

Bohlin et al. *BMC Evolutionary Biology* 2010, **10**:249  
<http://www.biomedcentral.com/1471-2148/10/249>



RESEARCH ARTICLE

Open Access

# Genomic comparisons of *Brucella* spp. and closely related bacteria using base compositional and proteome based methods

Jon Bohlin<sup>1\*</sup>, Lars Snipen<sup>2</sup>, Axel Cloeckert<sup>3</sup>, Karin Lagesen<sup>4</sup>, David Ussery<sup>5</sup>, Anja B Kristoffersen<sup>4,6</sup>, Jacques Godfroid<sup>7</sup>

## Abstract

**Background:** Classification of bacteria within the genus *Brucella* has been difficult due in part to considerable genomic homogeneity between the different species and biovars, in spite of clear differences in phenotypes. Therefore, many different methods have been used to assess *Brucella* taxonomy. In the current work, we examine 32 sequenced genomes from genus *Brucella* representing the six classical species, as well as more recently described species, using bioinformatical methods. Comparisons were made at the level of genomic DNA using oligonucleotide based methods (Markov chain based genomic signatures, genomic codon and amino acid frequencies based comparisons) and proteomes (all-against-all BLAST protein comparisons and pan-genomic analyses).

**Results:** We found that the oligonucleotide based methods gave different results compared to that of the proteome based methods. Differences were also found between the oligonucleotide based methods used. Whilst the Markov chain based genomic signatures grouped the different species in genus *Brucella* according to host preference, the codon and amino acid frequencies based methods reflected small differences between the *Brucella* species. Only minor differences could be detected between all genera included in this study using the codon and amino acid frequencies based methods.

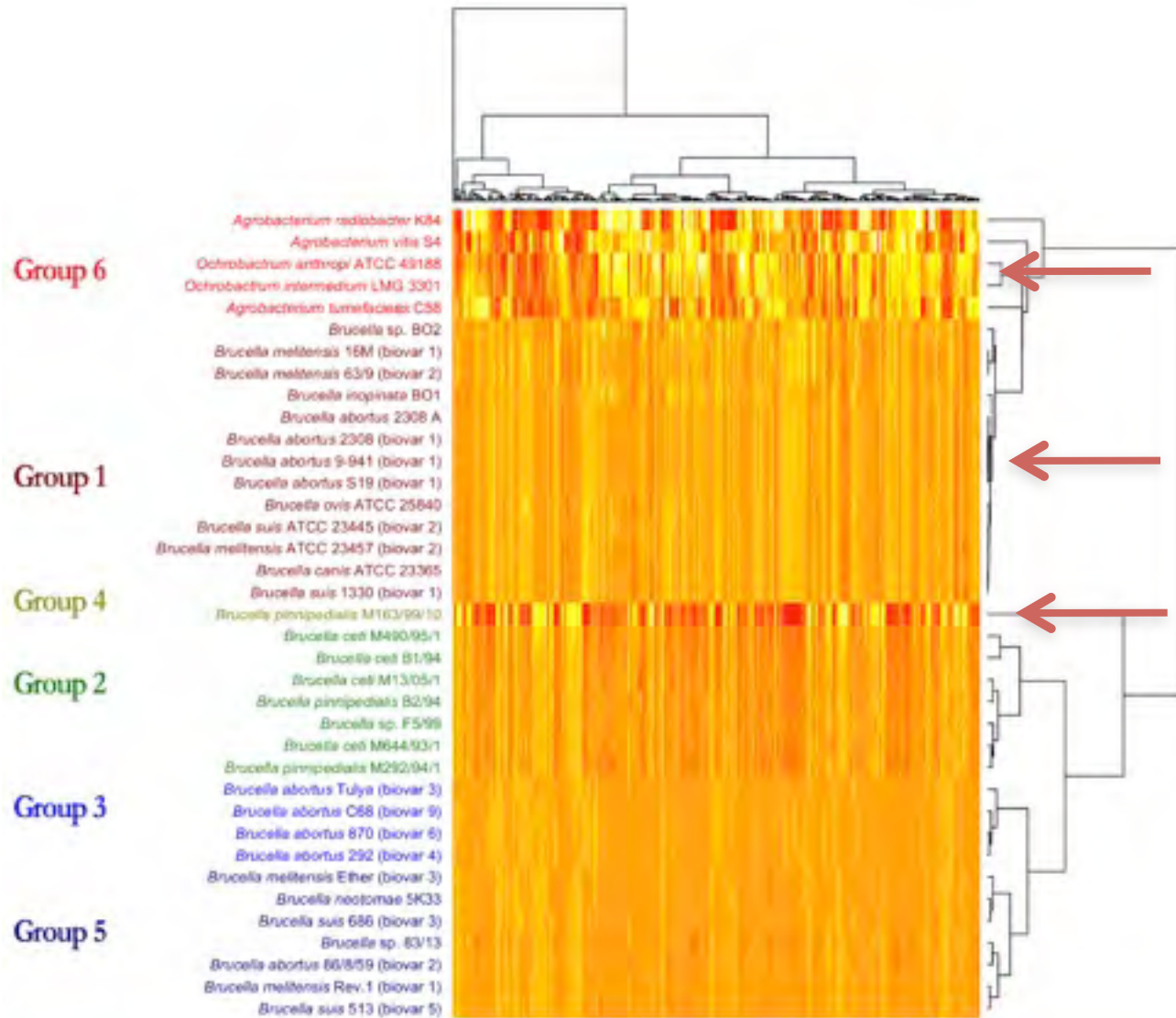
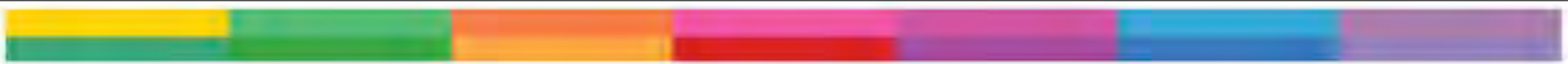
Proteome comparisons were found to be in strong accordance with current *Brucella* taxonomy indicating a remarkable association between gene gain or loss on one hand and mutations in marker genes on the other. The proteome based methods found greater similarity between *Brucella* species and *Ochrobactrum* species than between species within genus *Agrobacterium* compared to each other. In other words, proteome comparisons of species within genus *Agrobacterium* were found to be more diverse than proteome comparisons between species in genus *Brucella* and genus *Ochrobactrum*. Pan-genomic analyses indicated that uptake of DNA from outside genus *Brucella* appears to be limited.

**Conclusions:** While both the proteome based methods and the Markov chain based genomic signatures were able to reflect environmental diversity between the different species and strains of genus *Brucella*, the genomic codon and amino acid frequencies based comparisons were not found adequate for such comparisons. The proteome comparison based phylogenies of the species in genus *Brucella* showed a surprising consistency with current *Brucella* taxonomy.

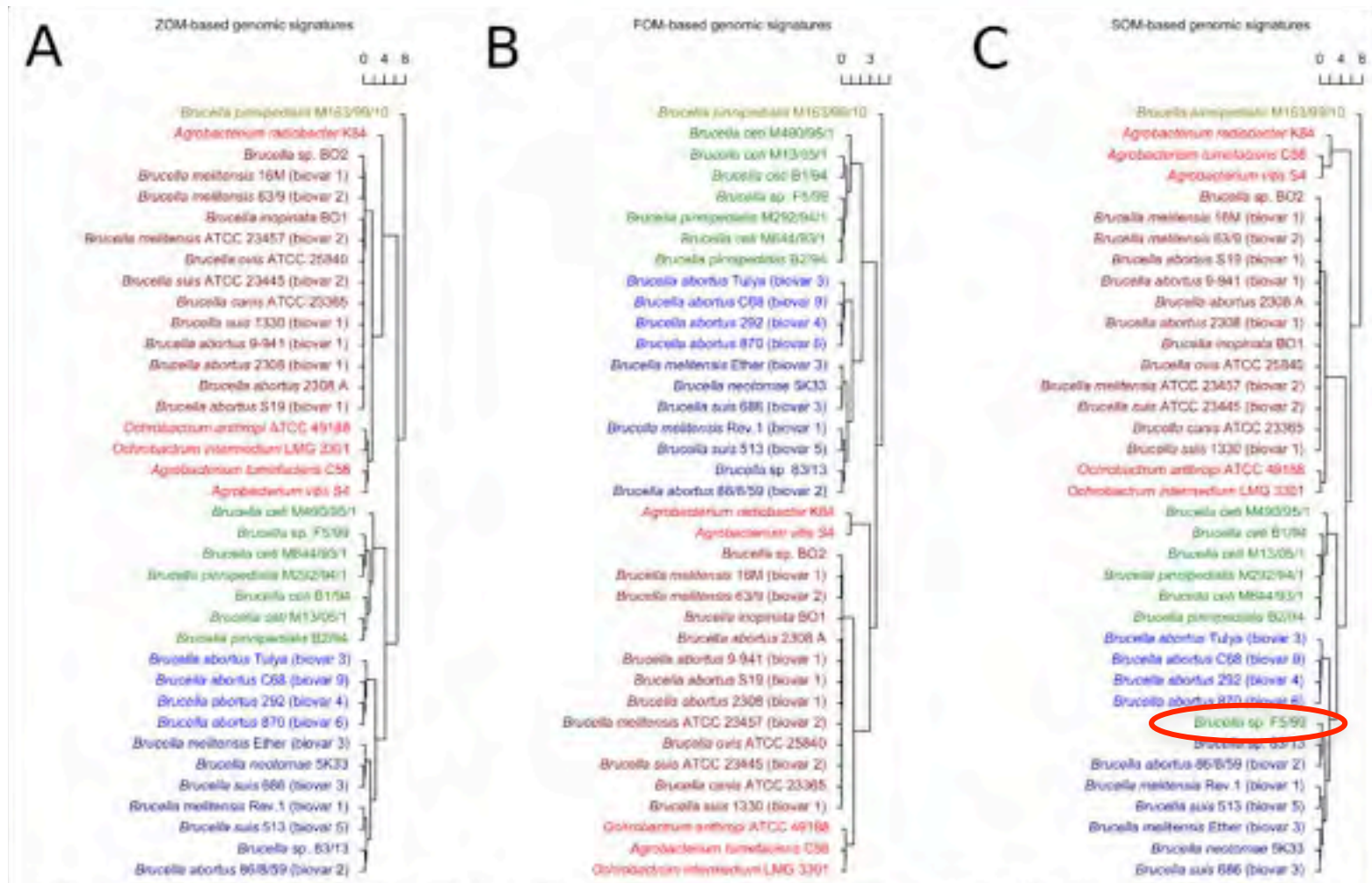
**Table 1 0<sup>th</sup> order Markov chain model based cluster groups of *Brucella* genomes**

Name	Accession	Database	Group	%GC	Size (mbp)	Host
<i>Brucella abortus</i> 2308 (biovar 1)	AM040264.1	Genbank/NCBI	1	57	3.28	Cattle
<i>Brucella abortus</i> 2308 A <sup>§</sup>	VF00022-VF00023	PATRIC	1	57	3.31	Cattle
<i>Brucella abortus</i> 9-941 (biovar 1)	AE017223.1	Genbank/NCBI	1	57	3.28	Cattle
<i>Brucella abortus</i> 519 (biovar 1)	CP000887.1	Genbank/NCBI	1	57	3.29	Cattle
<i>Brucella canis</i> ATCC 23365	CP000872.1	Genbank/NCBI	1	57	3.32	Dog
<i>Brucella inopinata</i> BO1 <sup>§</sup>	VF000041-VF000043	PATRIC	1	57	3.37	Human
<i>Brucella melitensis</i> 16M (biovar 1)	AE008917.1	Genbank/NCBI	1	57	3.32	Sheep, goat
<i>Brucella melitensis</i> 63/9 (biovar 2) <sup>§</sup>	ACEM01000000	Broad Institute	1	57	3.29	Sheep, goat
<i>Brucella melitensis</i> ATCC 23457 (biovar 2)	CP001488.1	Genbank/NCBI	1	57	3.28	Sheep, goat
<i>Brucella ovis</i> ATCC 25940	CP000709.1	Genbank/NCBI	1	57	3.28	Sheep
<i>Brucella</i> sp. BO2 <sup>§</sup>	VF00103-VF00105	PATRIC	1	57	3.28	Human
<i>Brucella suis</i> 1330 (biovar 1)	AE014291.4	Genbank/NCBI	1	57	3.31	Pig
<i>Brucella suis</i> ATCC 23445 (biovar 2)	CP000911.1	Genbank/NCBI	1	57	3.31	Pig, hare
<i>Brucella ceti</i> B1/94 <sup>§</sup>	ACE01000000	Broad Institute	2	58	3.34	Dolphin, porpoise
<i>Brucella ceti</i> M13/05/1 <sup>§</sup>	ACBP01000000	Broad Institute	2	58	3.34	Dolphin, porpoise
<i>Brucella ceti</i> M490/95/1 <sup>§</sup>	ACEJ01000000	Broad Institute	2	58	3.35	Dolphin, porpoise
<i>Brucella ceti</i> M644/93/1 <sup>§</sup>	ACBC01000000	Broad Institute	2	58	3.33	Dolphin, porpoise
<i>Brucella pinnipedalis</i> 80/94 <sup>§</sup>	ACBN01000000	Broad Institute	2	58	3.34	Seal
<i>Brucella pinnipedalis</i> M292/94/1 <sup>§</sup>	ACEF01000000	Broad Institute	2	58	3.37	Seal
<i>Brucella</i> sp. F5/99 <sup>§</sup>	ACFF01000000	Broad Institute	2	58	3.4	Dolphin
<i>Brucella abortus</i> 86/8/59 (biovar 2) <sup>§</sup>	ACBJ01000000	Broad Institute	3	58	3.32	Cattle
<i>Brucella melitensis</i> Ether (biovar 3) <sup>§</sup>	ACE01000000	Broad Institute	3	57	3.28	Sheep, goat
<i>Brucella melitensis</i> Rev.1 (biovar 1) <sup>§</sup>	ACEG01000000	Broad Institute	3	57	3.31	Sheep, goat
<i>Brucella neotomae</i> 5K33 <sup>§</sup>	ACEH01000000	Broad Institute	3	58	3.33	Rodent
<i>Brucella</i> sp. 83/13 <sup>§</sup>	ACBQ01000000	Broad Institute	3	58	3.29	Rodent
<i>Brucella suis</i> 513 (biovar 5) <sup>§</sup>	ACBK01000000	Broad Institute	3	58	3.15	Pig
<i>Brucella suis</i> 686 (biovar 3) <sup>§</sup>	ACBL01000000	Broad Institute	3	58	3.3	Pig
<i>Brucella pinnipedalis</i> M163/99/10 <sup>§</sup>	ACBA01000000	Broad Institute	4	59	3.41	Seal
<i>Brucella abortus</i> 292 (biovar 4) <sup>§</sup>	ACBH01000000	Broad Institute	5	58	3.28	Cattle
<i>Brucella abortus</i> 870 (biovar 4) <sup>§</sup>	ACBG01000000	Broad Institute	5	58	3.27	Cattle
<i>Brucella abortus</i> C68 (biovar 9) <sup>§</sup>	ACEI01000000	Broad Institute	5	58	3.27	Cattle
<i>Brucella abortus</i> Tulya (biovar 3) <sup>§</sup>	ACBI01000000	Broad Institute	5	58	3.31	Human
<i>Agrobacterium radiobacter</i> K84	CP000628.1	Genbank/NCBI	6	60	6.66	Plant
<i>Agrobacterium tumefaciens</i> C58	AE007869.2	Genbank/NCBI	6	59	4.92	Plant
<i>Agrobacterium vitis</i> 54	CP000633.1	Genbank/NCBI	6	58	5.01	Plant
<i>Ochrobactrum anthracis</i> ATCC 49188	CP000758.1	Genbank/NCBI	6	56	4.78	Human, plant
<i>Ochrobactrum intermedium</i> IMG 3301 <sup>§</sup>	VF000028-VF000031	PATRIC	6	58	4.73	Human

<sup>§</sup>Genomes not assembled; therefore GC content and genome size are only approximate values

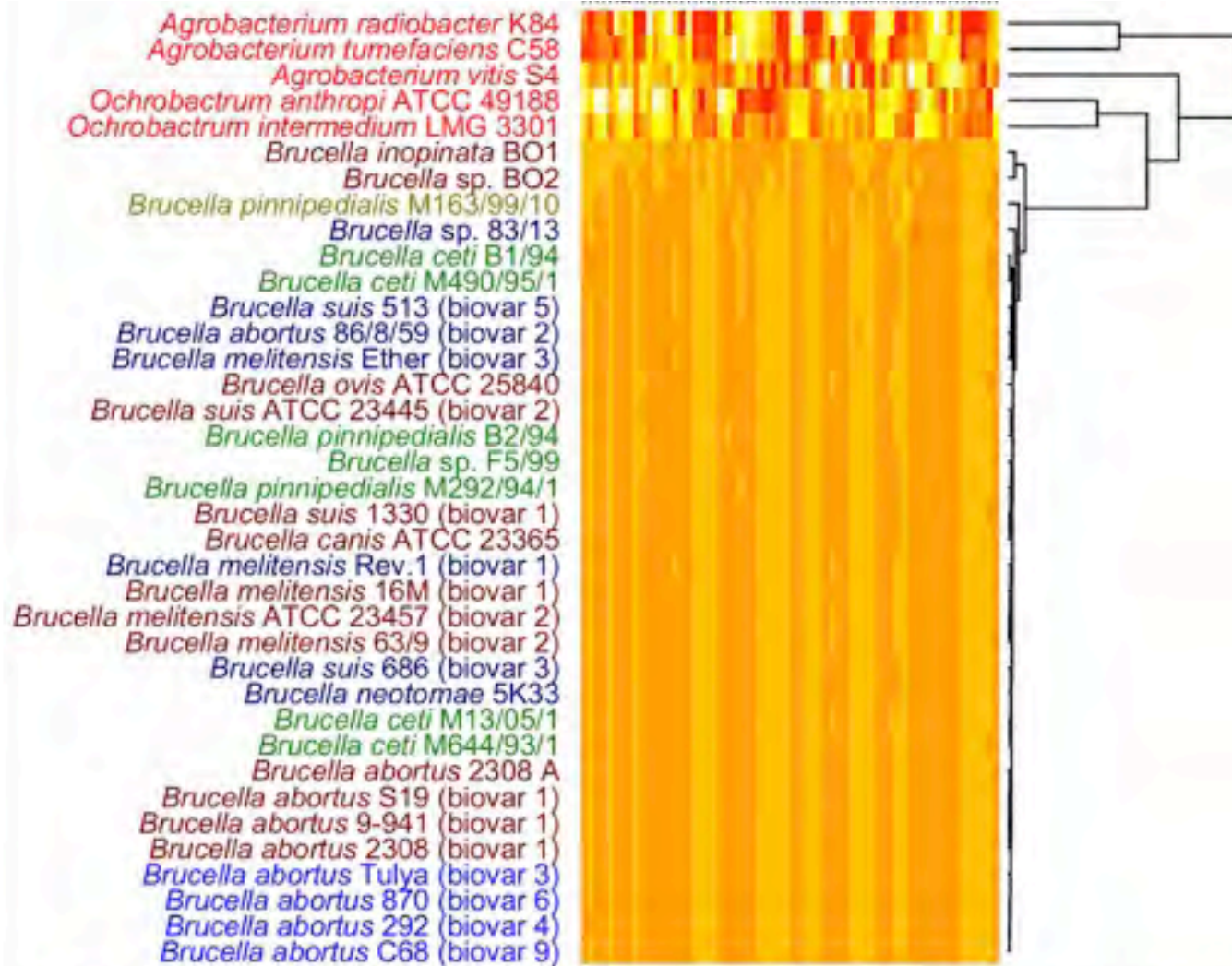


**Figure 1 ZOM based heatmap.** The ZOM Based heatmap shows all genomes compared using cluster analysis based on 0<sup>th</sup> order Markov chain model predicted tetranucleotide frequencies. It can be seen that the sequenced species from genus *Brucella* are very similar in terms of tetranucleotide usage patterns, with larger differences found in the more distantly related genera of *Agrobacterium* and *Ochrobactrum*. Although all species in genus *Brucella* are very similar in terms of base composition, as measured using the ZOM based method, several subgroups can be observed. For instance, marine associated (Groups 2 and 4) and terrestrial mammal associated (Groups 1, 3 and 5) species of genus *Brucella* are segregated into different groups.



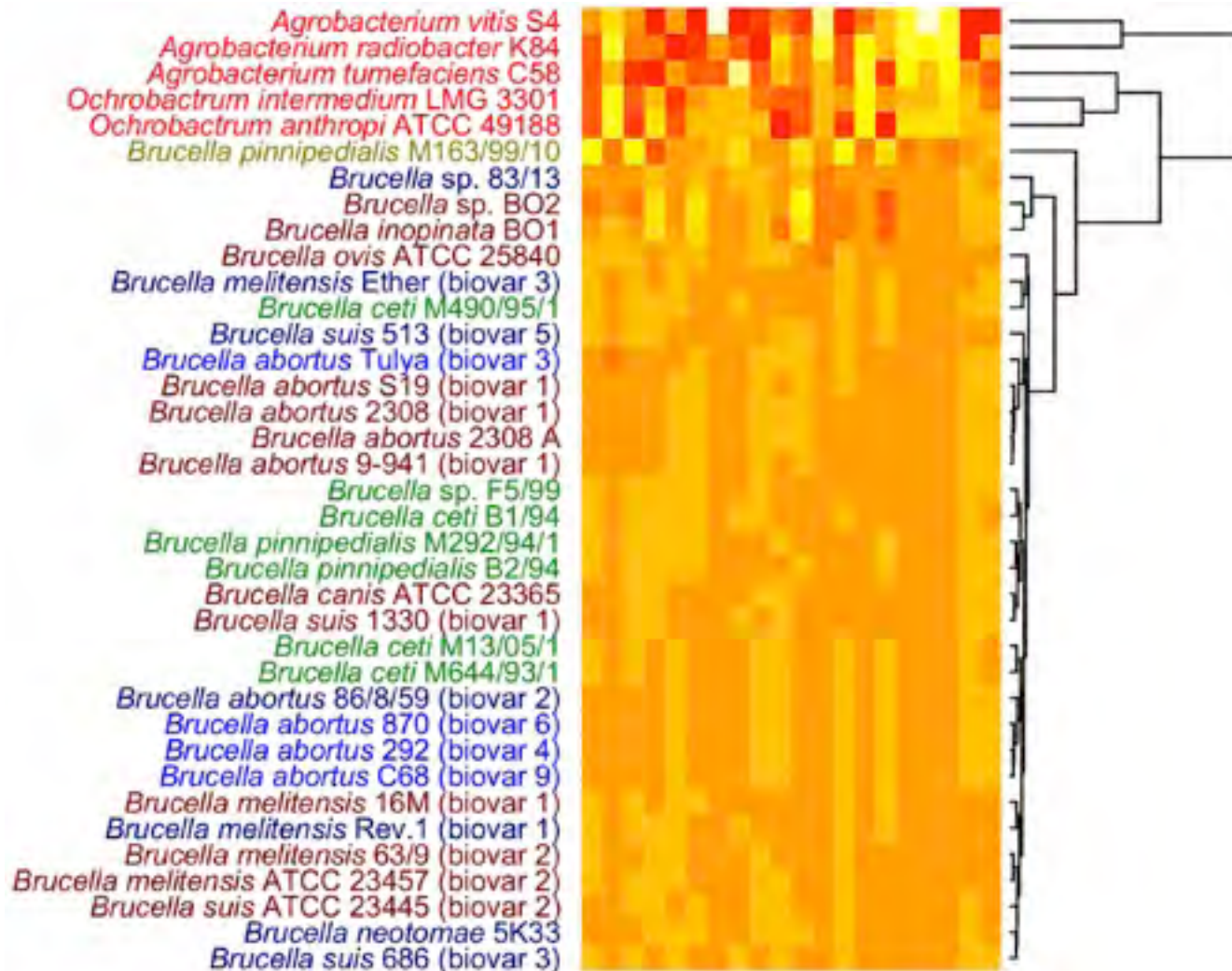
**Figure 2 Phylogenetic trees based on ZOM, FOM and SOM based genomic signatures.** The ZOM, FOM and SOM based trees (panels A, B, C, respectively) all show genomes compared using 0<sup>th</sup>, 1<sup>st</sup> and 2<sup>nd</sup> order Markov chain model based genomic signatures and clustered using average linkage with Euclidean distance. It can be seen that there is large agreement between the different Markov chain based genomic signatures. For all models, marine and terrestrial host based species of genus *Brucella* appear in distinctive groups, except for the SOM based tree (C) where the marine mammal associated strain *B. sp.* F5/99 is grouped with terrestrial mammal associated species. The more distantly related genera of *Agrobacterium* and *Ochrobactrum* appear in separate groups than the species of genus *Brucella* for all genomic signature based models.

# Codon frequency based clustering

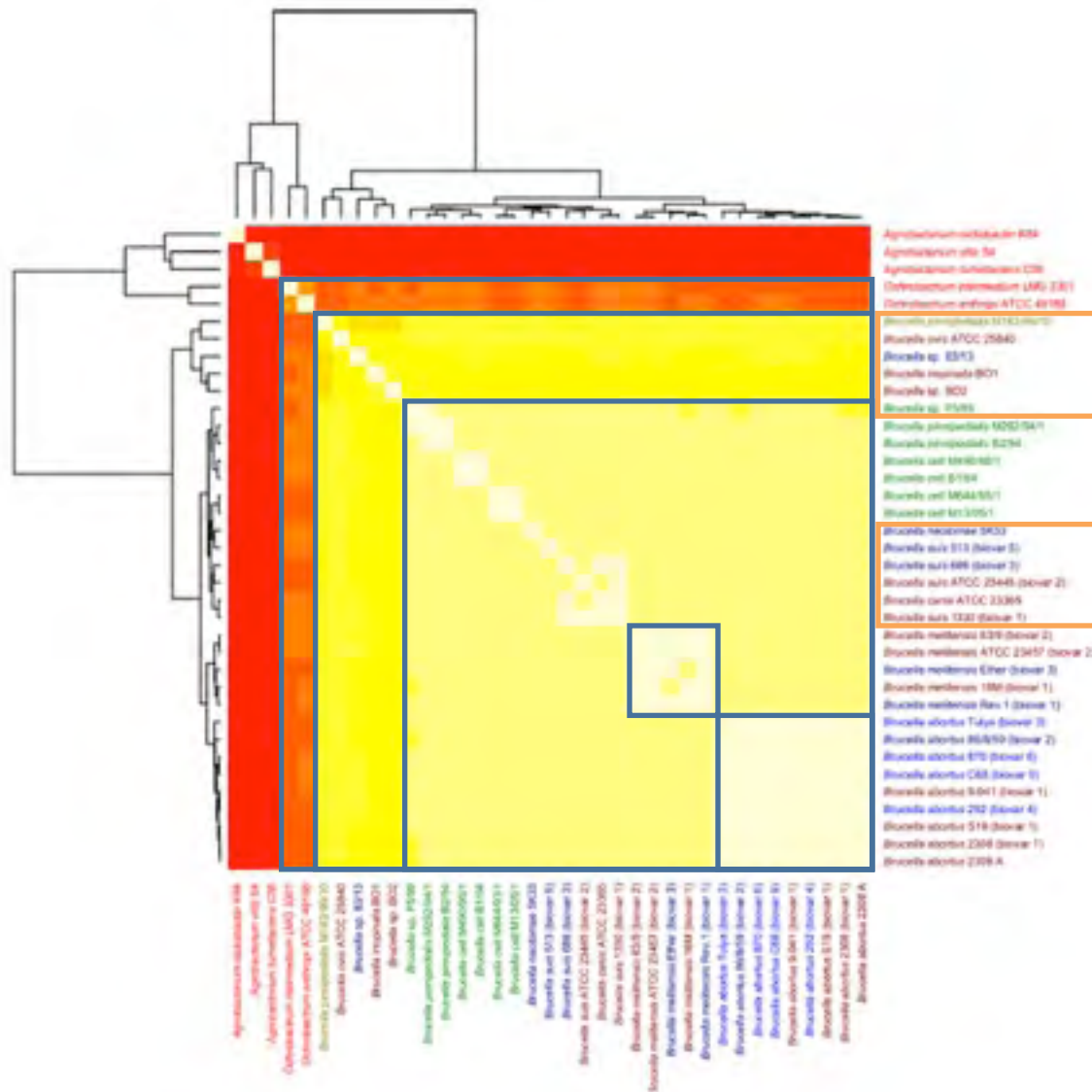


**Figure 3 Codon frequencies based heatmap.** The codon frequencies based heatmap is made from all the trinucleotide (codon) frequencies found in all predicted genes from the genomes of all species in the genera: *Brucella*, *Agrobacterium* and *Ochrobactrum*. It can be seen that there are small differences among the species in genus *Brucella* and only slightly more variance between the species found in the other genera.

# Amino acid clustering



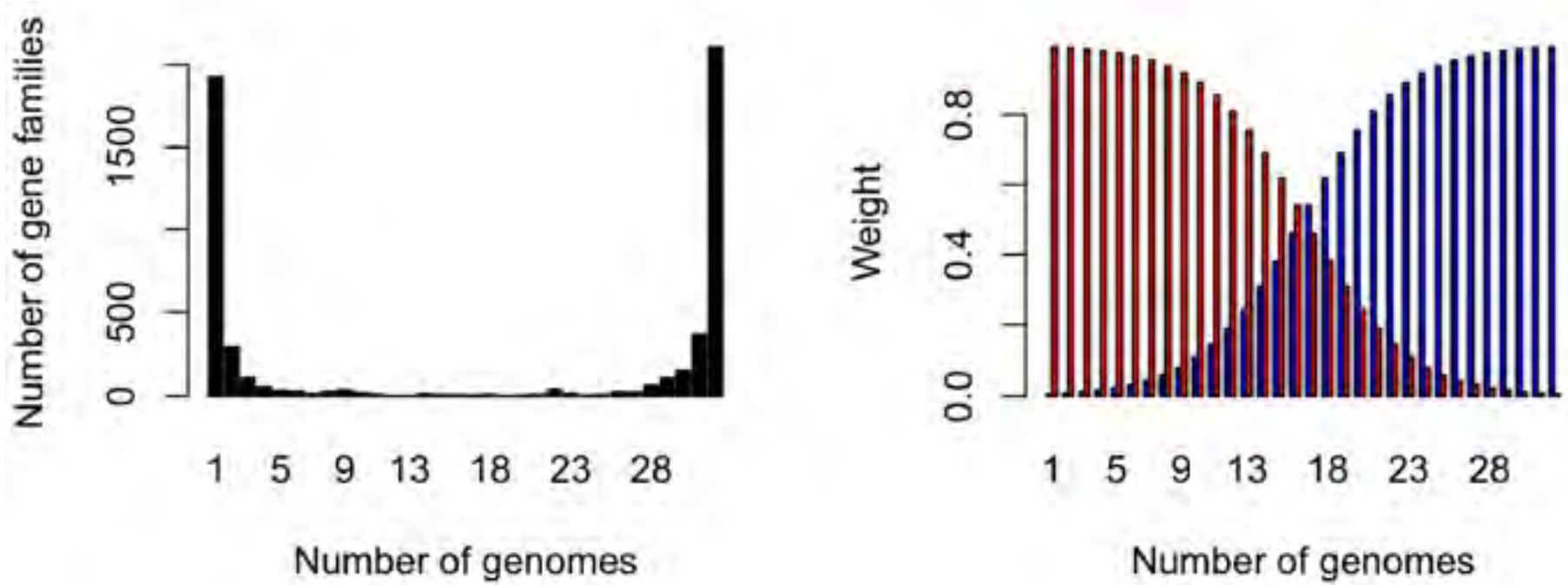
**Figure 4 Amino acid cluster diagram.** The amino acid frequencies based cluster diagram is made from the amino acid usage found in the predicted genes of the genomes in question. Amino acid frequencies correlate strongly with codon usage frequencies, but at a lower resolution. It can be seen that only small differences can be detected between species in genus *Brucella*.



**Figure 5 BLAST matrix.** Genomes are compared gene-wise using BLAST. All genes were converted to proteins and compared pair-wise all-against-all for each genome. Lighter color means closer similarity. Paralogs are removed when genomes are compared to them self which means that the hit score is less than 100%.



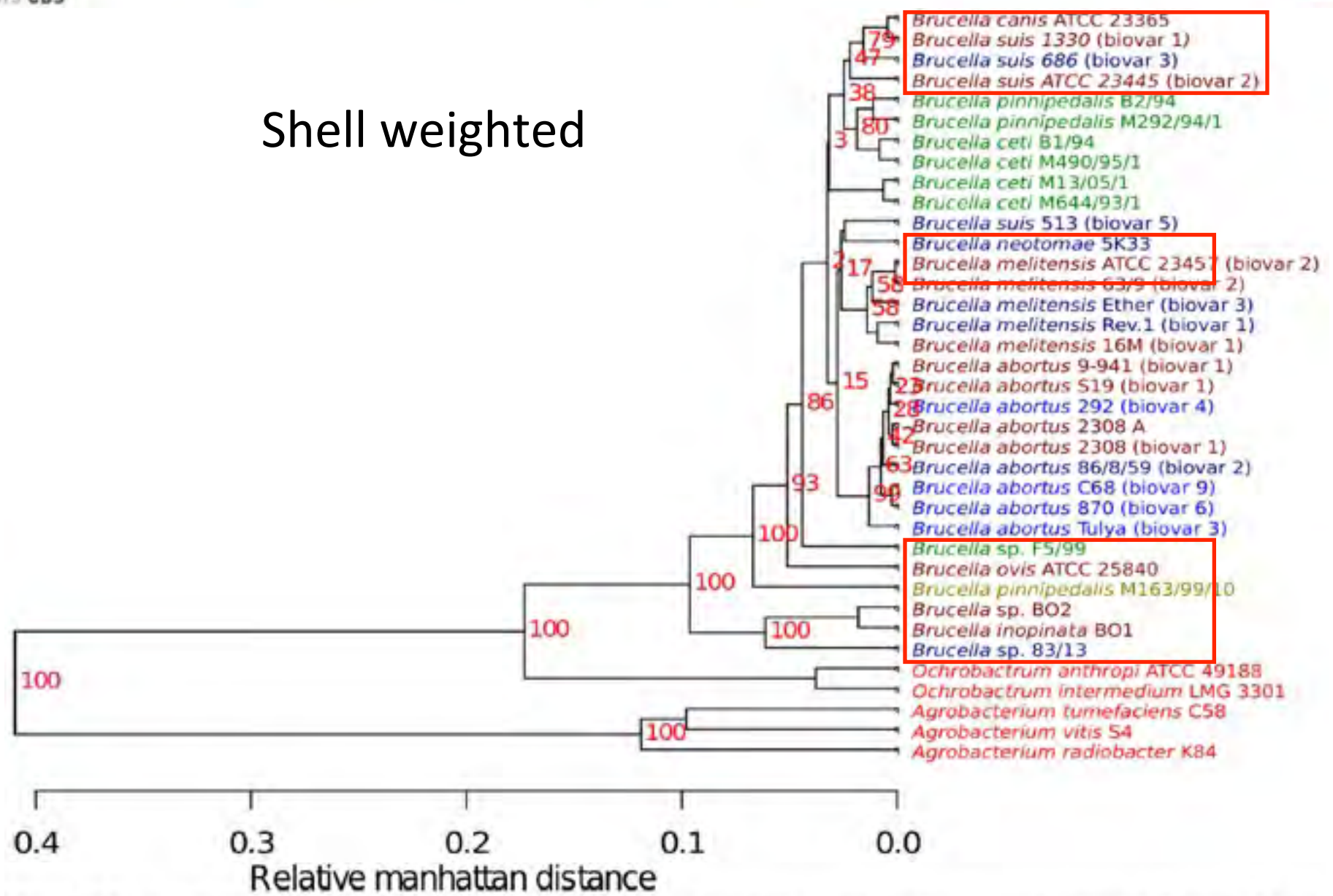
# Weighted pan-genomes



**Figure 8 Weighted pan-genomes.** The left panel is a bar-chart showing the number of gene families observed in 1, 2, ..., 32 gene families within the *Brucella* pan-genome (only genomes from species in genus *Brucella* are included). The gene families only found in a few genomes (left-wing bars) are called cloud genes, and those found in most genomes (right-wing bars) are called shell genes. The core genes are those that are found in all genomes (rightmost bar). The right panel indicates two different weighting strategies for computing the pan-genome tree. The red bars give more weight to cloud genes when comparing two genomes, while the blue bars emphasize the shell genes.



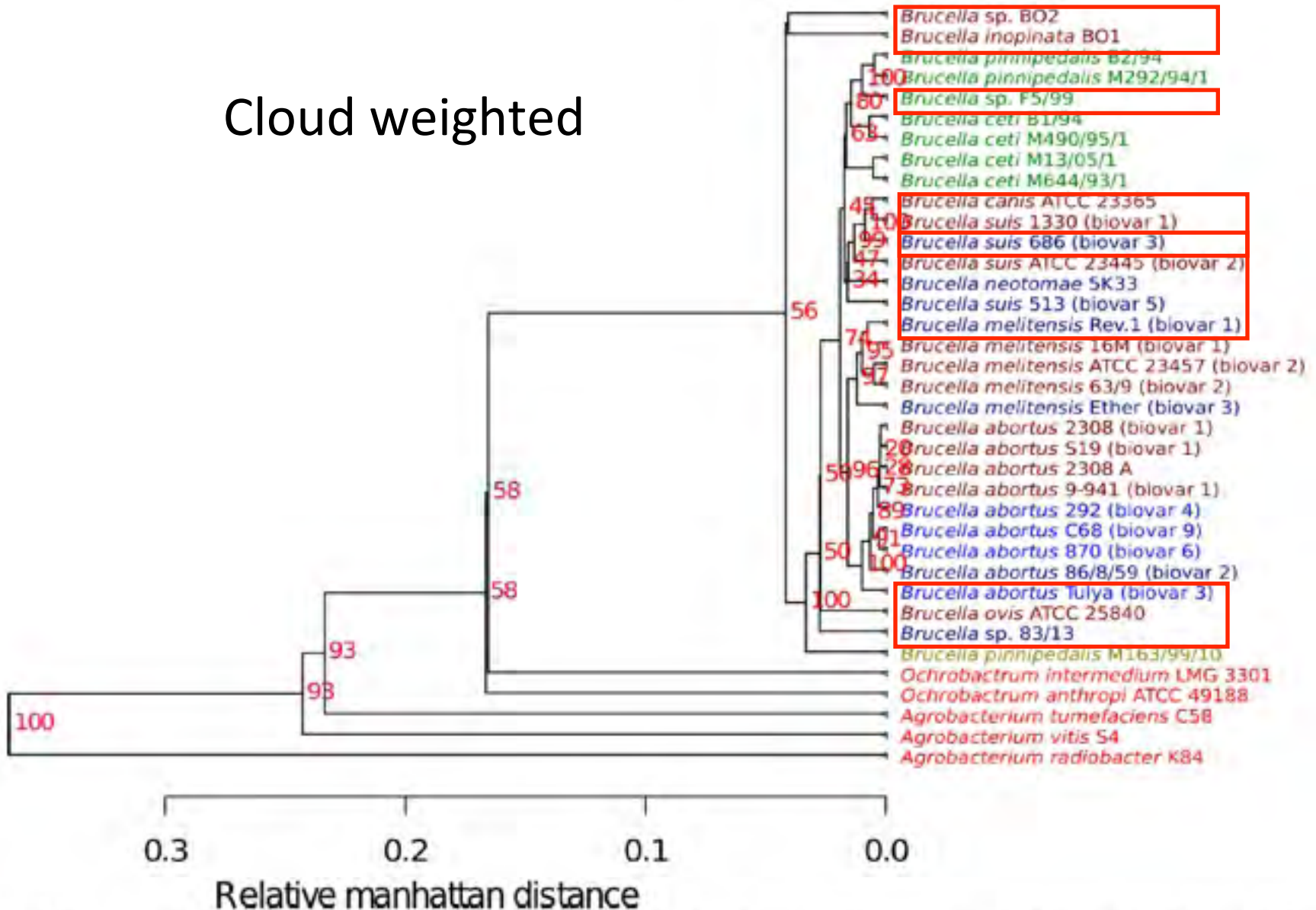
Shell weighted



**Figure 6 Pan-genomic shell tree.** The Figure shows a pan-genomic tree based on a weighting strategy emphasizing 'shell' genes. The red numbers are bootstrap-values, in percentage, based on 100 bootstrap samples.



# Cloud weighted



**Figure 7 Pan-genomic cloud tree.** The Figure shows a pan-genomic tree based on a weighting strategy emphasizing 'cloud' genes. The red numbers are bootstrap-values, in percentage, based on 100 bootstrap samples.

# Conclusion

- The genomes of the *Brucella* species are so alike, that genome sub-protein level methods are needed
- 0<sup>th</sup> Markov chain method is better at distinguishing between closely related species than other methods
- Pangenomic analysis indicate that very little or no transfer of DNA horizontally occurs
- Evolutionary indications

# Critics

## PROS

- Very well written and constructed
- Aware of the method's limitation

## CONS

- About mathematics, not biological
- Methods are not very well described for a non-mathematician
- No measure on figures
- Vague conclusions