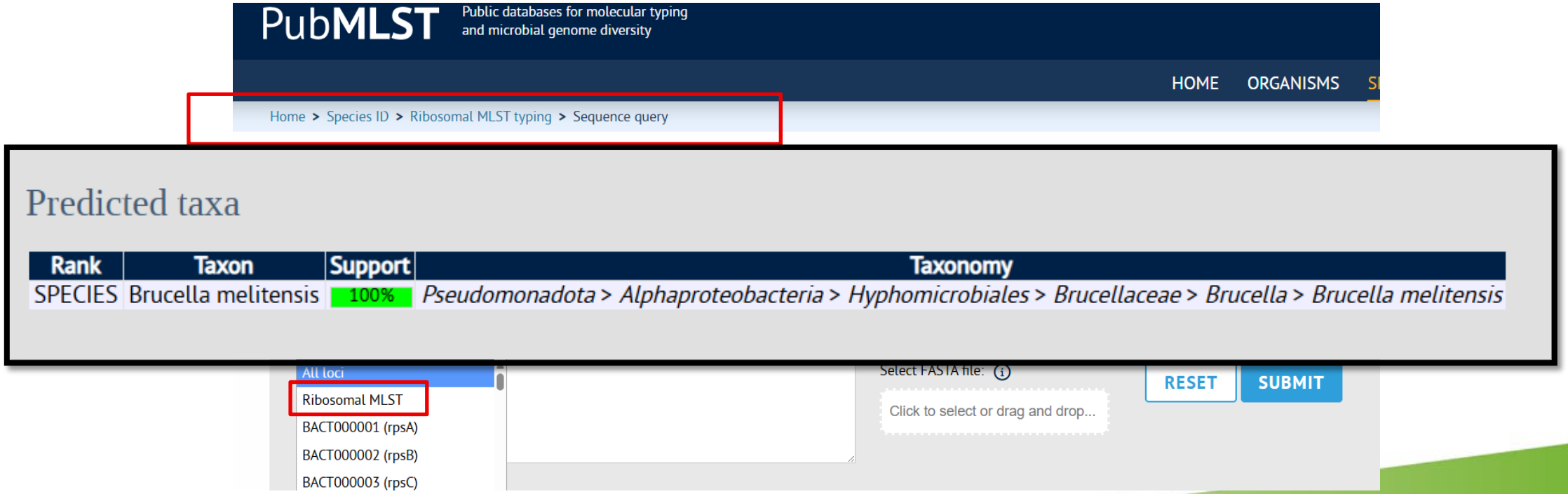
ST	gap	aroA	glk	dnaK	gyrB	trpE	cobQ	int_hyp	omp25
6	5	7	10	7	6	3	7	1	1
7	3	5	3	2	1	5	2	2	10
1	2	1	1	2	1	3	1	1	1
28	2	1	1	2	1	3	12	1	1

Whole-genome sequence-based tools [8]

rMLST (ribosomal protein subunits) → https://pubmlst.org/species-id

Used for species identification.

Same principle as MLST but the target genes are different and the amount is higher (~53 loci).



Whole-genome sequence-based tools

cgMLST (core genome)

Same principle as the previous two allelic schemes.

Used for typing.

Very high discriminatory power because there are **>1000 loci** in the schemes.

Importantly the output from cgMLST can be used for **cluster analysis**.

- This is achieved by calculating the allele dissimilarity between pairs of isolates.
- Pairs of isolates under a certain number of dissimilarities are genetically related → potentially clones, or part of the same outbreak.

More details about cgMLST-based cluster analysis on upcoming talks.

Whole-genome sequence-based tools [9-11]

cgMLST for *Brucella* spp. → https://www.cgmlst.org/ncs (database containing 'all' cgMLST schemes)

Scheme for *B. melitensis* – 2704 loci (Janowicz 2018)

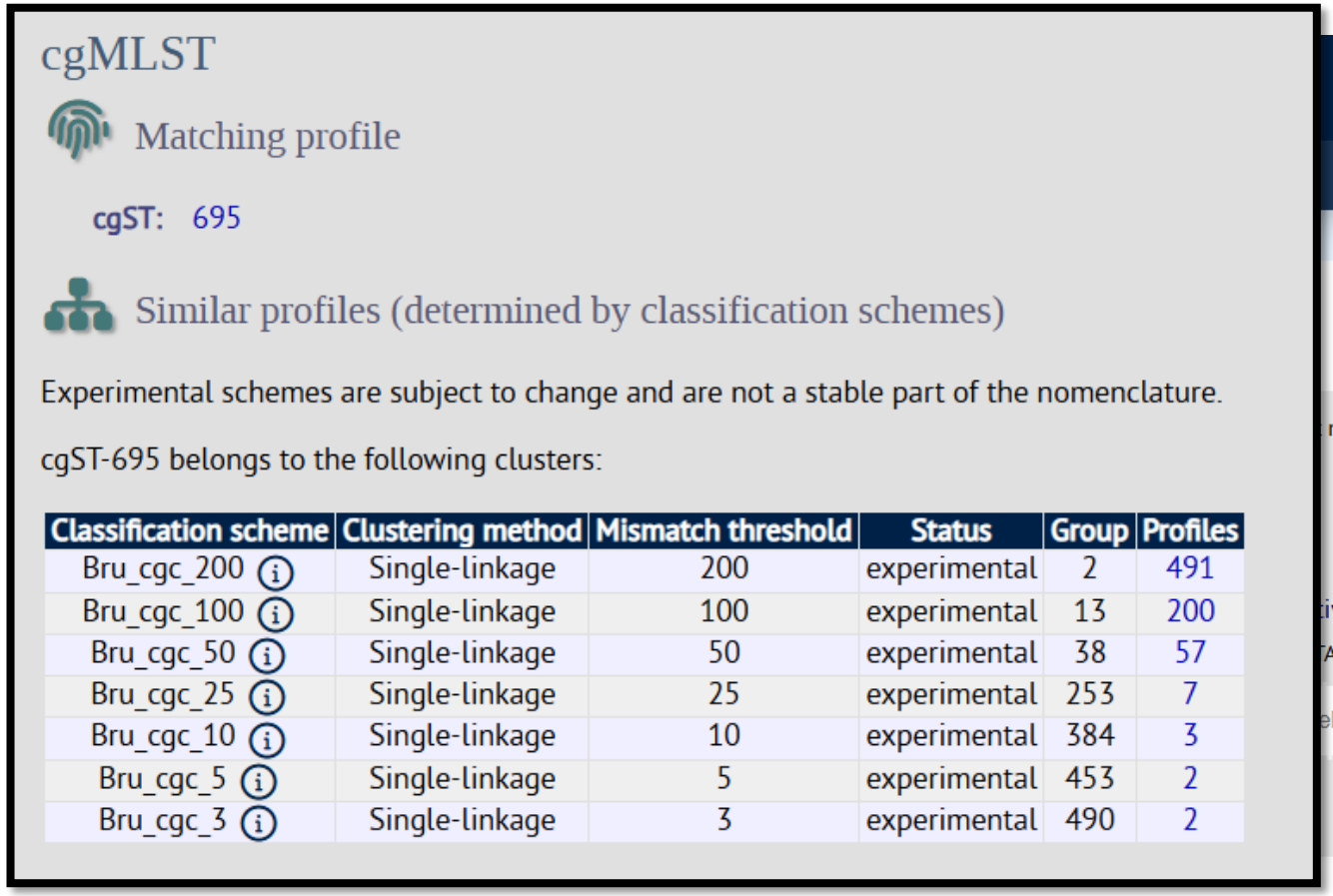
Scheme for *Brucella* spp. – 1764 loci (Abdel-Glil 2022)

Total 1.556.385 bp
All species
Available on pubMLST

Whole-genome sequence-based tools [9-11]

cgMLST for *Brucella* spp. → How to obtain a cgST for typing purposes

https://pubmlst.org/organisms/brucella-spp (same page as MLST)



cgMLST

Matching profile

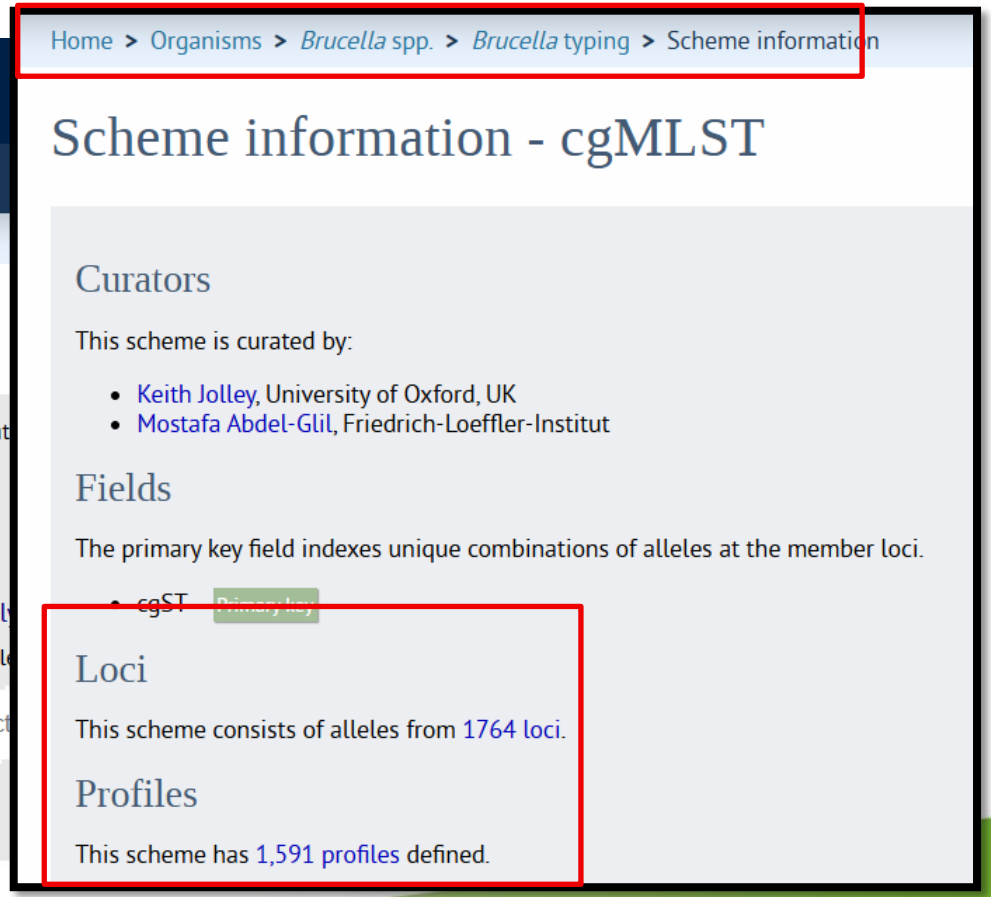
cgST: 695

Similar profiles (determined by classification schemes)

Experimental schemes are subject to change and are not a stable part of the nomenclature.

cgST-695 belongs to the following clusters:

Classification scheme	Clustering method	Mismatch threshold	Status	Group	Profiles
Bru_cgc_200	Single-linkage	200	experimental	2	491
Bru_cgc_100	Single-linkage	100	experimental	13	200
Bru_cgc_50	Single-linkage	50	experimental	38	57
Bru_cgc_25	Single-linkage	25	experimental	253	7
Bru_cgc_10	Single-linkage	10	experimental	384	3
Bru_cgc_5	Single-linkage	5	experimental	453	2
Bru_cgc_3	Single-linkage	3	experimental	490	2



Home > Organisms > *Brucella* spp. > *Brucella* typing > Scheme information

Scheme information - cgMLST

Curators

This scheme is curated by:

- Keith Jolley, University of Oxford, UK
- Mostafa Abdel-Glil, Friedrich-Loeffler-Institut

Fields

The primary key field indexes unique combinations of alleles at the member loci.

- cgST

Loci

This scheme consists of alleles from 1764 loci.

Profiles

This scheme has 1,591 profiles defined.

In summary

List of learning points in this session:

- *Brucella* are varied including differences in **metabolism** and in **host-preference**.
- Classically some *Brucella* species are further biotyped into **biovars**.
- Modern molecular methods incl. WGS allow for more strict differentiation of species and subtyping.
- Two methods were shown to **identify** *Brucella* isolates:
  - **KmerFinder** (alignment-based)
  - **rMLST** (gene-by-gene approach with a specific MLST scheme)
- Two methods were shown to **subtype** *Brucella* isolates:
  - Classic **MLST** (two schemes exist, with 9 and 21 loci)
  - **cgMLST** (two schemes exist, one for *B. melitensis* and another for all *Brucella* spp.)

References

1. Ficht T. *Brucella* **taxonomy and evolution.** Future Microbiol. 2010 Jun;5(6):859–866. doi: 10.2217/fmb.10.52
2. Whatmore AM. **Current understanding of the genetic diversity of Brucella, an expanding genus of zoonotic pathogens.** Infect Genet Evol. 2009 Dec;9(6):1168-84. doi: 10.1016/j.meegid.2009.07.001
3. Whatmore AM, Foster JT. **Emerging diversity and ongoing expansion of the genus Brucella.** Infect Genet Evol. 2021 Aug:92:104865. doi: 10.1016/j.meegid.2021.104865.
4. Clausen PTLC, et al. **Rapid and precise alignment of raw reads against redundant databases with KMA.** BMC Bioinformatics. 2018 Aug 29;19(1):307. doi: 10.1186/s12859-018-2336-6.
5. Jolley KA, et al. **Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications.** Wellcome Open Res. 2018 Sep 24:3:124. doi: 10.12688/wellcomeopenres.14826.1.
6. Whatmore AM, et al. **Characterisation of the genetic diversity of Brucella by multilocus sequencing.** BMC Microbiol. 2007 Apr 20:7:34. doi: 10.1186/1471-2180-7-34.
7. Whatmore AM, et al. **Extended Multilocus Sequence Analysis to Describe the Global Population Structure of the Genus Brucella: Phylogeography and Relationship to Biovars.** Front Microbiol. 2016 Dec 21:7:2049. doi: 10.3389/fmicb.2016.02049.
8. Jolley KA, et al. **Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain.** Microbiology. 2012 Apr;158(Pt 4):1005-1015. doi: 10.1099/mic.0.055459-0.
9. Zhou Z, et al. The **EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity.** Genome Res. 2020 Jan;30(1):138-152. doi: 10.1101/gr.251678.119.
10. Janowicz A, et al. **Core Genome Multilocus Sequence Typing and Single Nucleotide Polymorphism Analysis in the Epidemiology of Brucella melitensis Infections.** J Clin Microbiol. 2018 Aug 27;56(9):e00517-18. doi: 10.1128/JCM.00517-18.
11. Abdel-Glil MY, et al. **Core Genome Multilocus Sequence Typing Scheme for Improved Characterization and Epidemiological Surveillance of Pathogenic Brucella.** J Clin Microbiol. 2022 Aug 17;60(8):e0031122. doi: 10.1128/jcm.00311-22.

Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of Ana Rita Rebelo, DTU Food