

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Mapping of the seven AFLP clusters defined by (1) onto the weighted “MLST+AFLP” supertree of 2143 *B. cereus* group isolates. The supertree is based on phylogenetic information from five MLST schemes and three AFLP studies and was inferred following the Matrix Representation by Parsimony (MRP) strategy, with a weight of 40 or 68 given to the AFLP studies and a weight of 1 given to each of 26 MLST genes. For the sake of legibility isolate names are not shown. The seven major phylogenetic subdivisions defined by (1) are designated by romans numerals (I-VII) and 414 of the 425 *B. cereus* group isolates analyzed by AFLP in (1) are specifically colored (the remaining 11 isolates have conflicting phylogenetic data and are not included in HyperCAT analyses). Isolates belonging to a given subdivision are drawn in the same color. As can be seen, isolates from the same subdivision are all contained within a common monophyletic cluster in the supertree, with no mixing of isolates from different subdivisions, indicating that the overall structure of the AFLP tree from (1) is retained in the “MLST+AFLP” supertree. Branch support values (based on aLRT probabilities) for the major phylogenetic clusters are indicated when >95%. The image was generated using TreeDyn (2).

Supplementary Figure 2. Mapping of the ten AFLP clusters defined by (3) onto the weighted “MLST+AFLP” supertree of 2143 *B. cereus* group isolates based on phylogenetic information from five MLST schemes and three AFLP studies. For the sake of legibility isolate names are not shown. The seven major phylogenetic subdivisions defined by (1) are designated by romans numerals (I-VII). The 10 phylogenetic branches (A-K) defined by (3) based on AFLP analysis have been mapped onto the supertree and 318 of the 332 *B. cereus* group isolates analyzed in that study are specifically colored (the remaining 14 isolates have conflicting phylogenetic data and are not included in HyperCAT analyses). Isolates belonging to a given AFLP branch are drawn in the same color. As can be seen, isolates from the same AFLP group are all contained within a common monophyletic cluster in the supertree, with basically no mixing of isolates from different AFLP branches, with the exception of isolates from branches E and G which are included within branch F. The overall relationships between branches A-K in the supertree are also similar to that in the AFLP tree of (3). Altogether, this indicates that the overall structure of the AFLP

tree of (3) is retained in the “MLST+AFLP” supertree. Branch support values (based on aLRT probabilities) for the major phylogenetic clusters and AFLP branches are indicated when >95%. The image was generated using TreeDyn (2).

Supplementary Figure 3. Mapping of the nine MLST clusters defined by (4) onto the weighted “MLST+AFLP” supertree of 2143 *B. cereus* group isolates based on phylogenetic information from five MLST schemes and three AFLP studies. For the sake of legibility isolate names are not shown. The seven major phylogenetic subdivisions defined by (1) are designated by romans numerals (I-VII). The 9 phylogenetic clades defined by (4) based on MLST analysis have been mapped onto the supertree and a set of 257 *B. cereus* group isolates belonging to the 59 sequence types (STs) described in that study were taken from the PubMLST (<http://pubmlst.org/bcereus/>) database and are specifically colored. Isolates belonging to a given MLST clade are drawn in the same color. As can be seen, isolates from the same MLST group are all contained within a common monophyletic cluster in the supertree, with basically no mixing of isolates from different MLST clades, with the exception of some isolates from clade “Sotto” which are included within clade “Kurstaki”. The overall relationships between clades in the supertree are also similar to that in the MLST tree of (4). Altogether, this indicates that the overall structure of the MLST tree of (4) is retained in the “MLST+AFLP” supertree. Branch support values (based on aLRT probabilities) for the major phylogenetic clusters and MLST clades are indicated when >95%. The image was generated using TreeDyn (2).

Supplementary Figure 4. Impact of missing data on supertree reconstruction. The impact of missing data was tested by including in the supertree four replicates of the 34 strains for which MLST, AFLP, and MLEE data are available, using either all data or data from one of the three typing methods. The strain replicates are included in a weighted “MLST+AFLP+MLEE” supertree containing a total of 2315 *B. cereus* group isolates based on phylogenetic information from five MLST schemes, three AFLP studies, and one MLEE study. The supertree was inferred following the Matrix Representation by Parsimony (MRP) strategy, with a weight of 40 or 68 given to the AFLP studies, a weight of 1 given to each of 26 MLST genes, and a weight of 3 to MLEE data. For the sake of legibility isolate names are not shown. The seven major

phylogenetic subdivisions defined by (1) are designated by romans numerals (I-VII). The replicates of a given strain are drawn in the same color. Panel A shows the results for 17 of the 34 strains, while the other 17 strains are shown in panel B. As can be seen, for all 34 strains, the four replicates were always located in the same phylogenetic cluster and were mostly located in the same subtree (for 28 of the 34 strains the replicates were separated by less than a dozen internal nodes), indicating that missing data did not bias the general positioning of the strains. However, variations within clusters did occur, suggesting that missing data may occasionally limit the resolution of precise relationships within clusters. The image was generated using TreeDyn (2).

1. Guinebretière, M.H., Thompson, F.L., Sorokin, A. *et al.* (2008) Ecological diversification in the *Bacillus cereus* Group. *Environ Microbiol*, **10**, 851-65.
2. Chevenet, F., Brun, C., Banuls, A.L. *et al.* (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics*, **7**, 439.
3. Hill, K.K., Ticknor, L.O., Okinaka, R.T. *et al.* (2004) Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. *Appl Environ Microbiol*, **70**, 1068-80.
4. Priest, F.G., Barker, M., Baillie, L.W. *et al.* (2004) Population structure and evolution of the *Bacillus cereus* group. *J Bacteriol*, **186**, 7959-70.