**Supple Data 2.4. Algorithm for suspicious database sequence sorting**

Suspicious sequences, which may be phylogenetically placed within the ingroup but are not annotated as such, are included in the sequence-set. This process is implemented to prevent potential misinterpretations of the phylogenetic tree that may arise from missing sequences (Fig. S2.5). We defined suspicious sequences as those that cannot be confidently classified as either ingroup or outgroup. Let *ak* denote an suspicious sequence, and *a* represent the taxonomic group of the suspicious sequence. The set of suspicious sequences, *ak* is determined using the following condition:

 

**Fig. S2.4.** Infographics illustrating the algorithm for suspicious database sequence sorting.(A) Assumed scenario for suspicious sequence inclusion. (B) Comparison of the effects when suspicious sequences are either included or excluded from the sequence set.