Supplementary Materials

Supplementary Table 1. Excel file of 3,313 putative PKS sequences (spanning 2,786 NCBI sequence records) identified using the BLAST-based scan for ≥3 clustered KS domains.

Supplementary Table 2. Putative PKSs identified by the BLAST scan but not antiSMASH2.

Supplementary Table 3. Excel file of all 1,236 identified non-identical PKS gene clusters, listing for each PKS: NCBI sequence accession number, Sequence description, Date deposited in NCBI, Species name, Cluster number within this sequence, Link to graphical Antismash view, Cluster coordinates, Cluster length (bp), Cluster GC content (%), Number of KSs, Number of ATs, Number of ACPs, Number of KRs, Number of DHs, Number of ERs, Number of Cs, Number of As, Interesting Antismash-predicted domains, Antismash-predicted cluster type, gene clusters with identical sequence, and gene clusters with identical domain architecture in this species (but with non-identical sequence).

Supplementary Table 4. The 1,236 PKS gene clusters defined in Supplementary Table 3 were refined to 885 non-redundant PKS gene clusters as defined by a 90% similarity score cutoff. The redundant clusters for each row are listed in the final column.

Supplementary Figure 1. Dendrogram comparing all PKS clusters. The label for each PKS lists the Genbank ID and cluster number, date the sequence was deposited in NCBI, number of KS, AT, C, and A domains, and sequence description. Scale bar displays the distance between gene clusters (between 0 and 1).

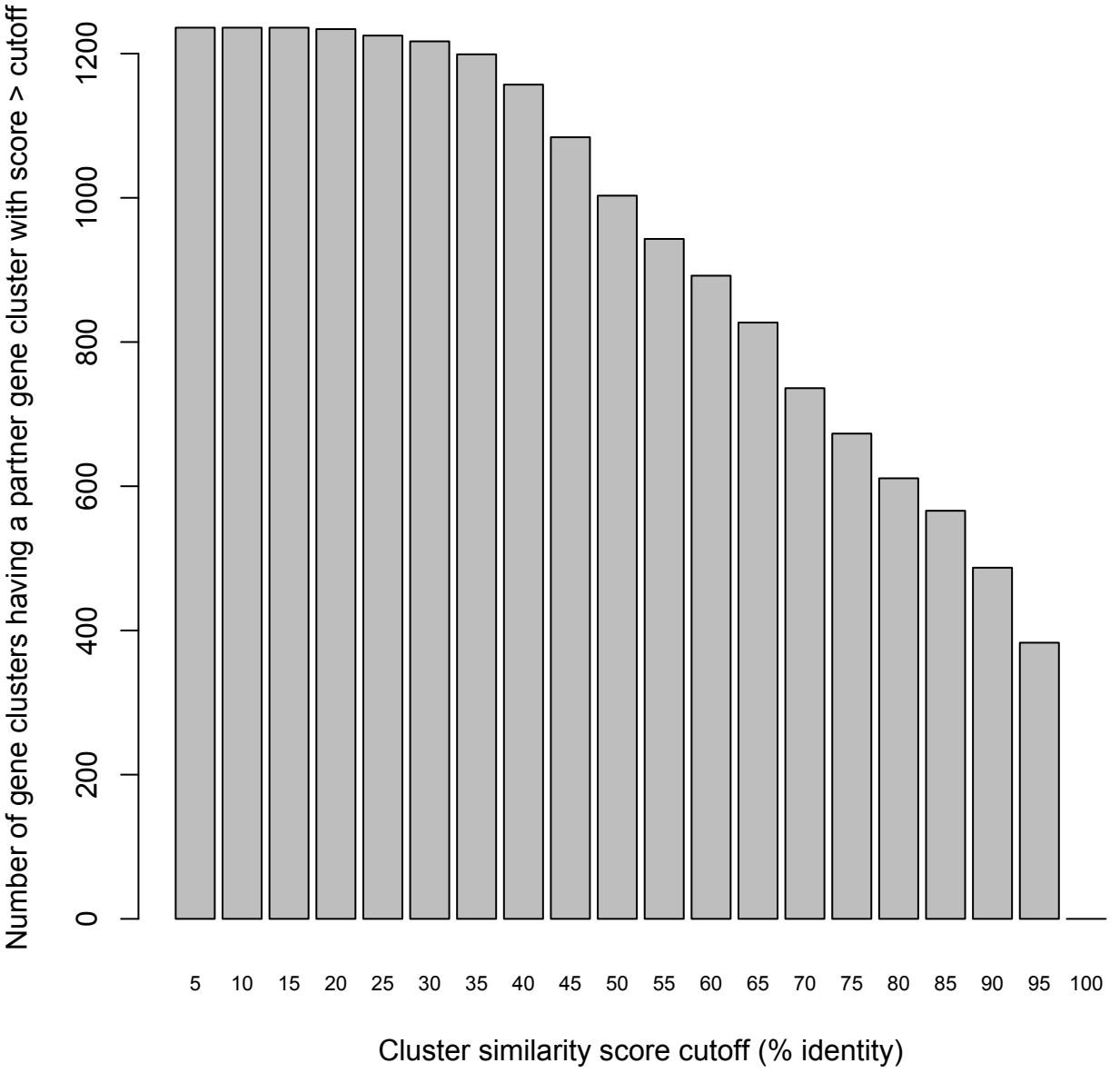
Supplementary Figure 2. Histograms of KS domains and GC content per gene cluster.

Supplementary Figure 3. Number of clusters having a partner gene cluster with similarity at the given cutoff. Gene clusters having a partner with similarity >90% were defined as redundant.

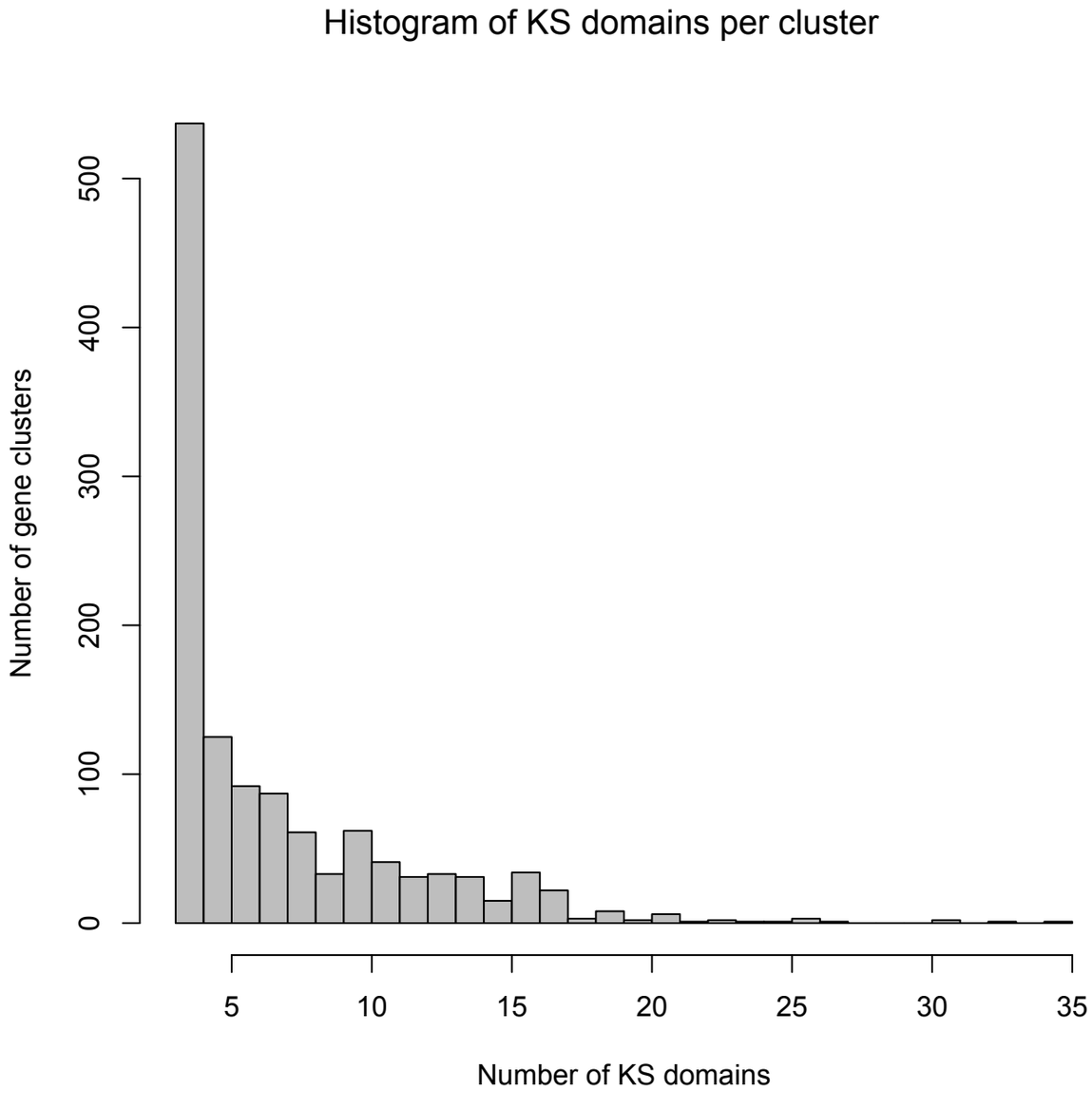
Supplementary Figure 4. Phylogenetic analysis of the KS domains from the 62 gene clusters in Figure 2. KS domains from the PKS-NRPS hybrid clade in Figure 2 are in color: KS domains from the trans-AT clade are colored red and KS domains from the cis-AT clade are colored blue. The remaining KS domains are in black.

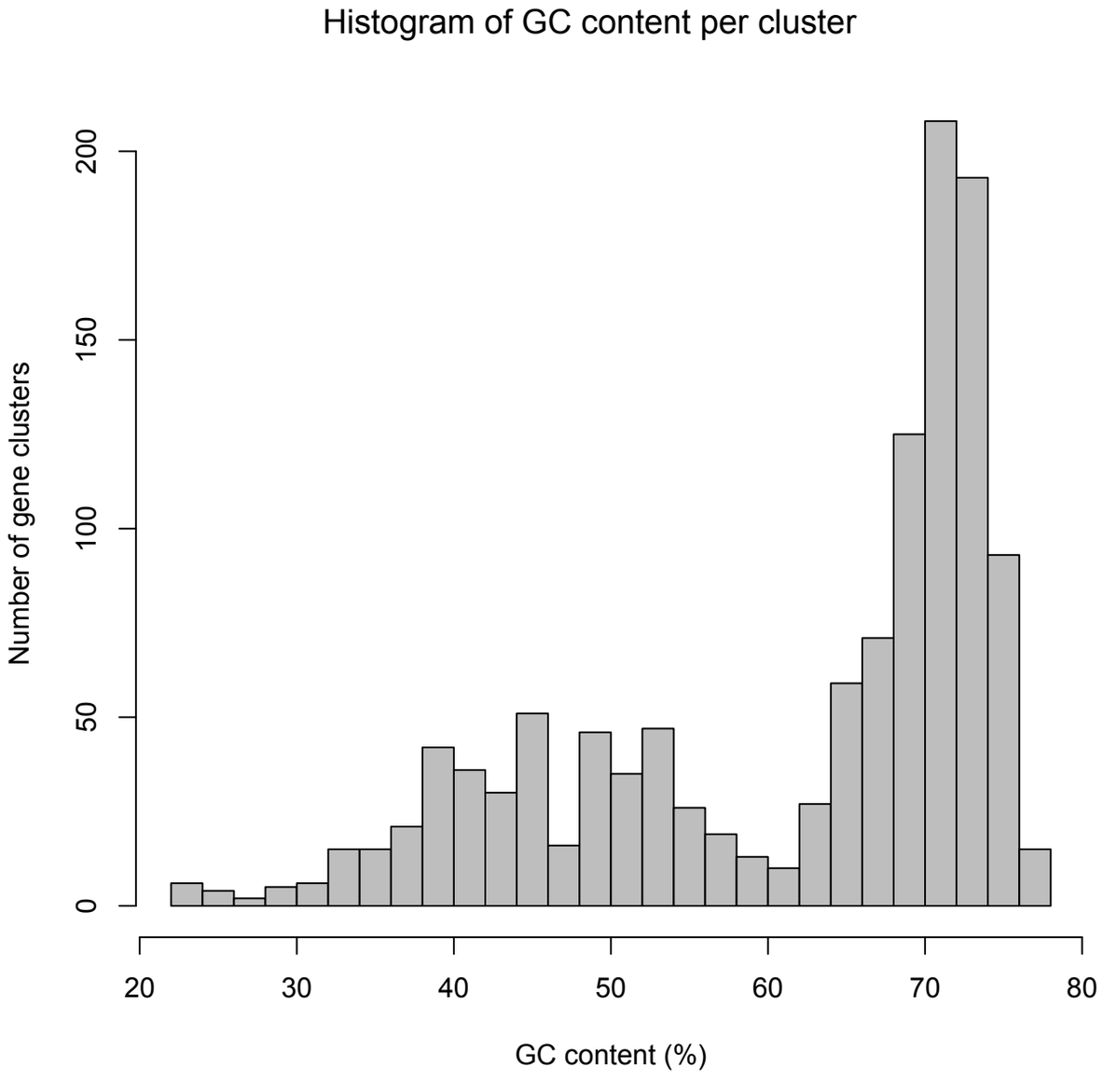
Supplementary Figure 5. (A.) Phylogenetic trees of domains of the *C. elegans* ("worm") cluster with bacterial domains and *C. elegans* fatty acid synthase (FASN) domains. Trees were created using clustalw, with default settings; shown are neighbour-joining trees without distance corrections. Abbreviations: SAV939 = avermectin cluster, dsz = disorazole cluster, epo = epothilone cluster, kir = kirromycin cluster, DEBS = erythromycin cluster, pik = pikromycin cluster, tyl = tylosin cluster. (B.) The conserved PKS genes in nematodes, clustered using our *ad hoc* PKS sequence similarity score. Here we show the putative domain architectures (as predicted by antiSMASH) for several homologs of the *C. elegans* PKS described in Figure 6. A site of possible duplication (as suggested by the KS phylogenetics in part A) is highlighted.

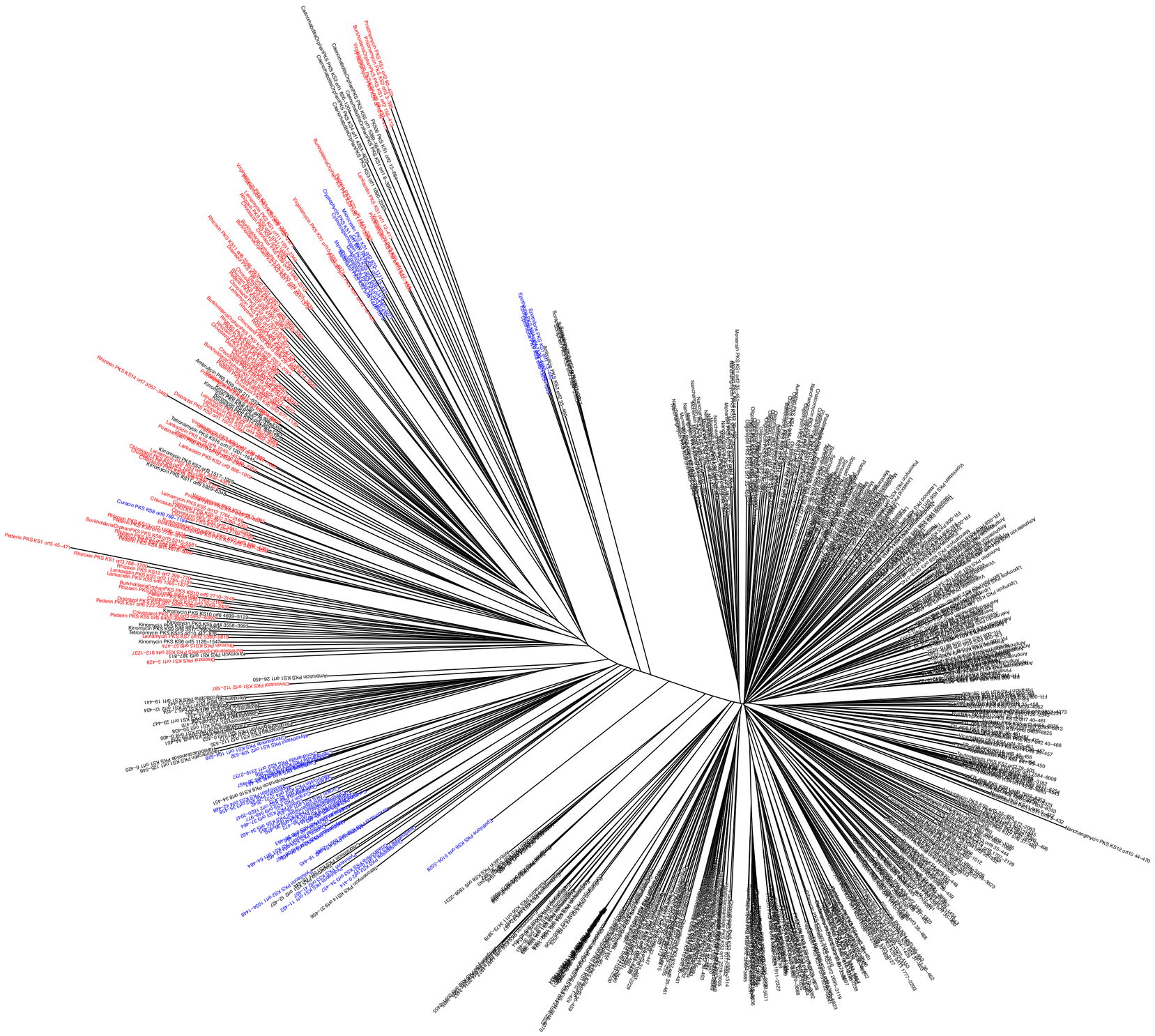
Supplementary Figure 3



Supplementary Figure 2

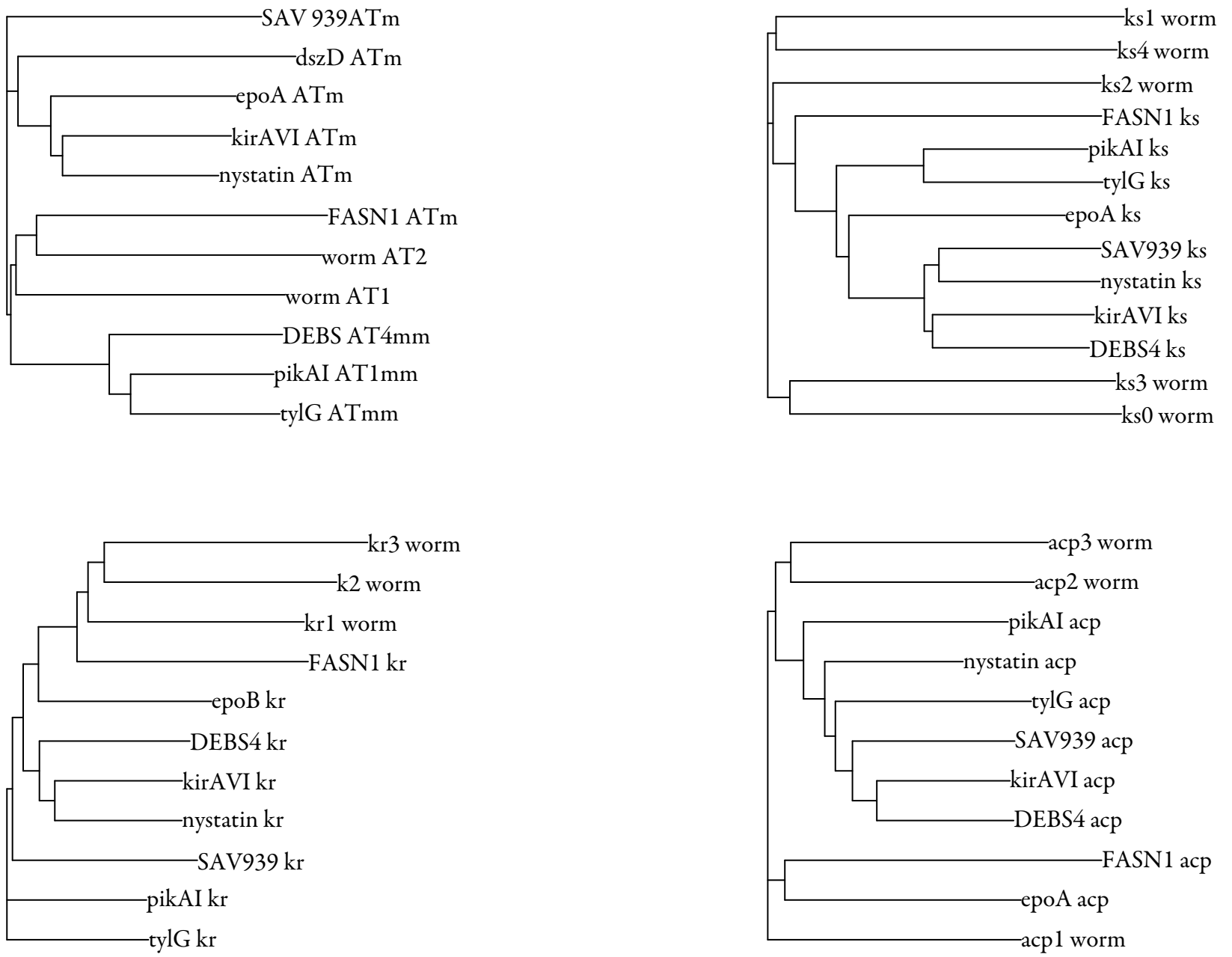
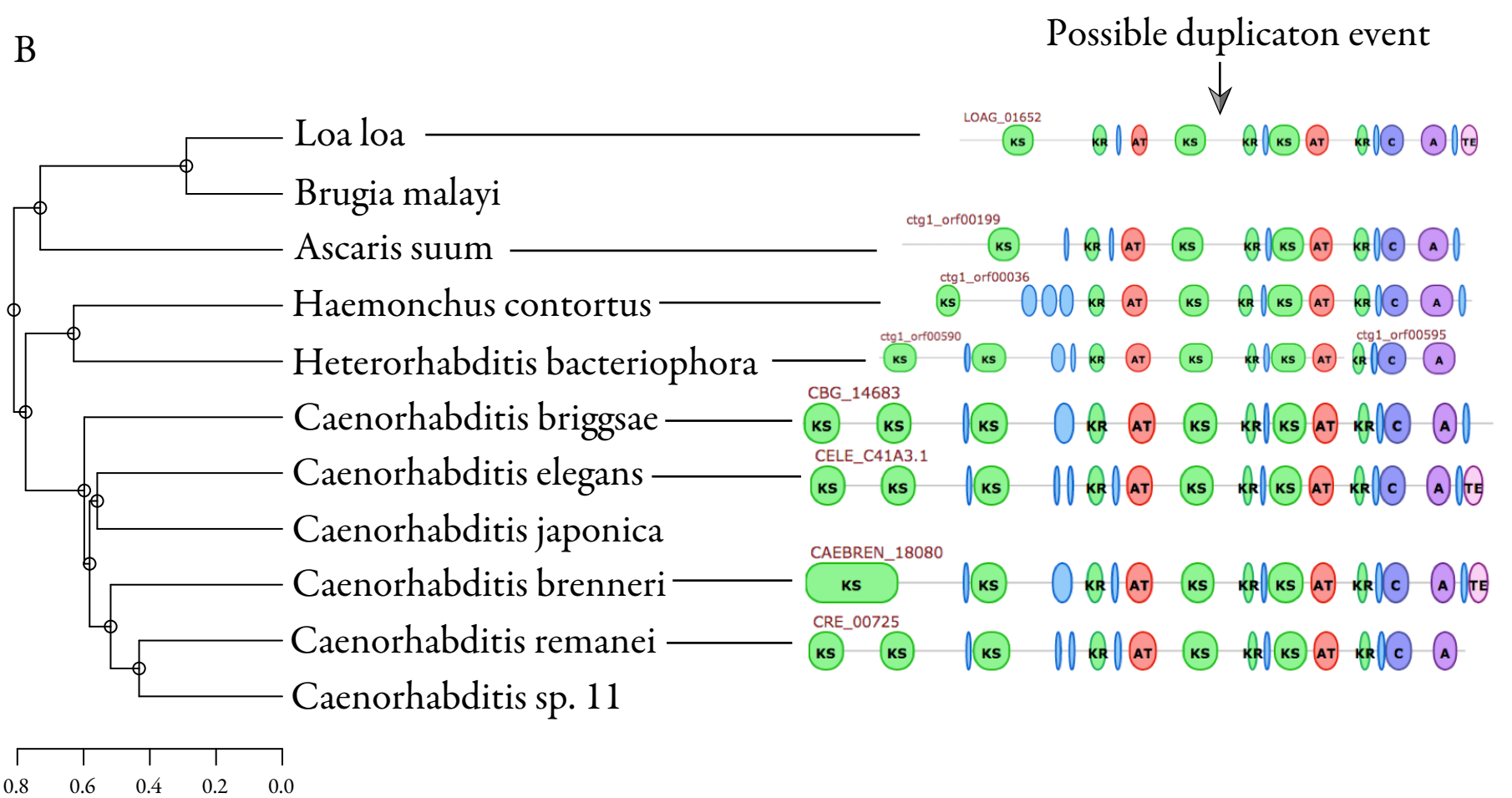






Supplementary Figure 4

A



B