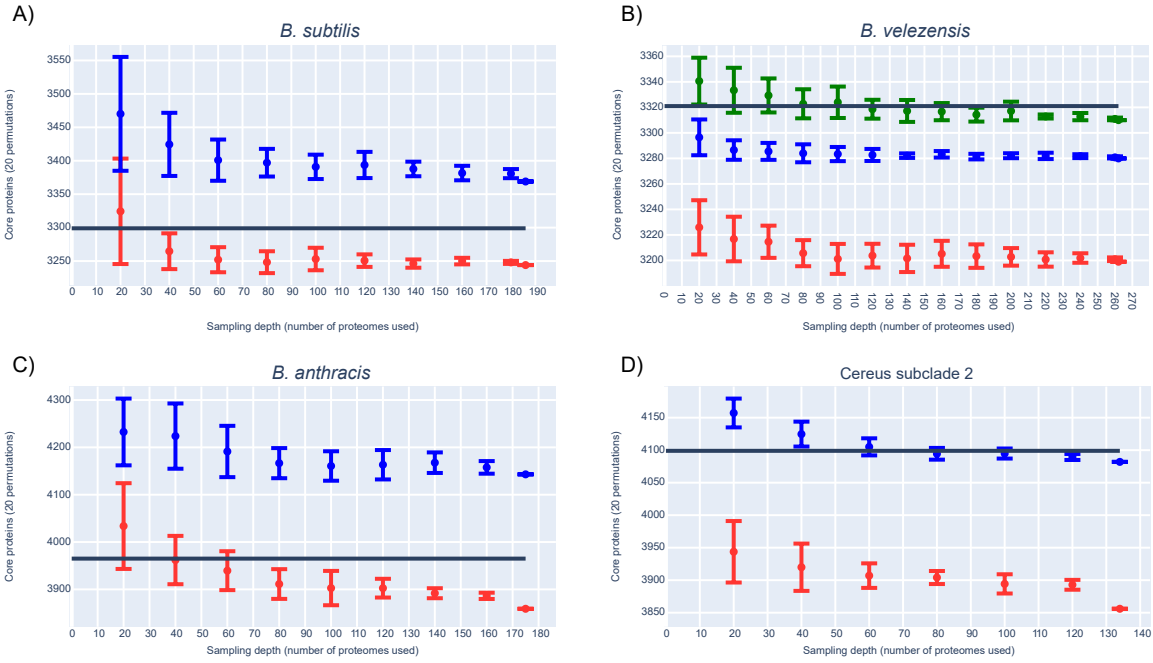
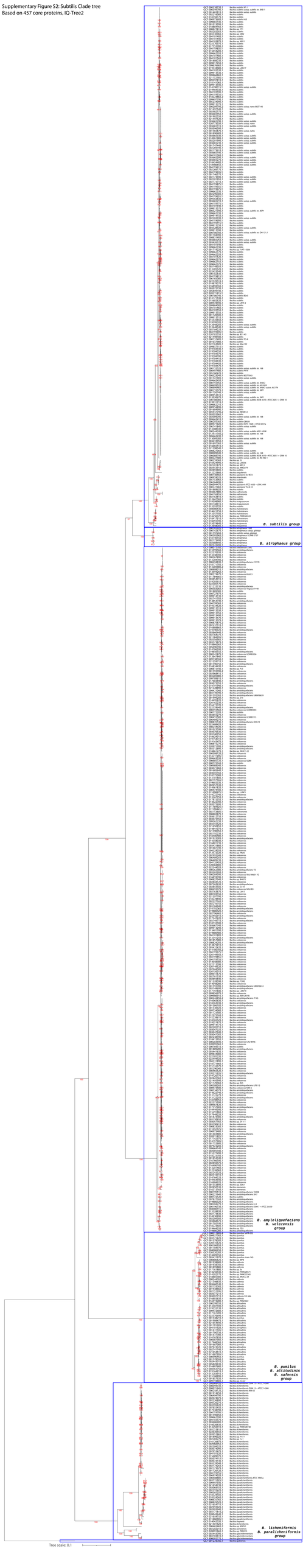
Supplementary Figure S1

Saturation curves of soft core proteins at 85%, 90% and 95% depending on the number of genomes sampled

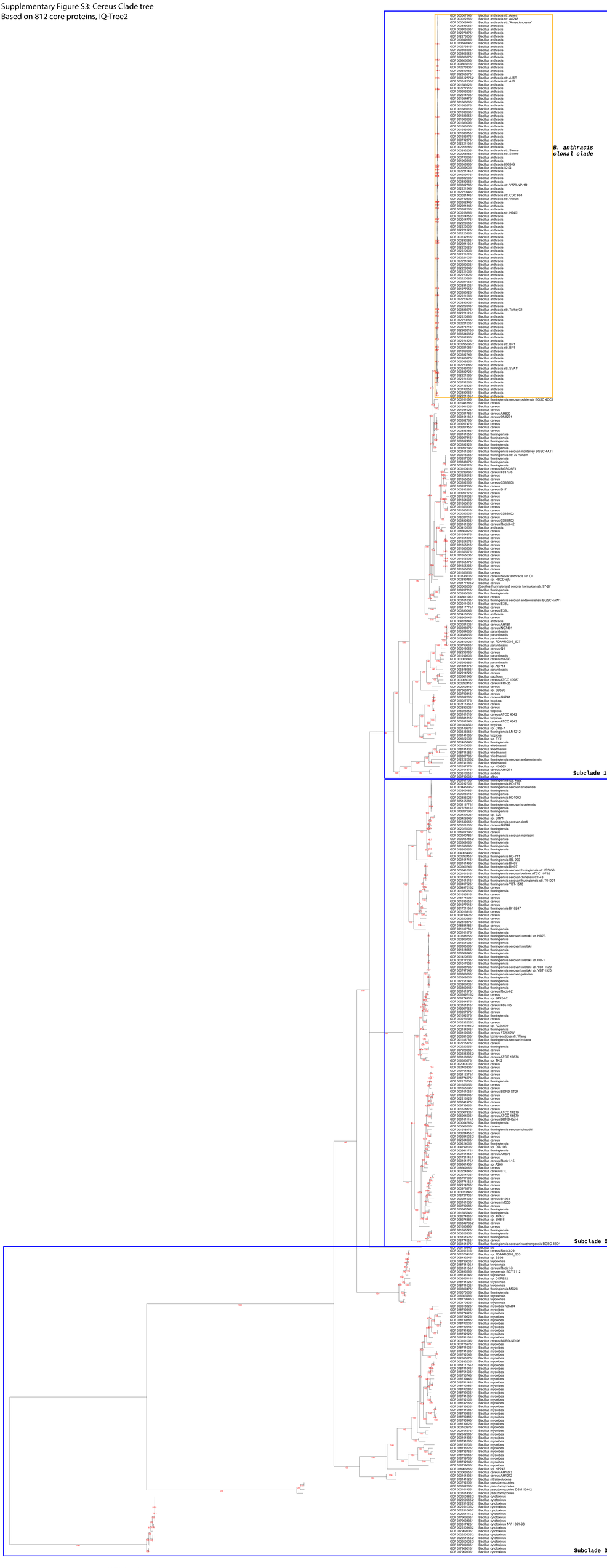


Black bar corresponds to the number of core proteins calculated from the normalized dataset (5 organisms per species)

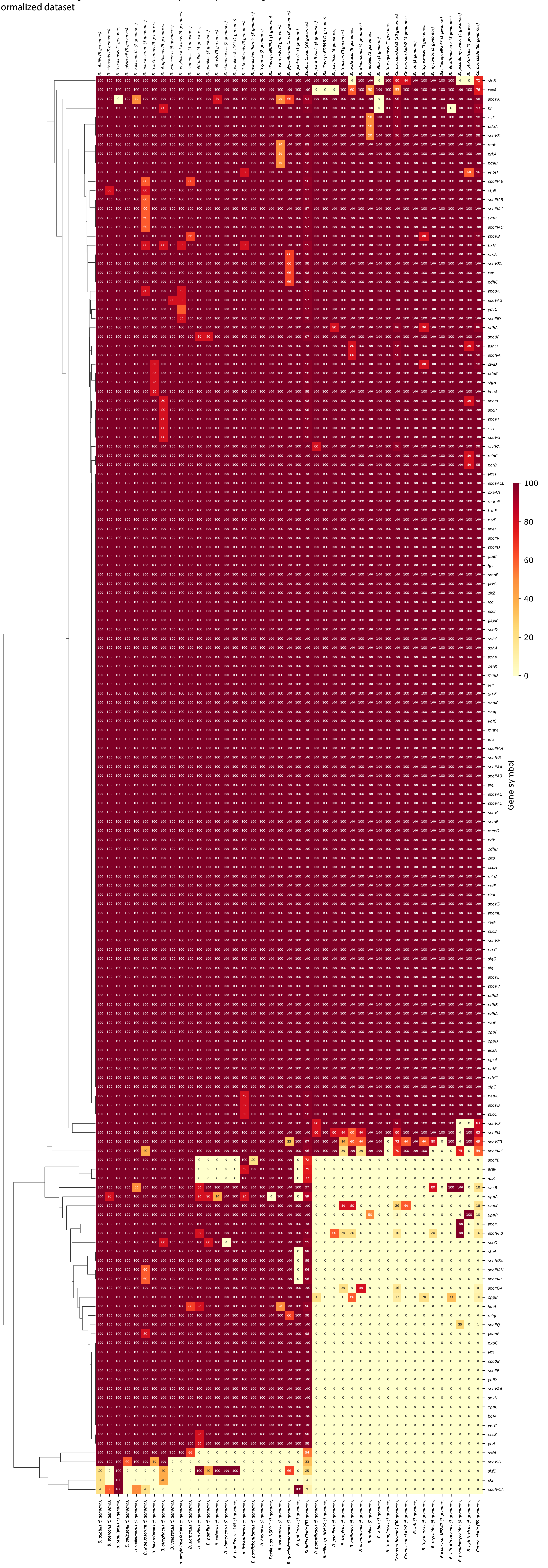
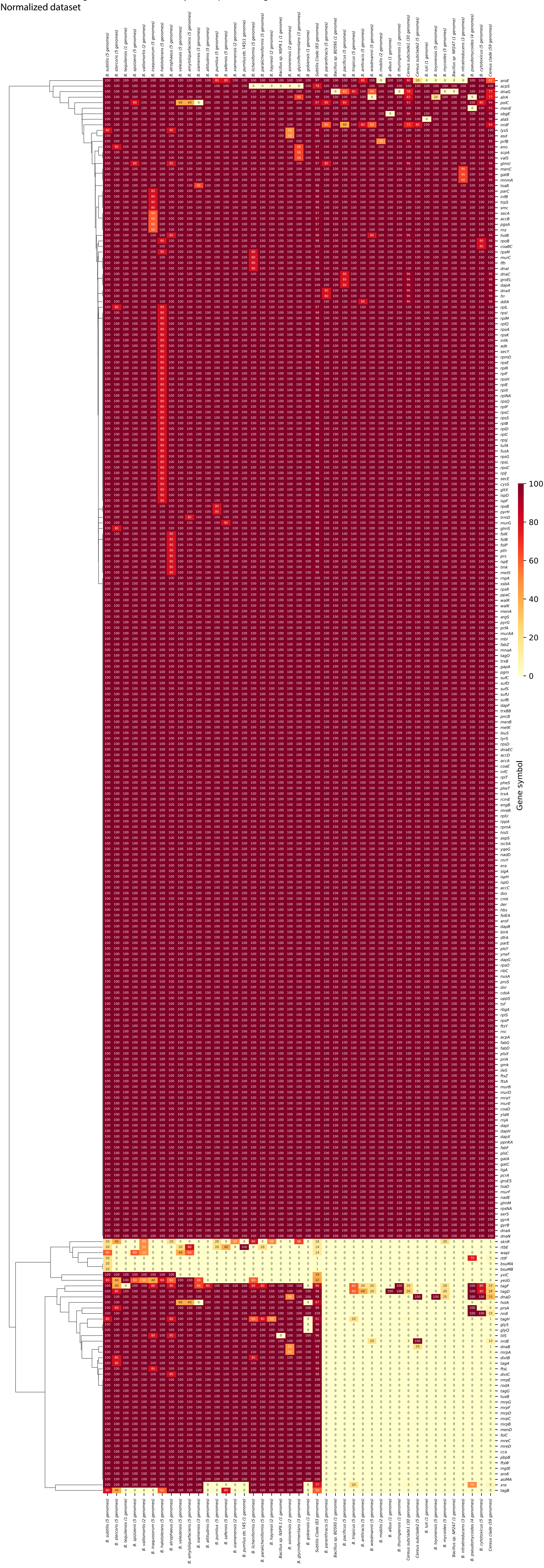
Supplementary Figure S2: Subtilis Clade tree
Based on 457 core proteins, IQ-Tree2



Supplementary Figure S3: Cereus Clade tree
Based on 812 core proteins, IQ-Tree2



Supplementary Figure S4: Sporulation gene homologues
Presence of homologues with 50% aa identity at 50% protein length
Normalized dataset



Supplementary Figure S5: Essential gene homologues
Presence of homologues with 50% aa identity a 50% protein length
Normalized dataset