**Supple Data 2.5. Algorithm for outgroup database sequence sorting** (Fig. S2.5)

The most appropriate outgroup, denoted as *t*, is automatically selected from the database. It is the closest group of sequences to the assigned taxonomic group, while still being distinctly different from assigned ingroup taxonomic group. Let *tk* represent the sequences of the outgroup taxonomic group, *ix* the sequence of the ingroup taxonomic group, *Bpq* the bitscore value between sequences *p* and *q*, and *c* the outgroup offset. The outgroup taxonomic group *t* was chosen based on the following criterion:



**Fig. S2.5.** Infographics illustrating the algorithm for outgroup database sequence sorting.