Phylogenomic resolution of the bacterial genus *Pantoea* and its relationships with *Erwinia* and *Tatumella*
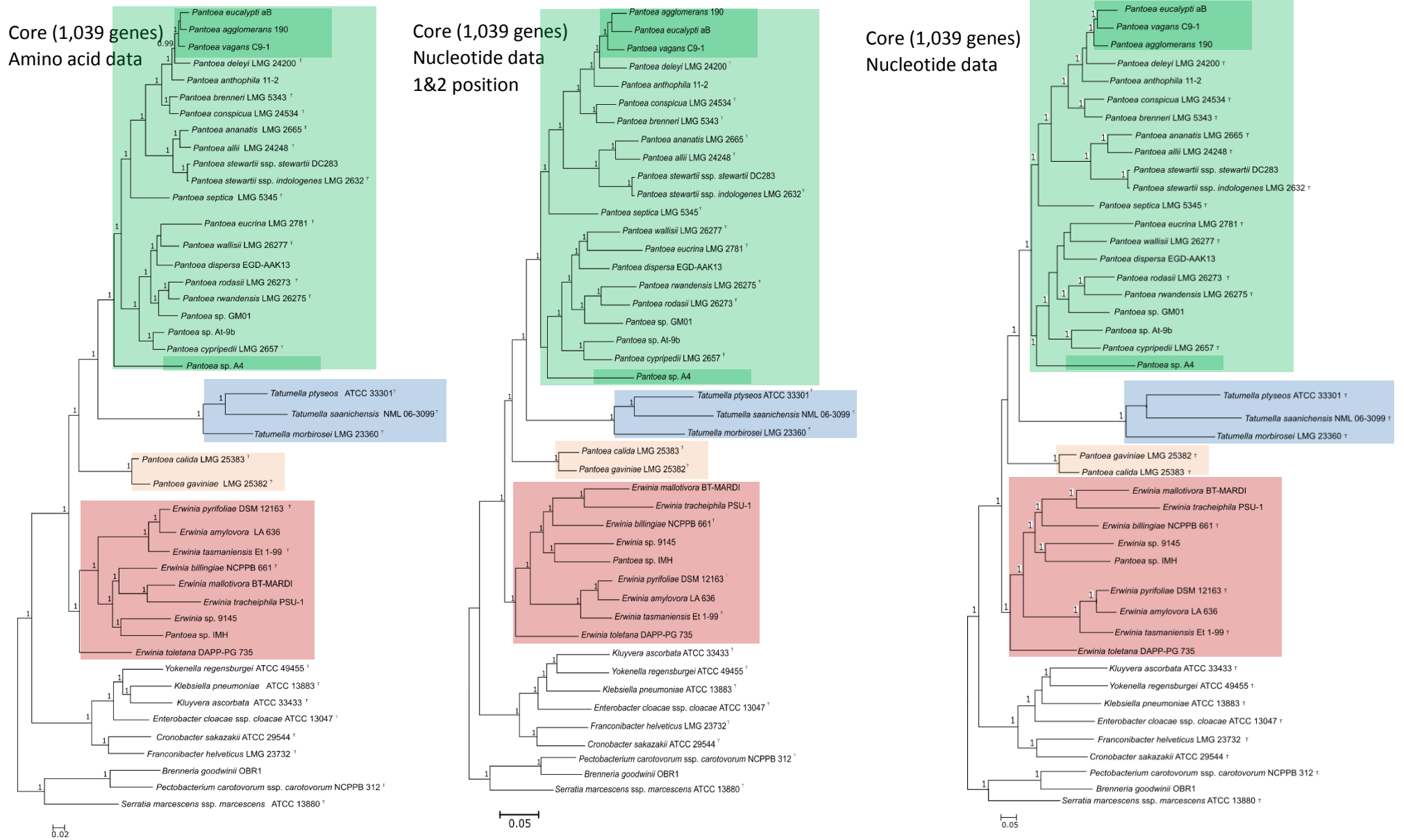
Marike Palmer[1] *, Emma T. Steenkamp[1], Martin P. A. Coetzee [2], Wai-Yin Chan[1] Elritha van Zyl[1], Pieter De Maayer[3], Teresa A. Coutinho[2], Jochen Blom[4], Theo H. M. Smits[5], Brion Duffy[6], Stephanus N. Venter[1]

Supplementary Figure S1. Comparison of the full dataset (1,039 sequences) maximum likelihood trees (RAxML) with support values inferred with FastTree (amino acid sequences, 3rd nucleotide position excluded, full nucleotide dataset). Darker blocks indicate differences in the relationships of the ingroup taxa between these three topologies.
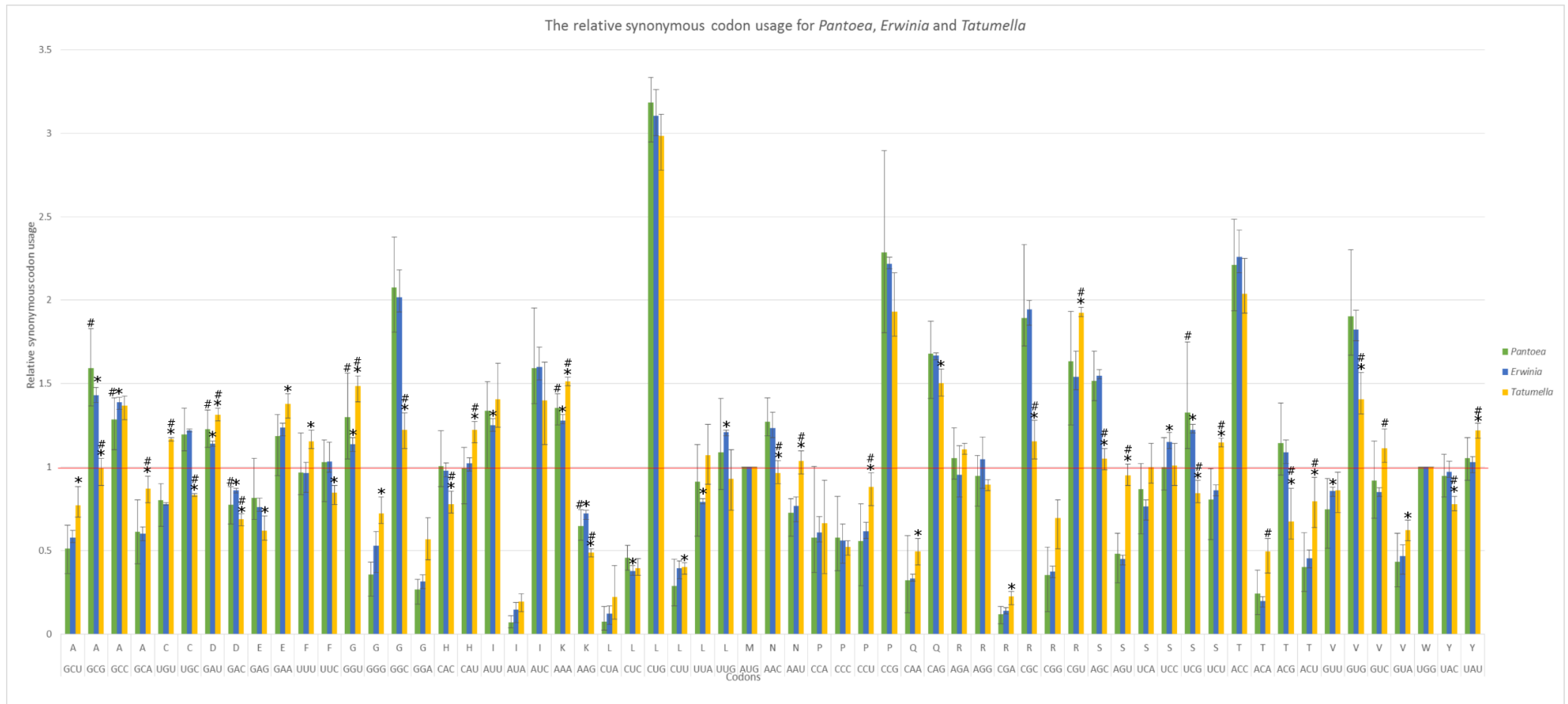
Supplementary Figure S2. The relative synonymous codon usage for *Erwinia* **(blue)**, *Pantoea* **(green)** and *Tatumella* **(yellow)**. * Indicate values significantly different from *Pantoea* and # indicate values significantly different from *Erwinia*, based on two-tailed unpaired t-tests (p<0.05).
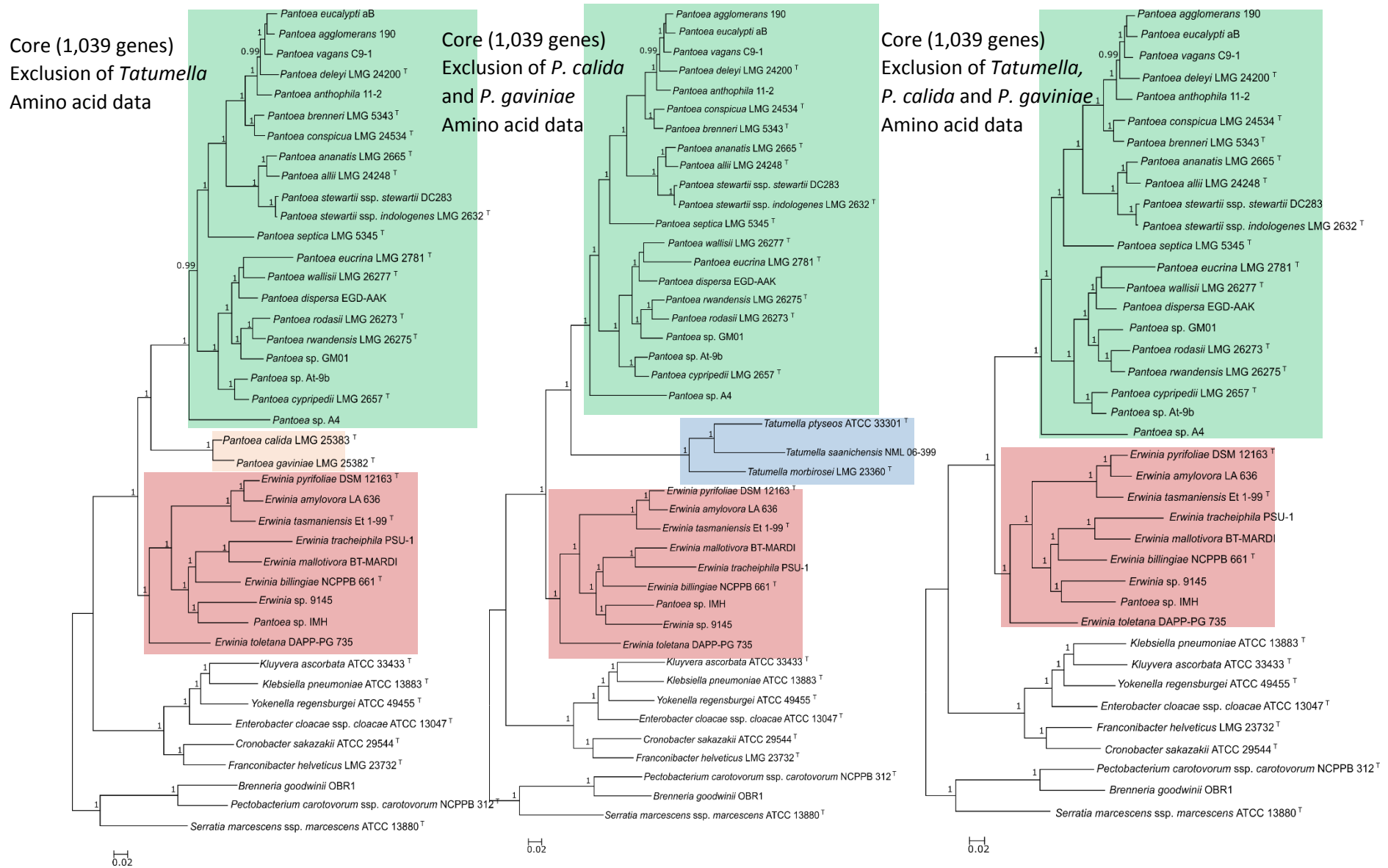
Supplementary Figure S3. Amino acid based maximum likelihood trees with the exclusion of *Tatumella*, *Pantoea calida* and *Pantoea gaviniae*, and *Tatumella* with *Pantoea calida* and *Pantoea gaviniae*, respectively. SH-support values were calculated in FastTree and are indicated at the nodes.
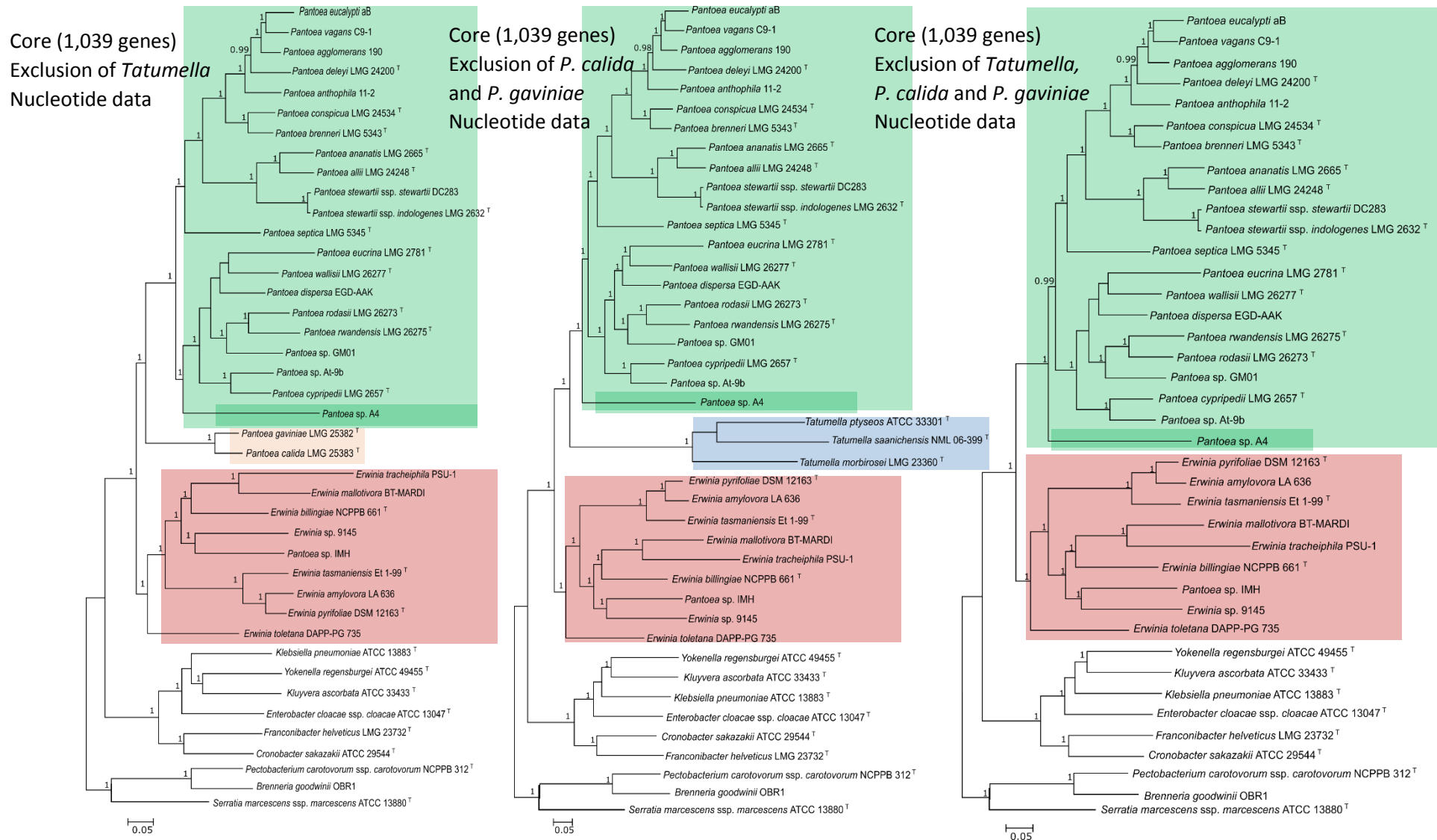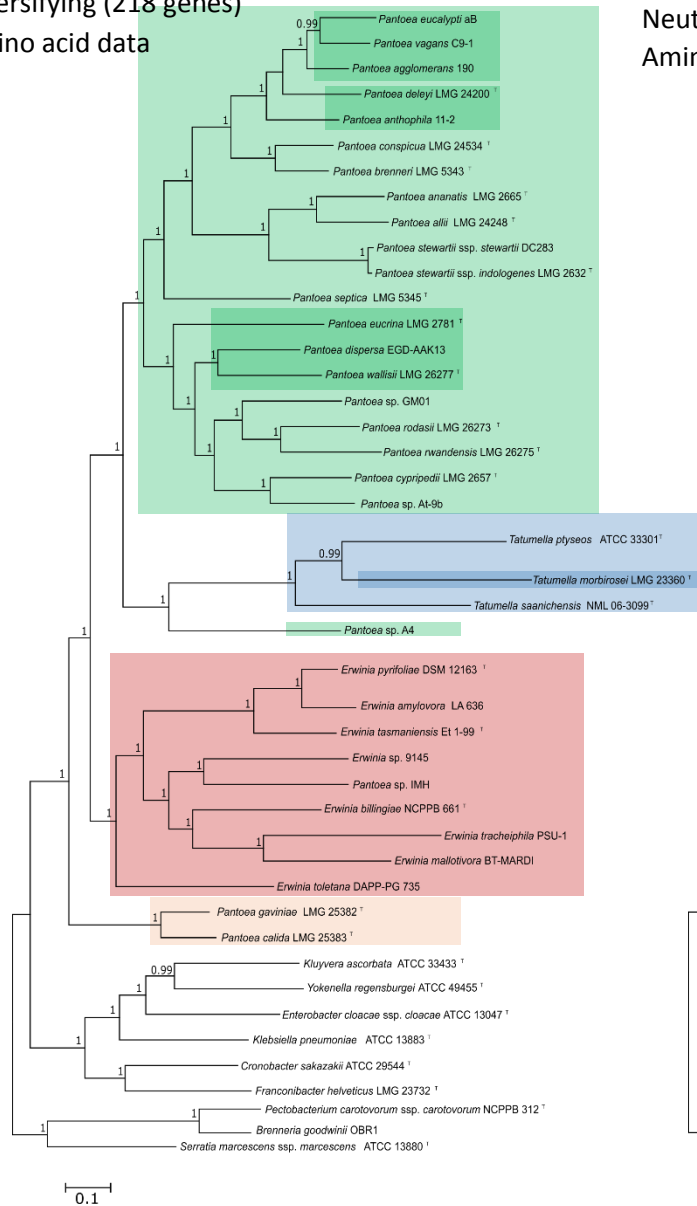
Supplementary Figure 4. Nucleotide based maximum likelihood trees with the exclusion of *Tatumella*, *Pantoea calida* and *Pantoea gaviniae*, and *Tatumella* with *Pantoea calida* and *Pantoea gaviniae*, respectively. SH-support values were calculated in FastTree and are indicated at the nodes. The position of *Pantoea* sp. A4 differs with the inclusion of *Pantoea calida* and *Pantoea gaviniae*. Darker blocks indicate differences between the relationships of the ingroup taxa.
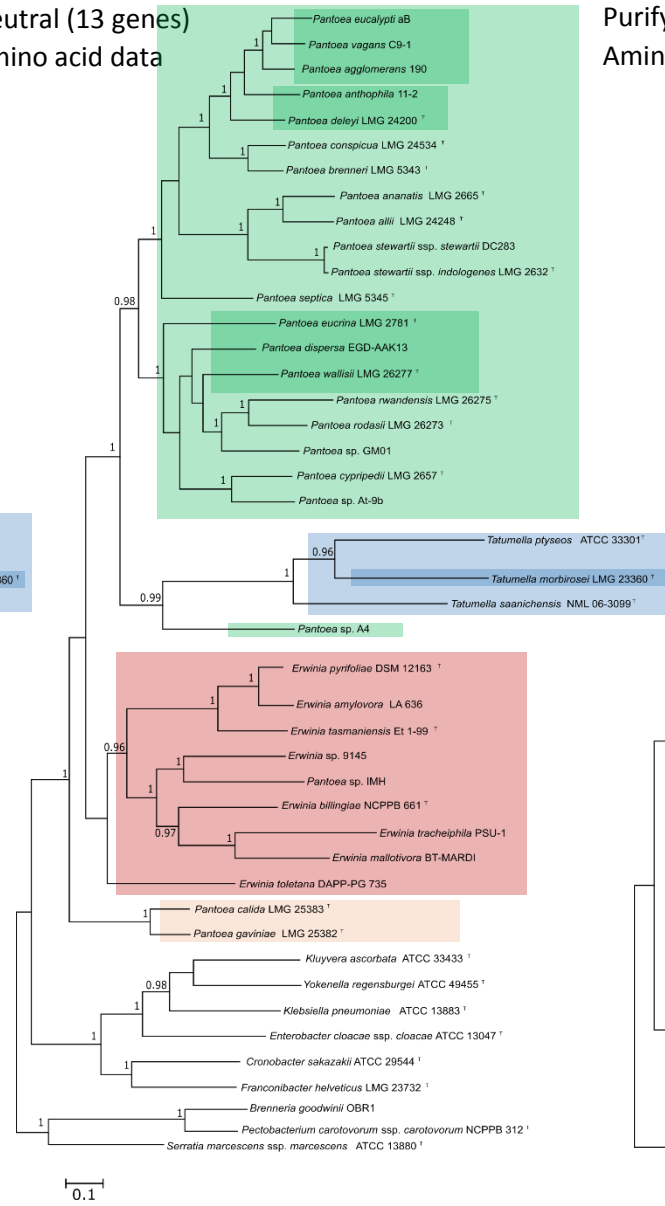
Supplementary Figure S5. Gene-wide dN/dS values indicated as a radial line graph with functional categories of core genes indicated in colour on the inner track. The blue circle indicate a dN/dS value of 1.
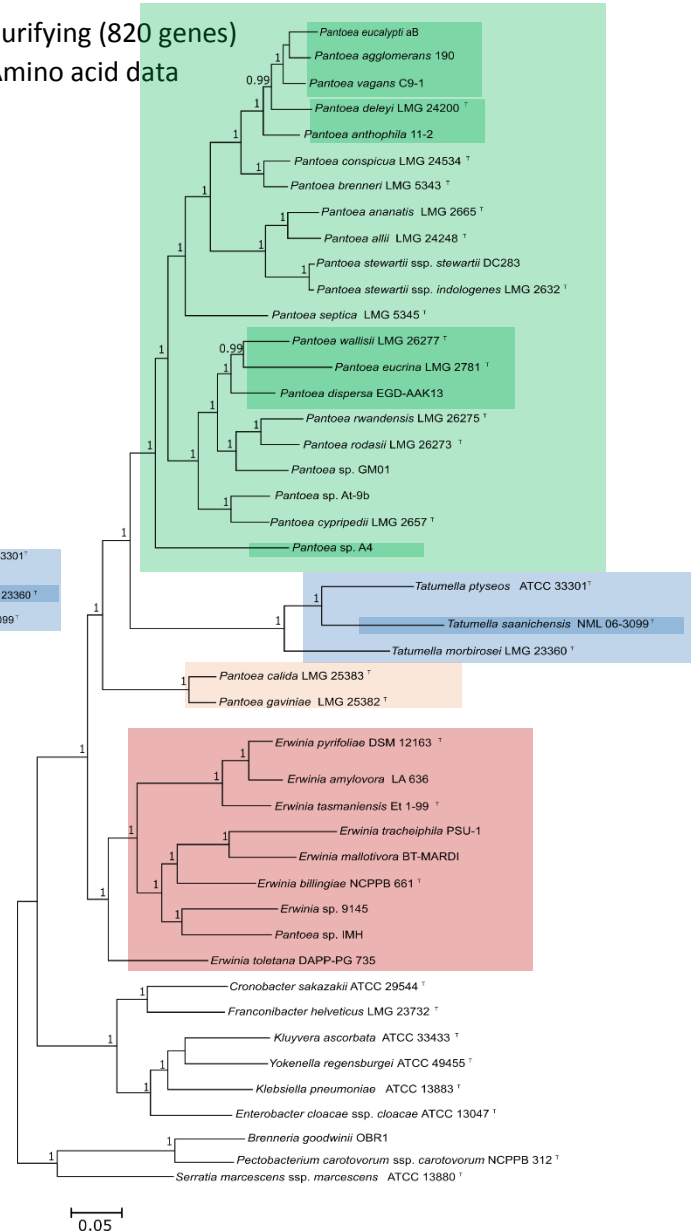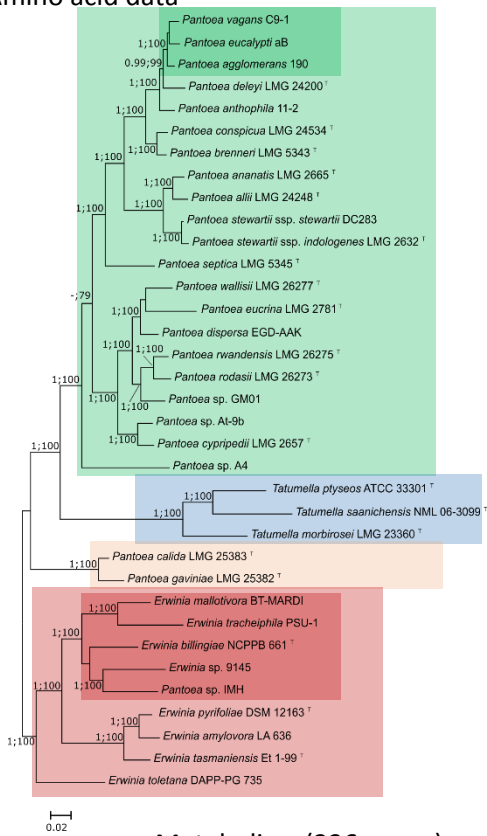
Supplementary Figure S6. Comparison of all nucleotide selection subset trees constructed with FastTree. Darker blocks indicate differences between the relationships of the ingroup taxa in the different topologies.
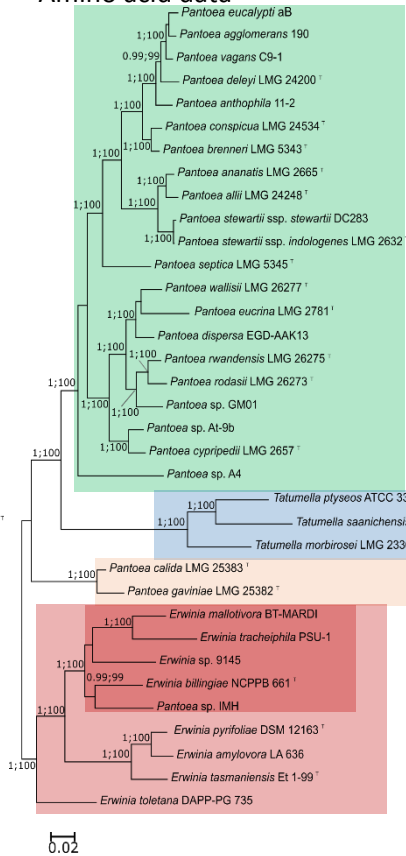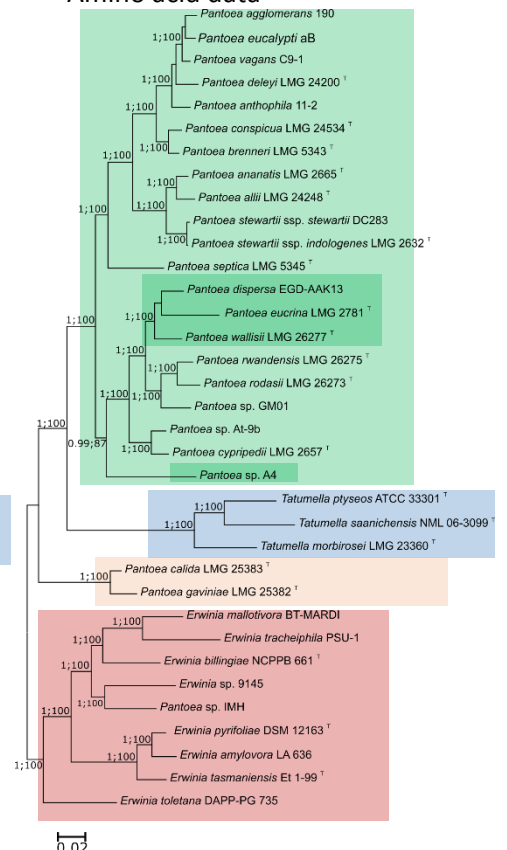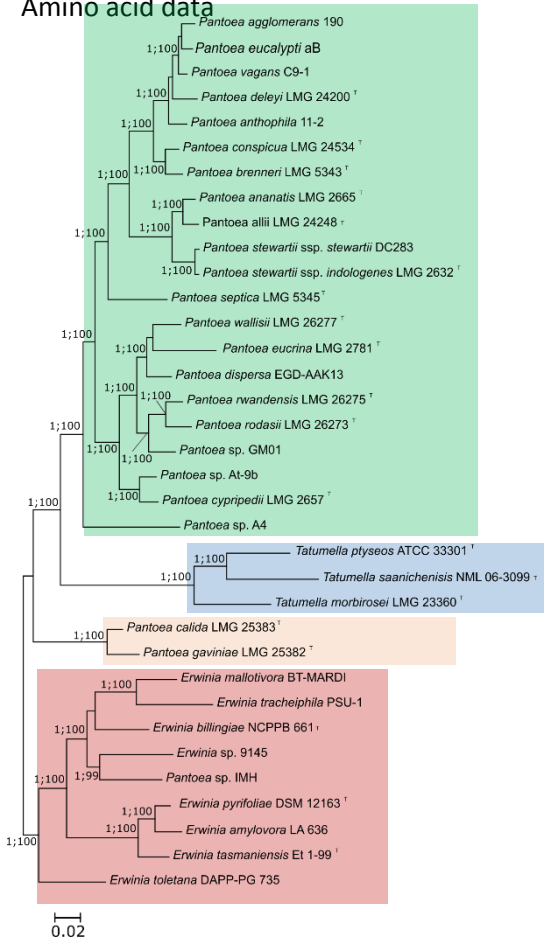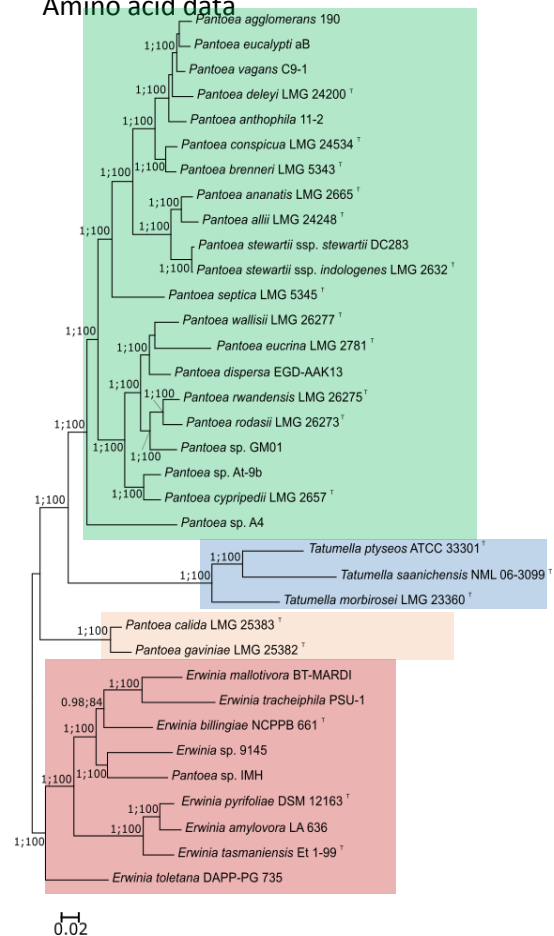
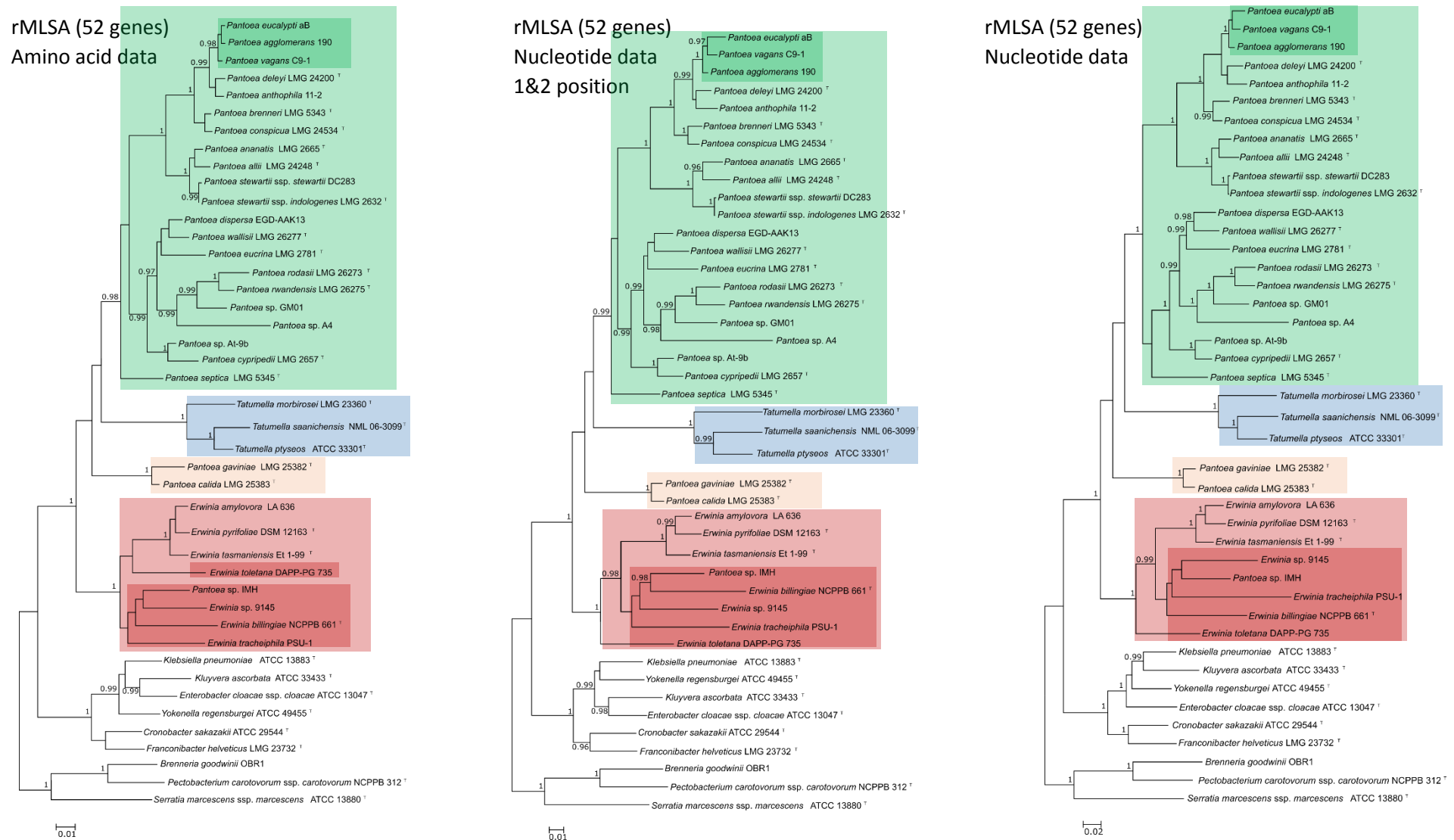**Supplementary Figure S7.** Comparison of all amino acid functional subset trees constructed with FastTree. Darker blocks indicate differences in the relationships of the ingroup taxa in comparison to the amino acid 1,039-gene tree.
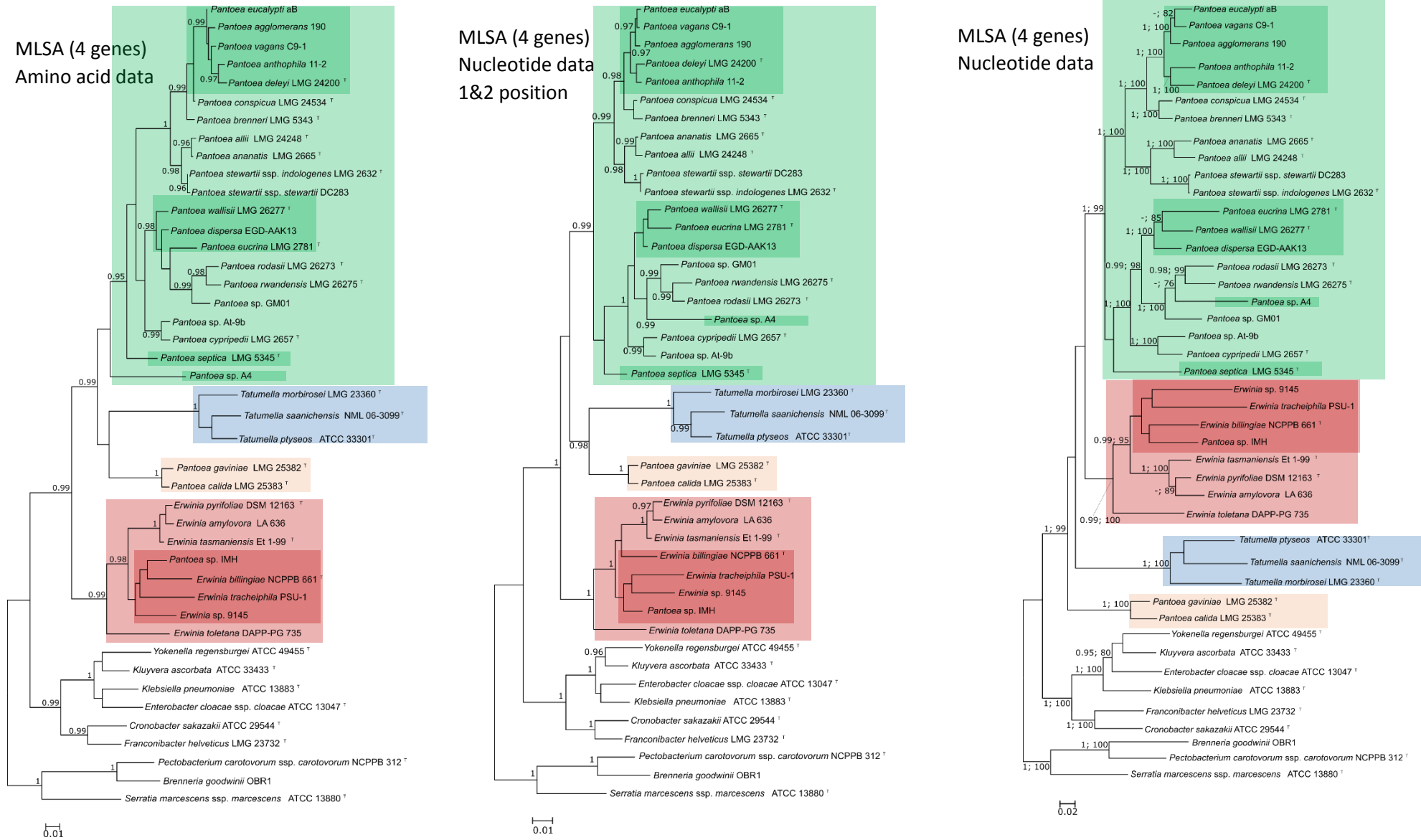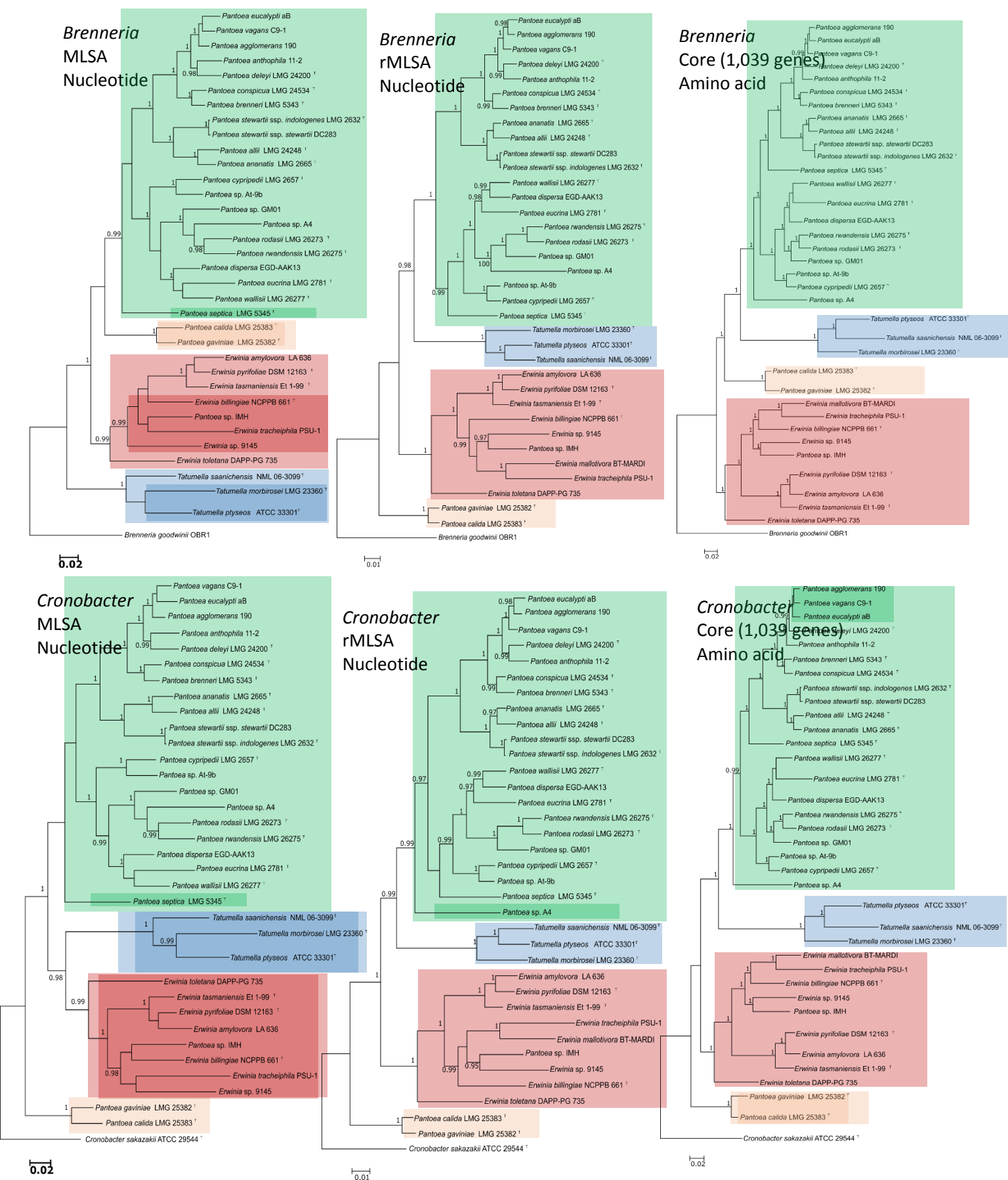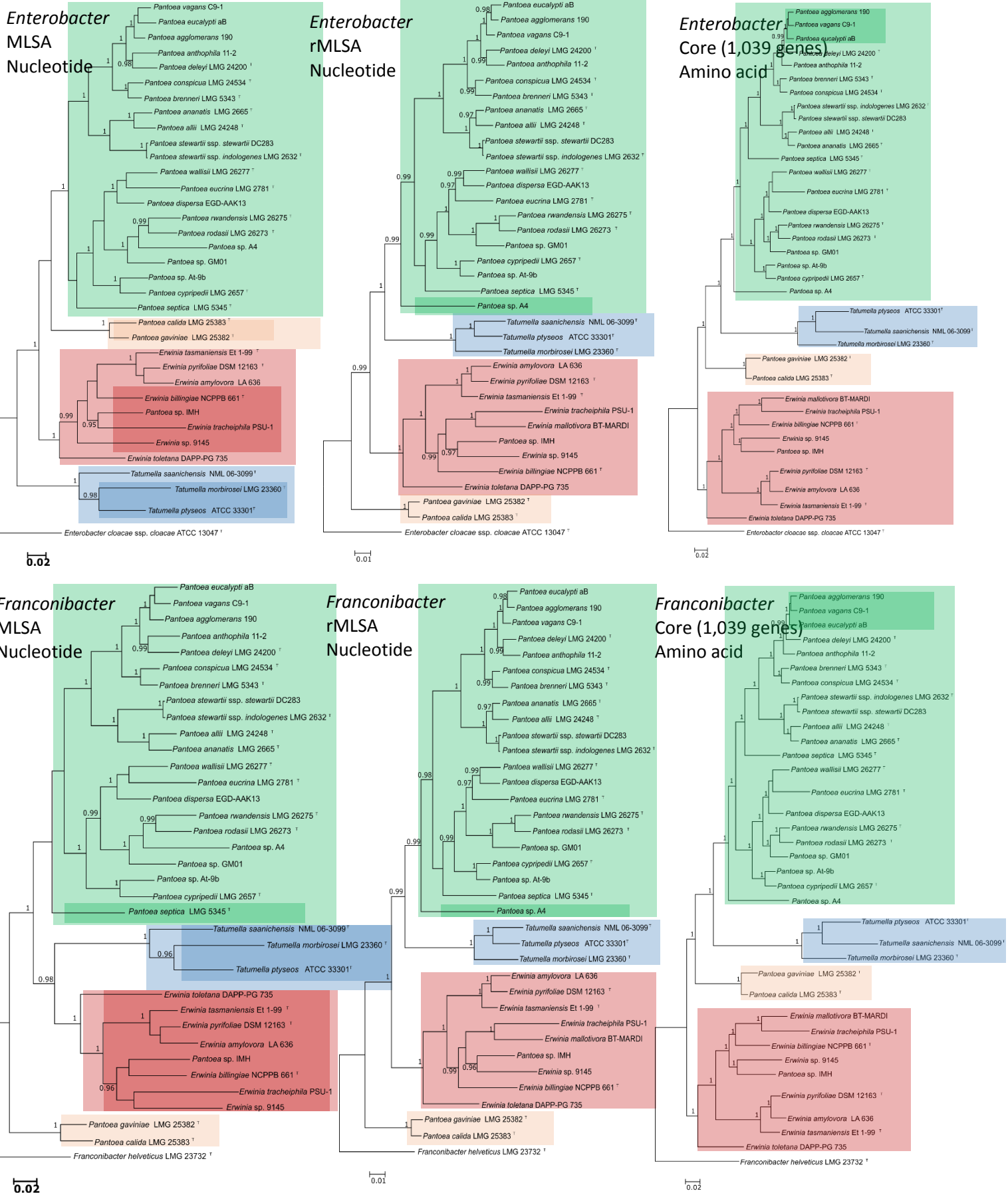
Supplementary Figure S8. Comparison between the maximum likelihood MLSA trees constructed with RAxML with branch support from FastTree (amino acid sequences, 3rd nucleotide excluded, full nucleotide dataset). Darker blocks indicate differences between the relationship of the ingroup taxa when compared to each other.

Supplementary Figure S9. Comparison between the maximum likelihood rMLSA trees constructed with RAxML with branch support from FastTree (amino acid sequences, 3rd nucleotide excluded, full nucleotide dataset). Darker blocks indicate differences between the relationships of the ingroup taxa when compared amongst each other.
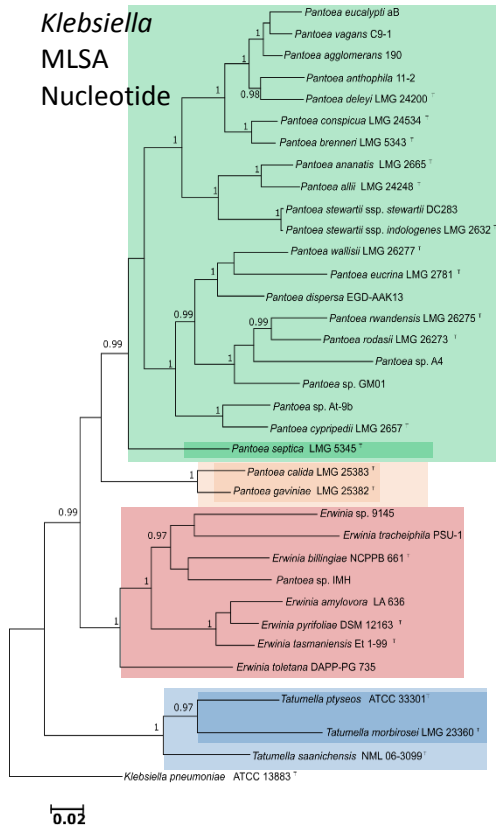
Supplementary Figure S10. The MLSA (nucleotide), rMLSA (nucleotide) and 1,039-gene core (amino acid) datasets with each of the different outgroup taxa respectively. Darker blocks indicate differences within the ingroup relationships between the single outgroup trees and the respective nucleotide or amino acid topologies obtained from the original dataset.
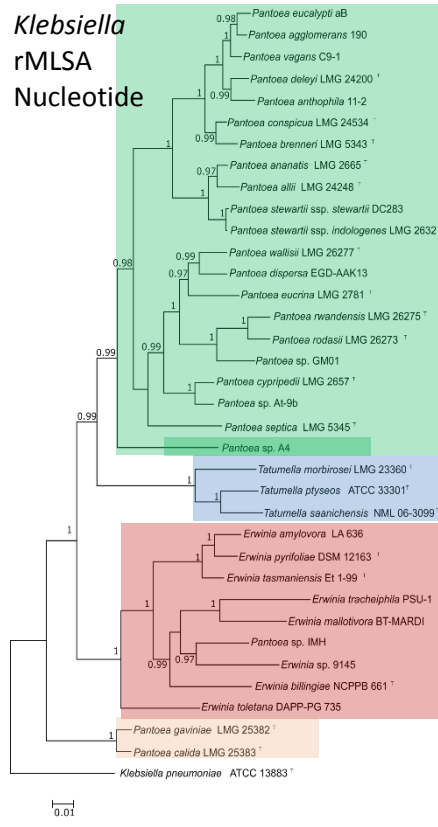
Supplementary Figure S10. The MLSA (nucleotide), rMLSA (nucleotide) and 1,039-gene core (amino acid) datasets with each of the different outgroup taxa respectively. Darker blocks indicate differences within the ingroup relationships between the single outgroup trees and the respective nucleotide or amino acid topologies obtained from the original dataset.
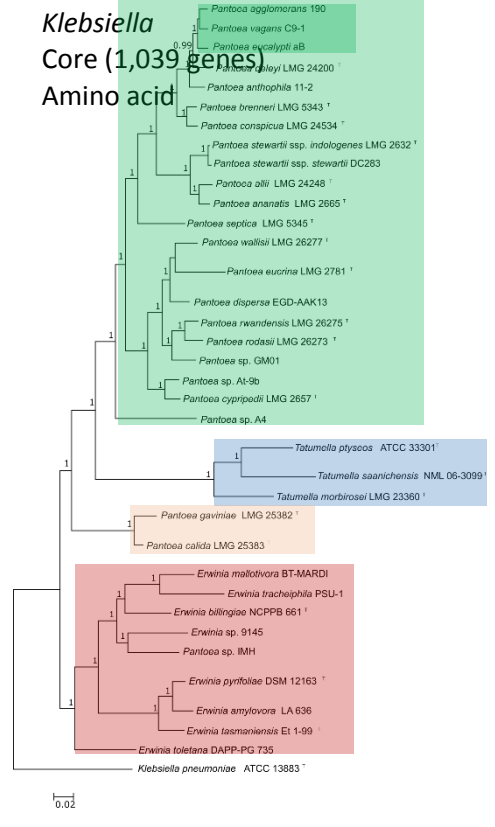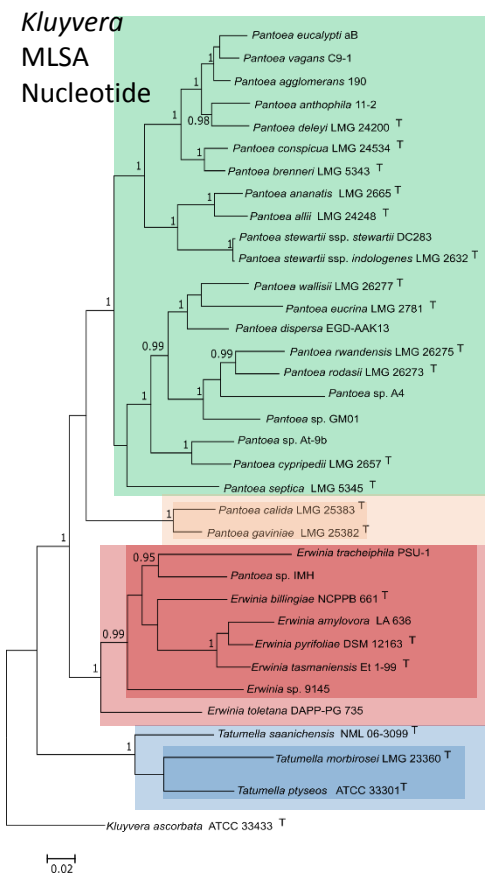
Supplementary Figure S10. The MLSA (nucleotide), rMLSA (nucleotide) and 1,039-gene core (amino acid) datasets with each of the different outgroup taxa respectively. Darker blocks indicate differences within the ingroup relationships between the single outgroup trees and the respective nucleotide or amino acid topologies obtained from the original dataset.
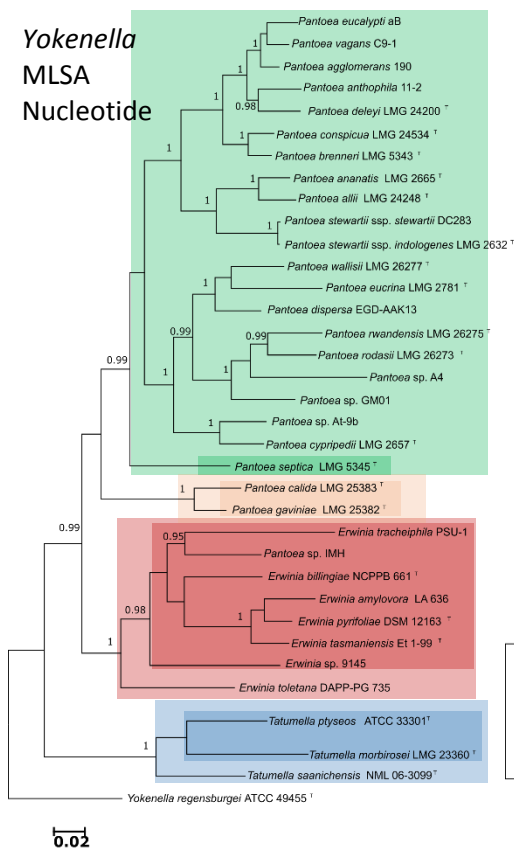
Supplementary Figure S10. The MLSA (nucleotide), rMLSA (nucleotide) and 1,039-gene core (amino acid) datasets with each of the different outgroup taxa respectively. Darker blocks indicate differences within the ingroup relationships between the single outgroup trees and the respective nucleotide or amino acid topologies obtained from the original dataset.
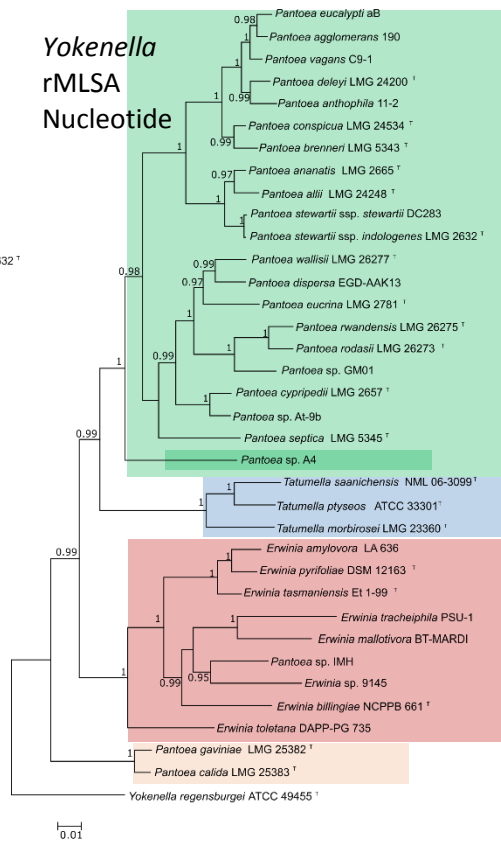
Supplementary Figure S10. The MLSA (nucleotide), rMLSA (nucleotide) and 1,039-gene core (amino acid) datasets with each of the different outgroup taxa respectively. Darker blocks indicate differences within the ingroup relationships between the single outgroup trees and the respective nucleotide or amino acid topol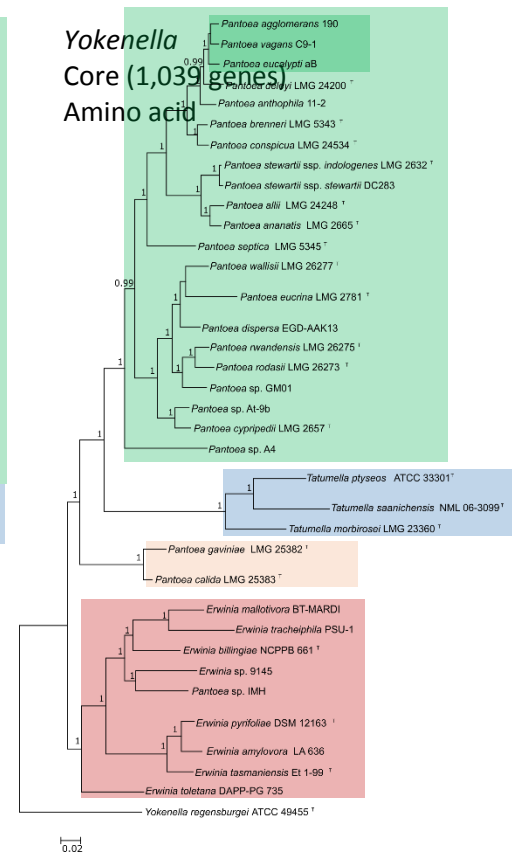ogies obtained from the original dataset.