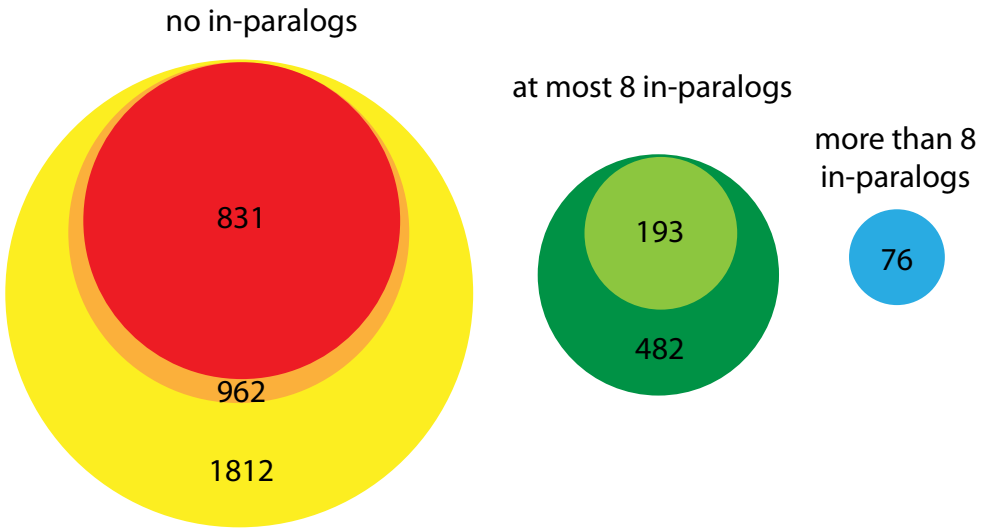
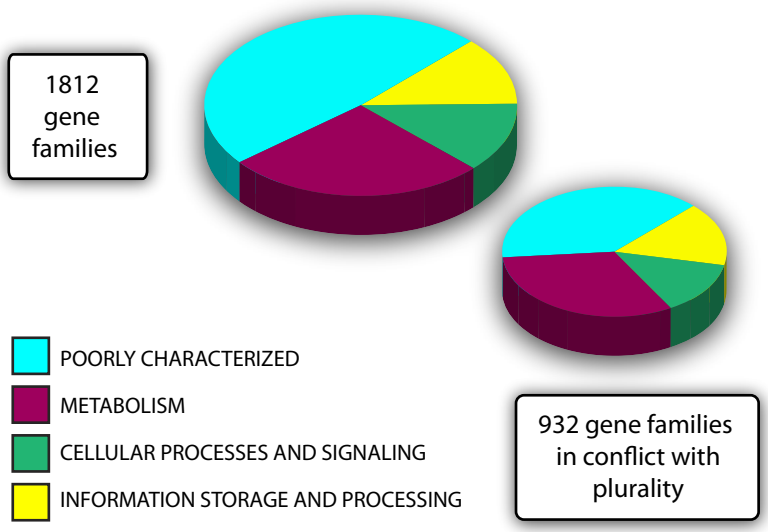
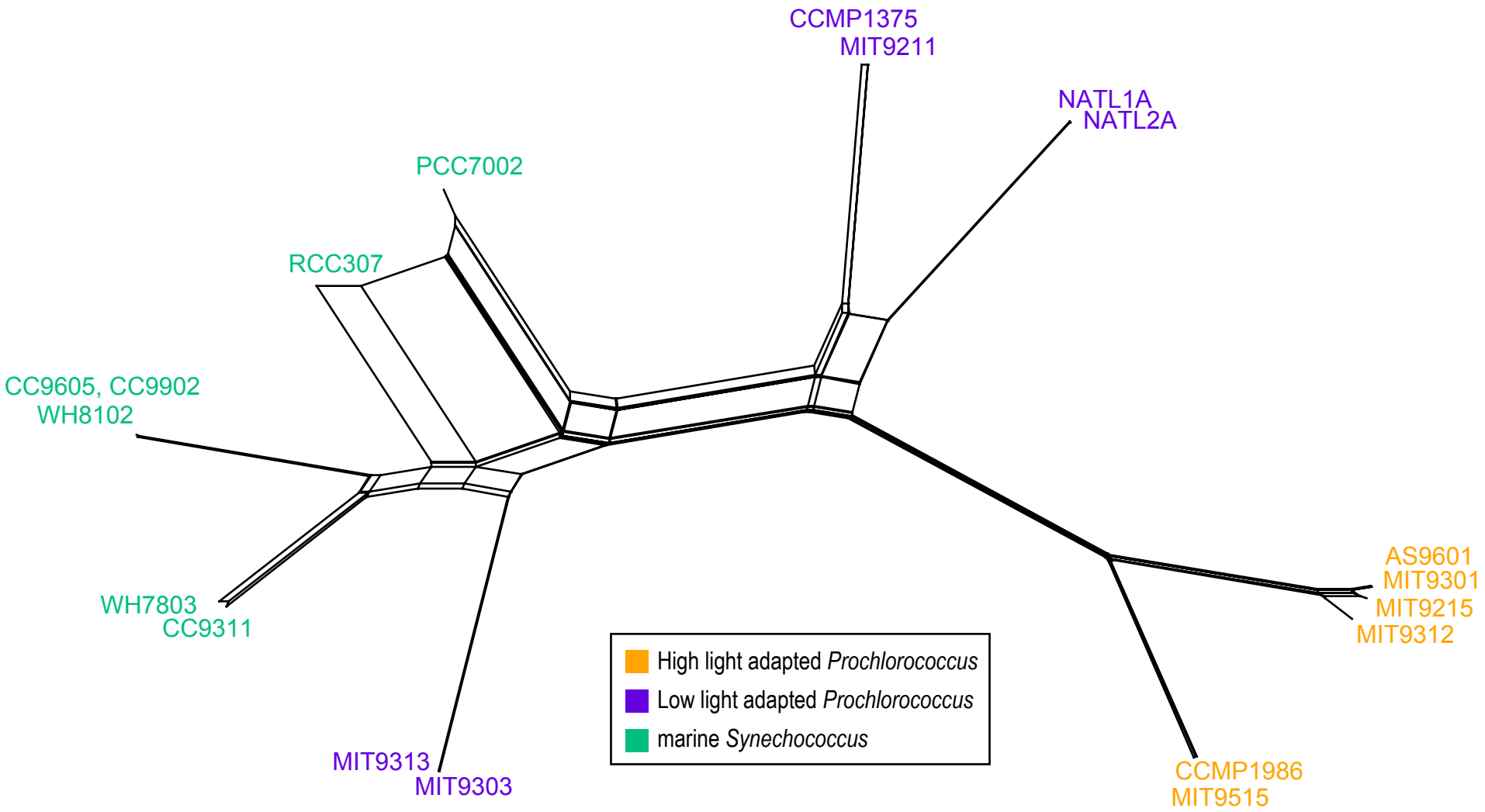


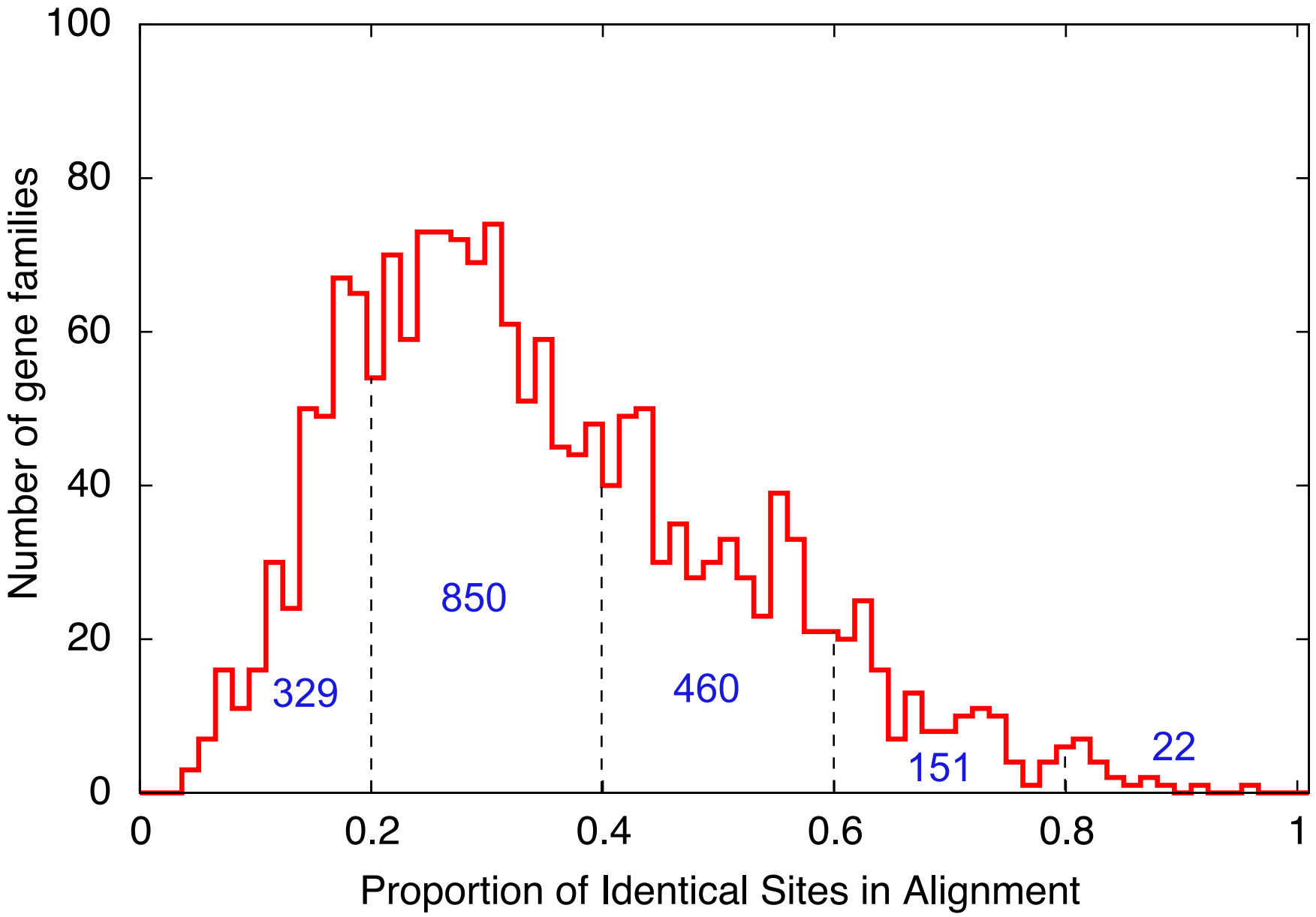
Supplementary Figure 1. Division of gene families into three sets, depending on how many in-paralogs per family was present. Within the "No in-paralogs" set (yellow) genes were subdivided into "core genes" (red) and "core genes with an ortholog in the outgroup genome" (orange) subsets. In "At most 8 in-paralogs" set (green) "core genes" subset (light green) was identified. "More than 8 in-paralogs" set (blue) was excluded from further analyses.



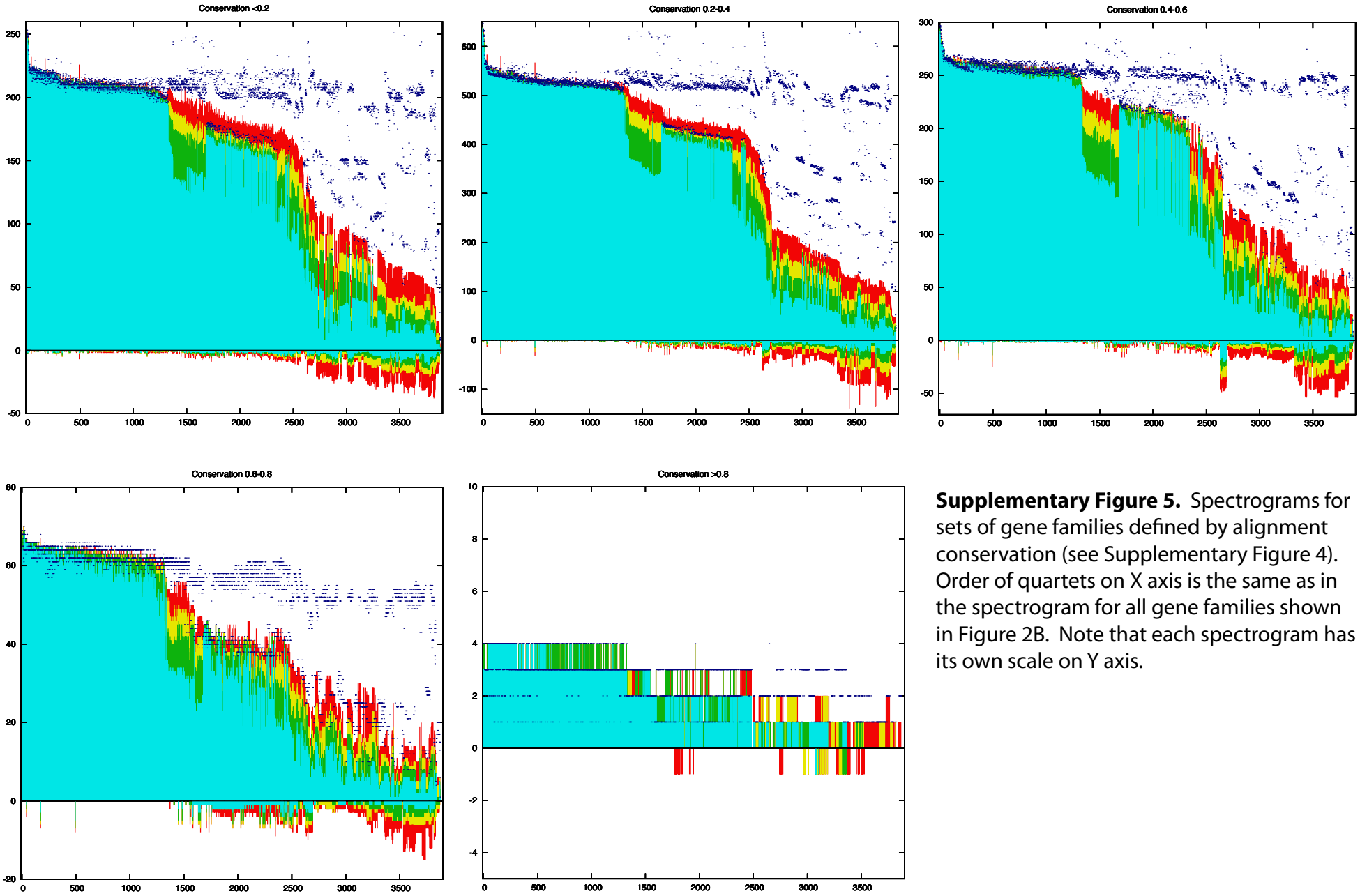
Supplementary Figure 2.
Distribution of gene families without in-paralogs across functional categories. The four super-categories are defined by COG database (see Materials and Methods). Notably, genes of informational storage and processing are represented in equal proportions in genes in conflict with plurality as compared to all 1812 gene families, which contradicts complexity hypothesis (Jain et al. 1999). Metabolic genes appear to be overrepresented in the gene family pool which conflicts with plurality.



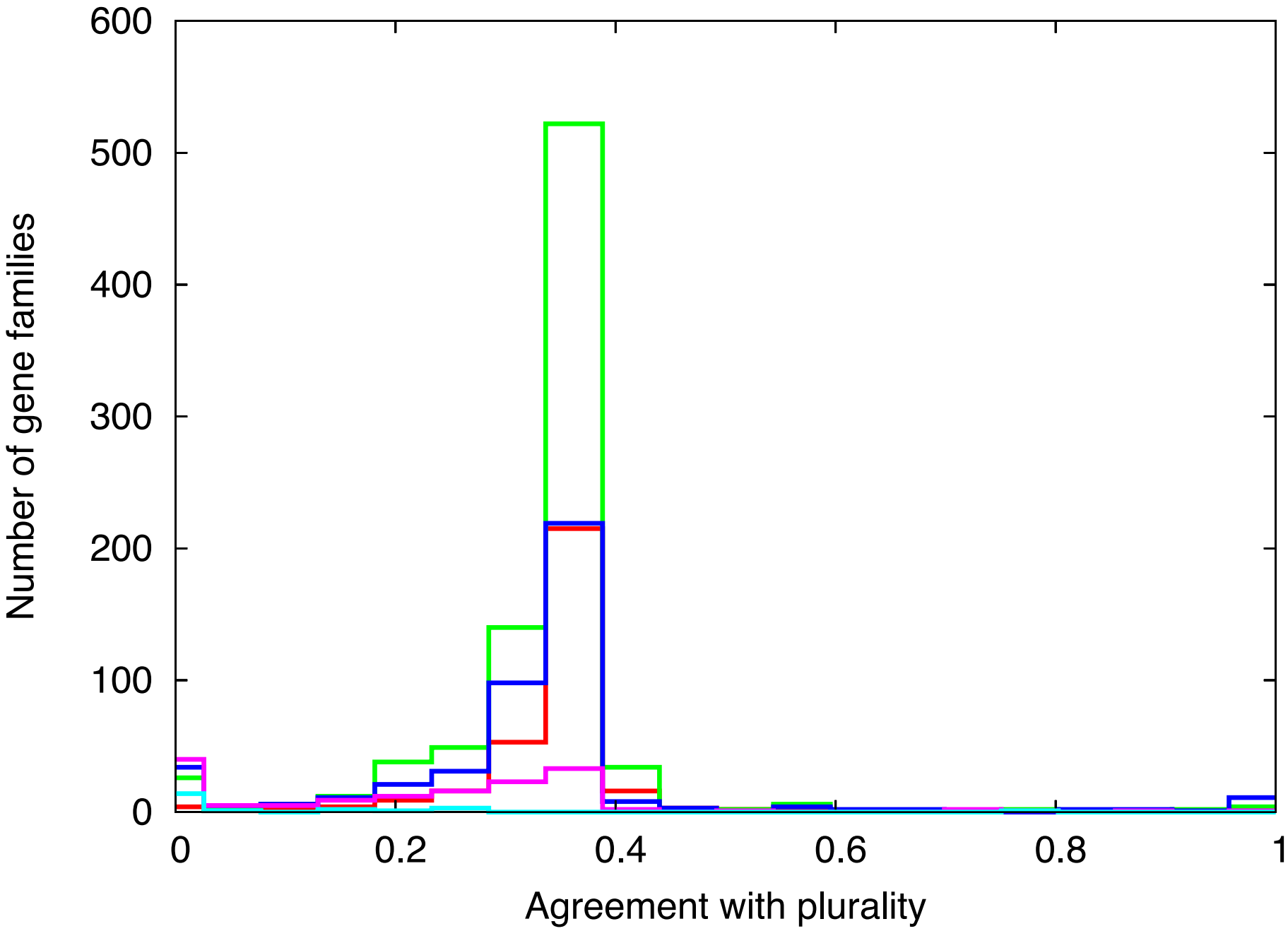
Supplementary Figure 3. NeighborNet reconstruction of MRP matrix constructed from quartets supported by the plurality of gene families without paralogs (compare to Figure 1A and 1B). 79 quartets that conflict strictly bifurcating tree shown in Figure 1A account for alternative branchings shown on this NeighborNet diagram.



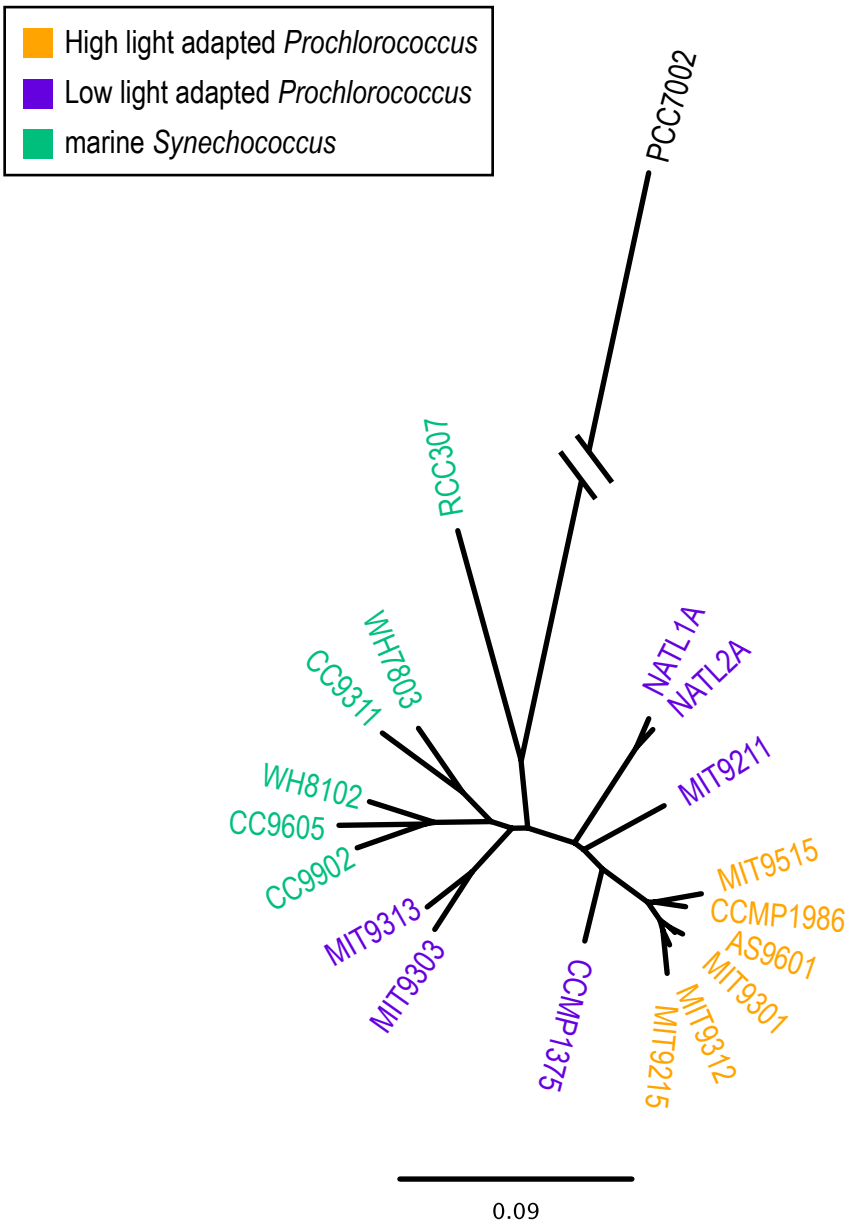
Supplementary Figure 4. Division of gene families by alignment conservation. Numbers in blue indicate the number of gene families in each area of the histogram.



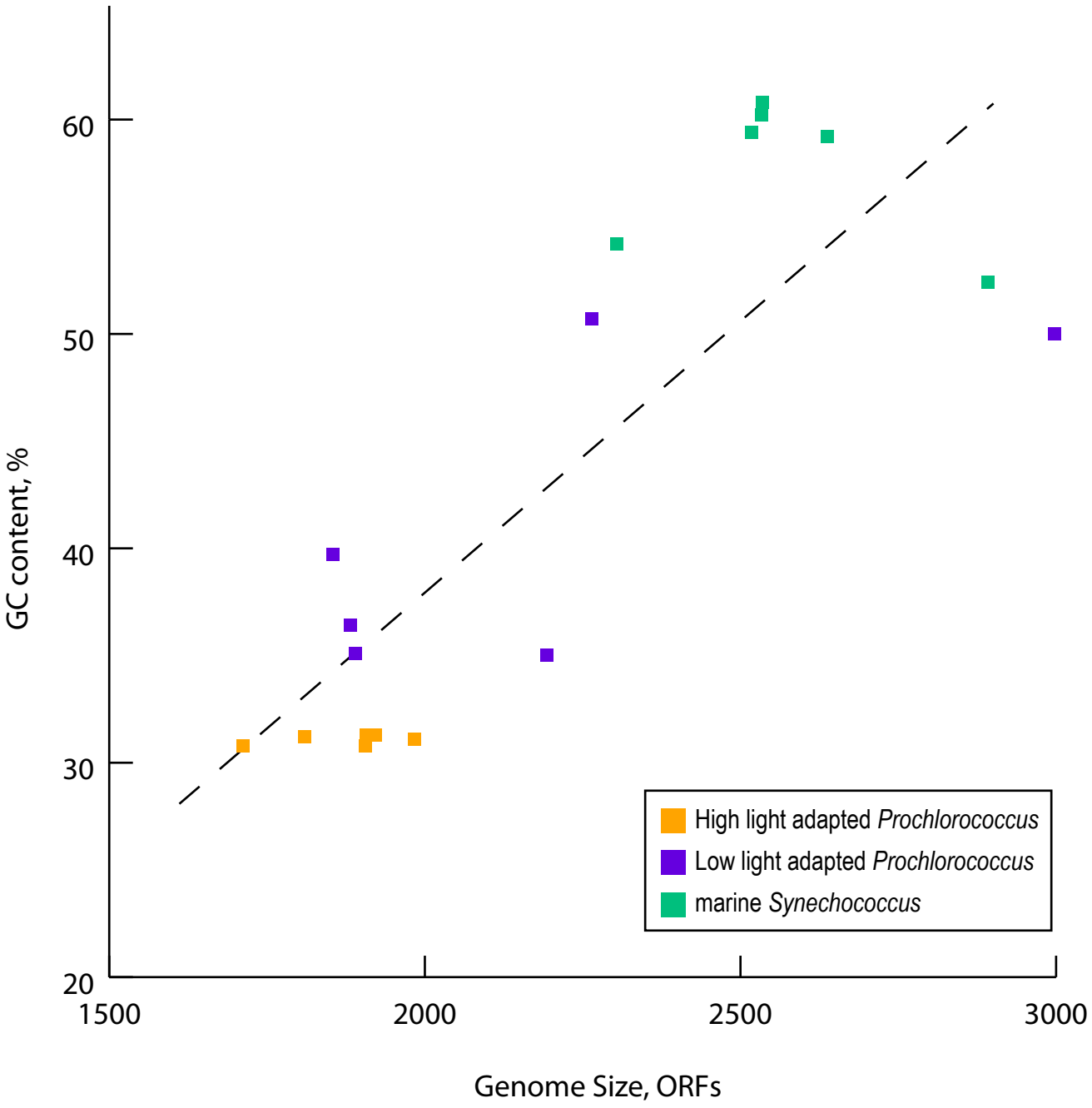
Supplementary Figure 5. Spectrograms for sets of gene families defined by alignment conservation (see Supplementary Figure 4). Order of quartets on X axis is the same as in the spectrogram for all gene families shown in Figure 2B. Note that each spectrogram has its own scale on Y axis.



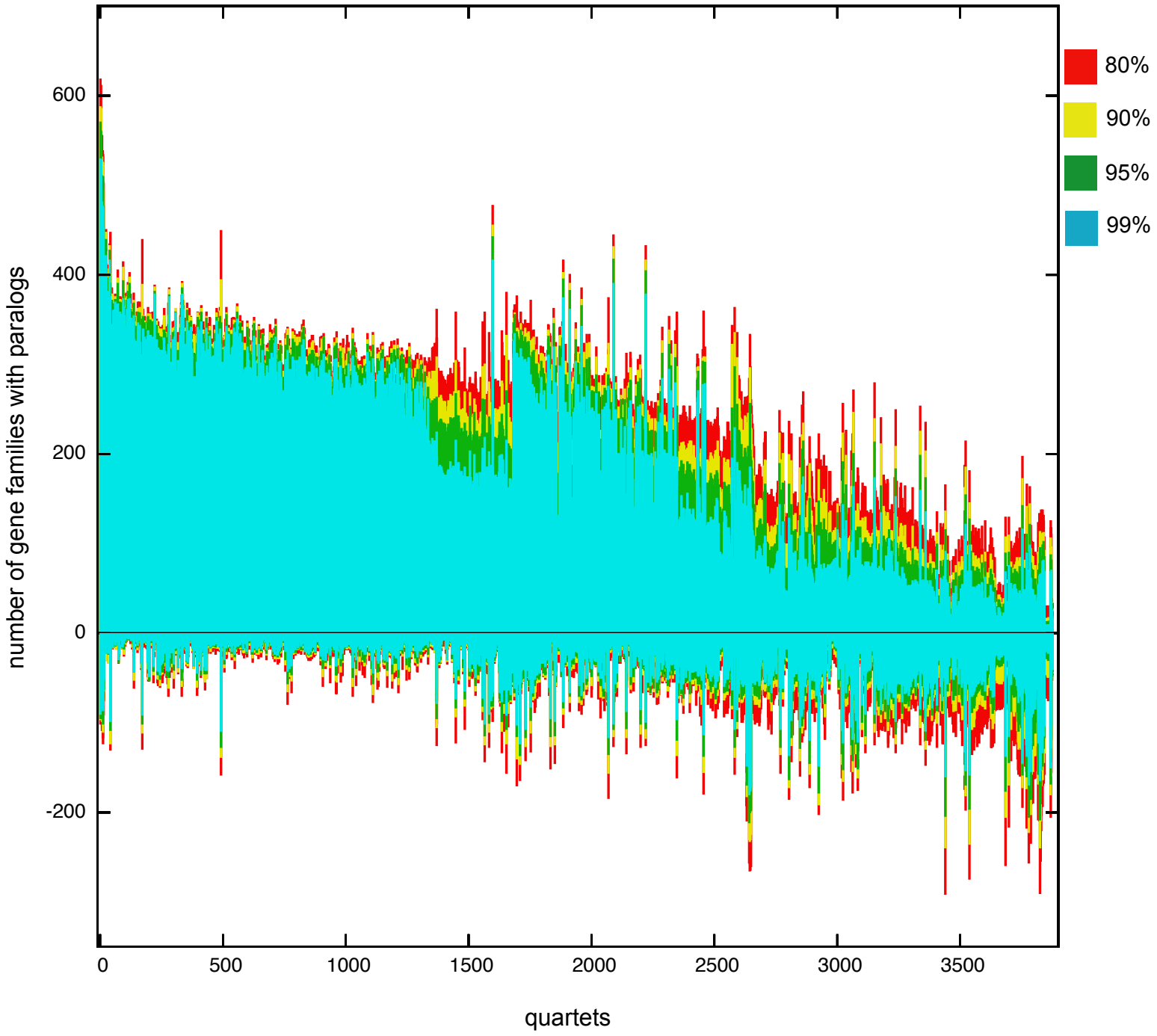
Supplementary Figure 6. Agreement of gene families grouped by alignment conservation with plurality tree. Each histogram corresponds to family set defined in Supplementary Figure 4. For details on agreement score calculation see Materials and Methods. Compare this figure with Figure 4.



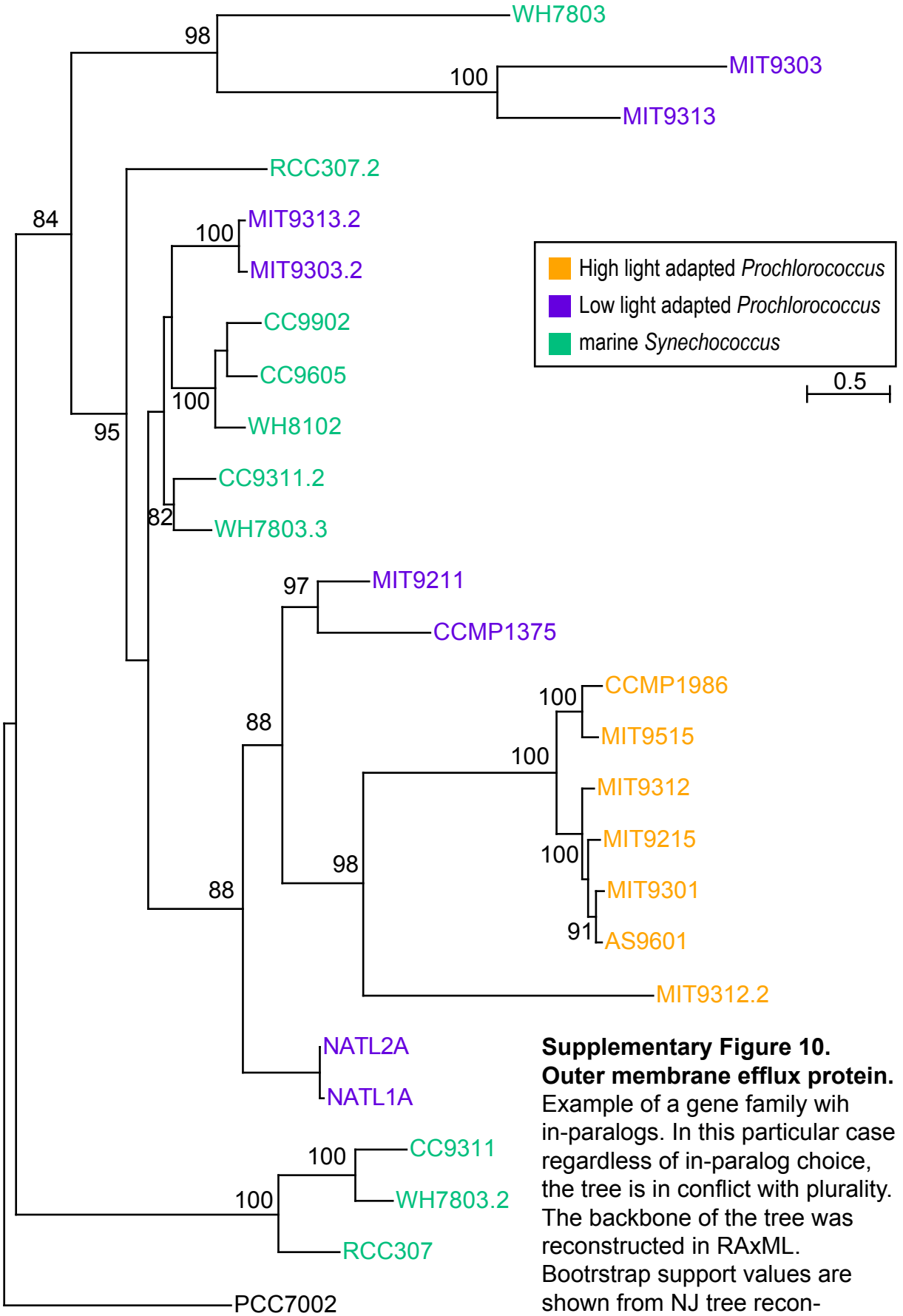
Supplementary Figure 7. Phylogenetic tree reconstructed from re-arrangement distances.



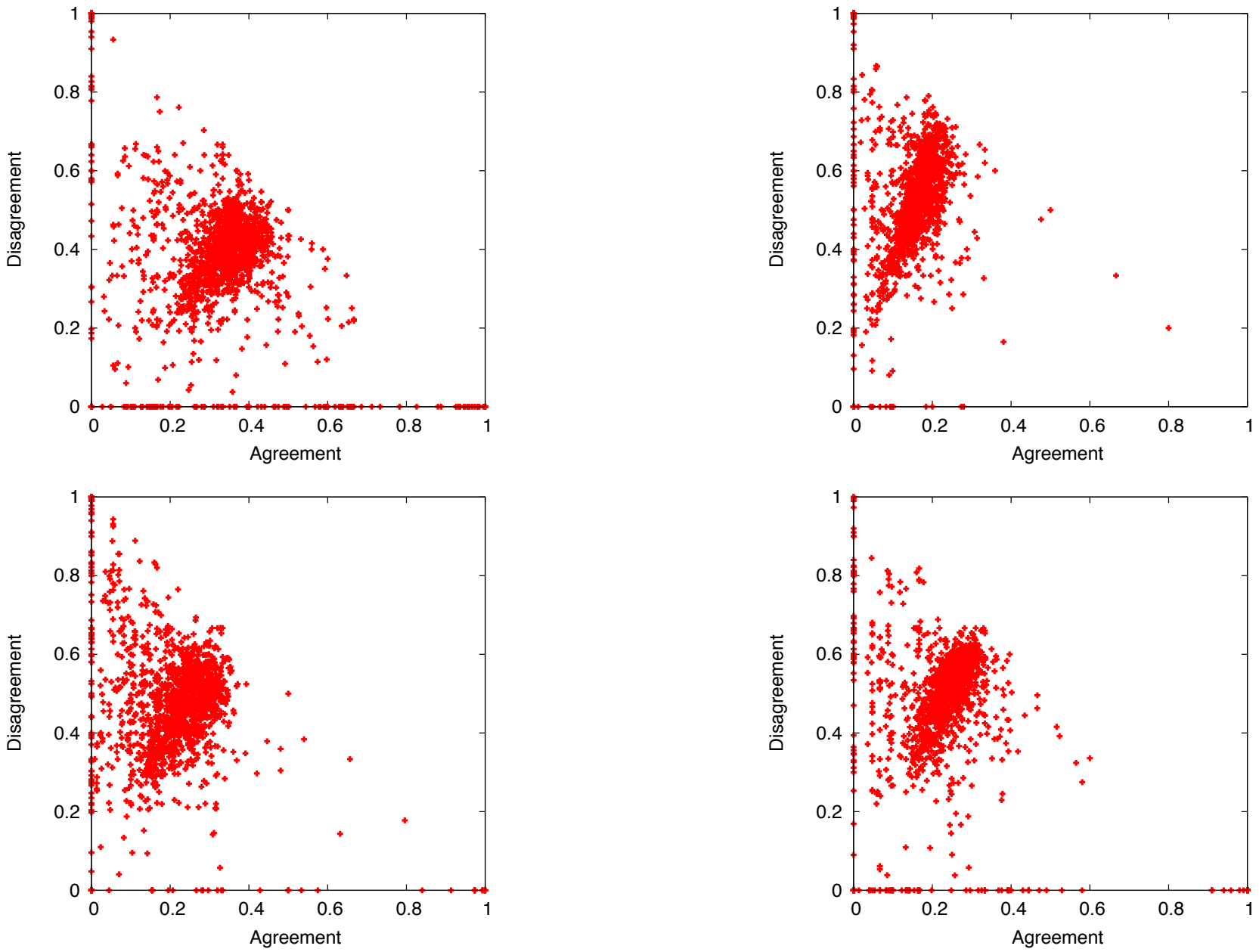
Supplementary Figure 8. Correlation between genome size and GC content among 19 analyzed genomes. Trendline y=0.0252x-12.433, r^2=0.68 is shown as a dashed line.



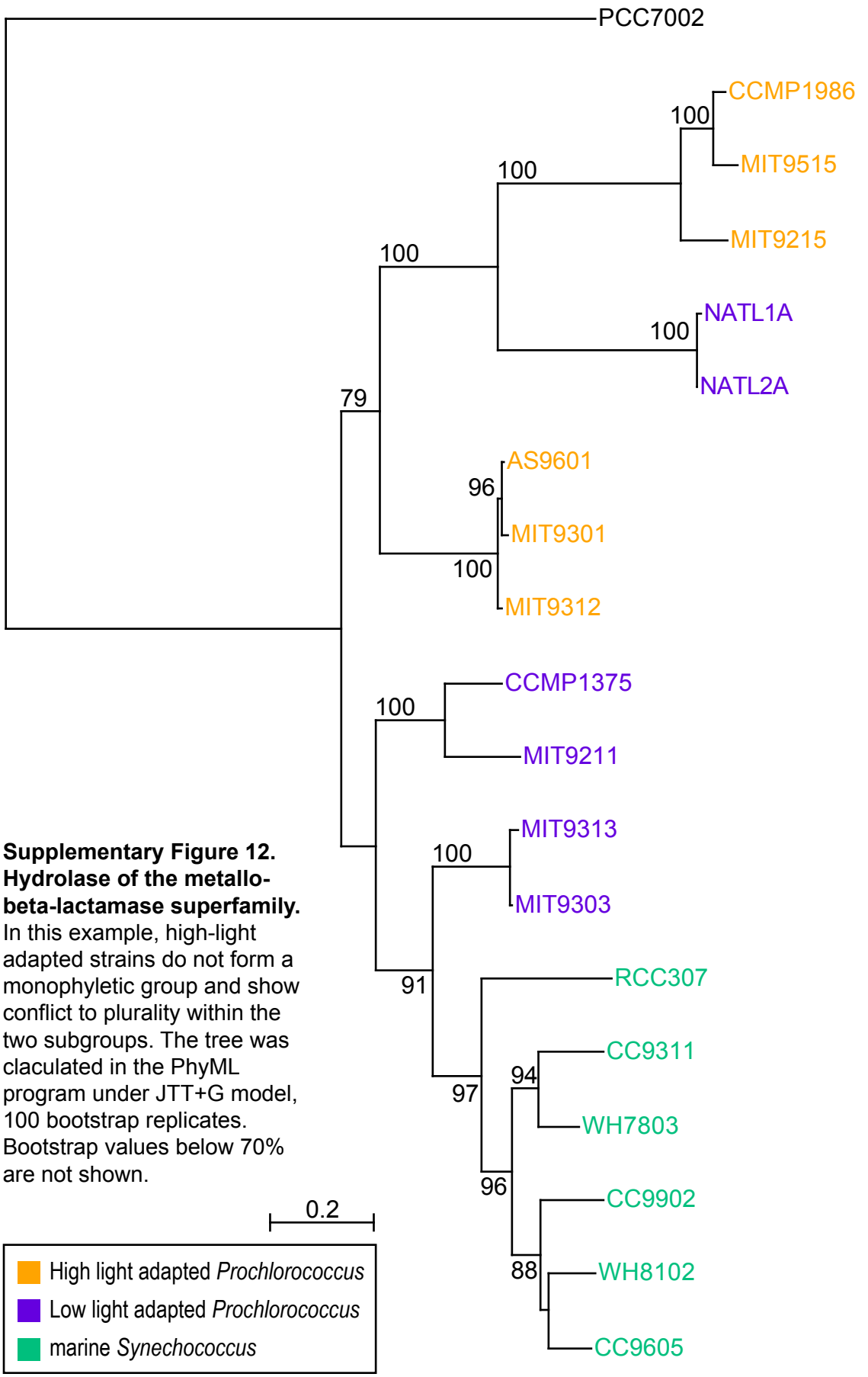
Supplementary Figure 9. Quartet decomposition analysis of 482 gene families with paralogs. Columns are in the order of their appearnce in Figure 2B. The embedded quartets were evaluated to their agreement with quartet's plurality topology (Figure 1A). Each quartet corresponding to in-paralogs was evaluated independently. Note that number of gene families on Y axis is amplified by presence of in-paralogs, as the embedded quartets containing them are counted independently within each gene family. For figure notations see legend to Figure 2B.



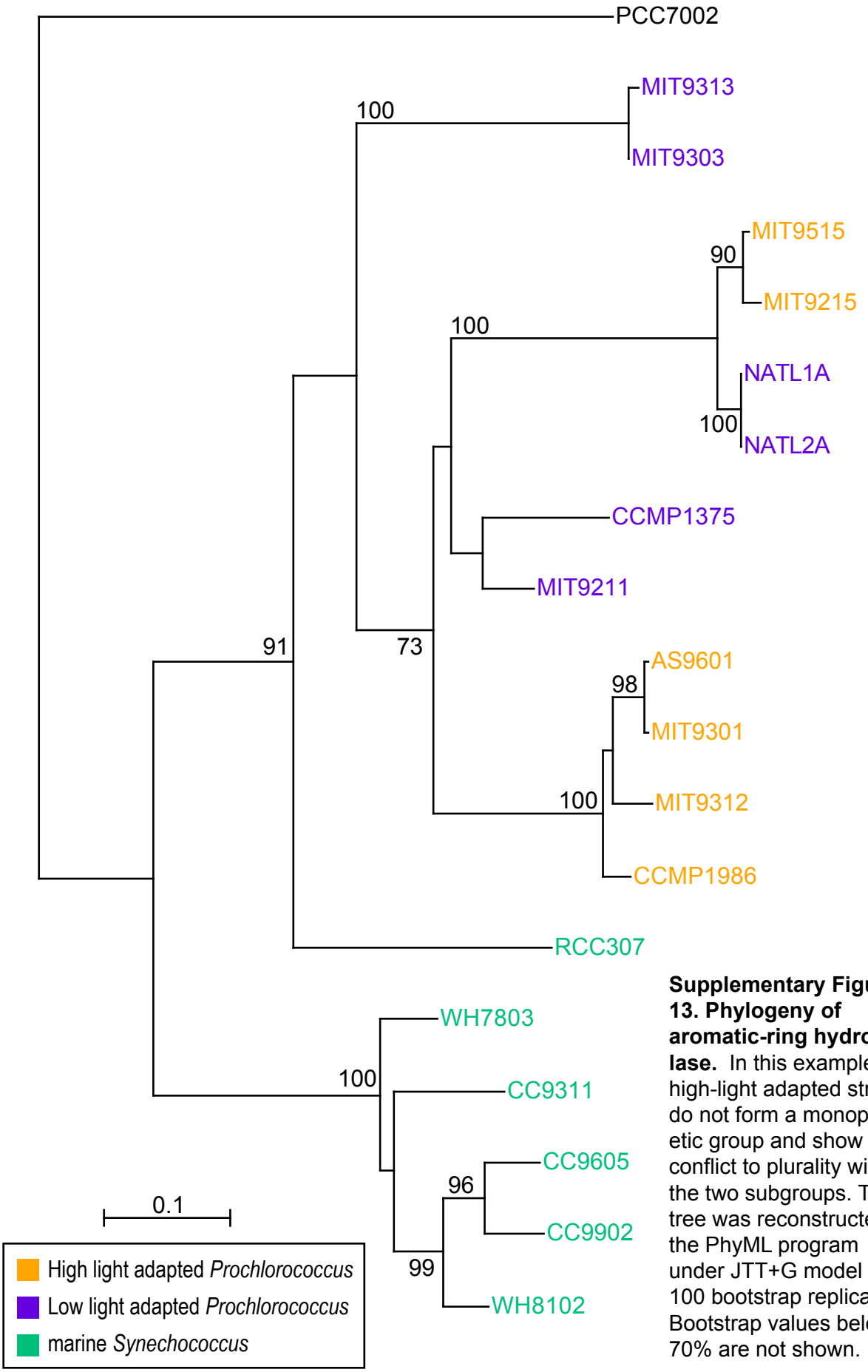
Supplementary Figure 10. Outer membrane efflux protein. Example of a gene family wih in-paralogs. In this particular case regardless of in-paralog choice, the tree is in conflict with plurality. The backbone of the tree was reconstructed in RAxML. Bootrstrap support values are shown from NJ tree reconstructed from the TREE-PUZZLE distances.



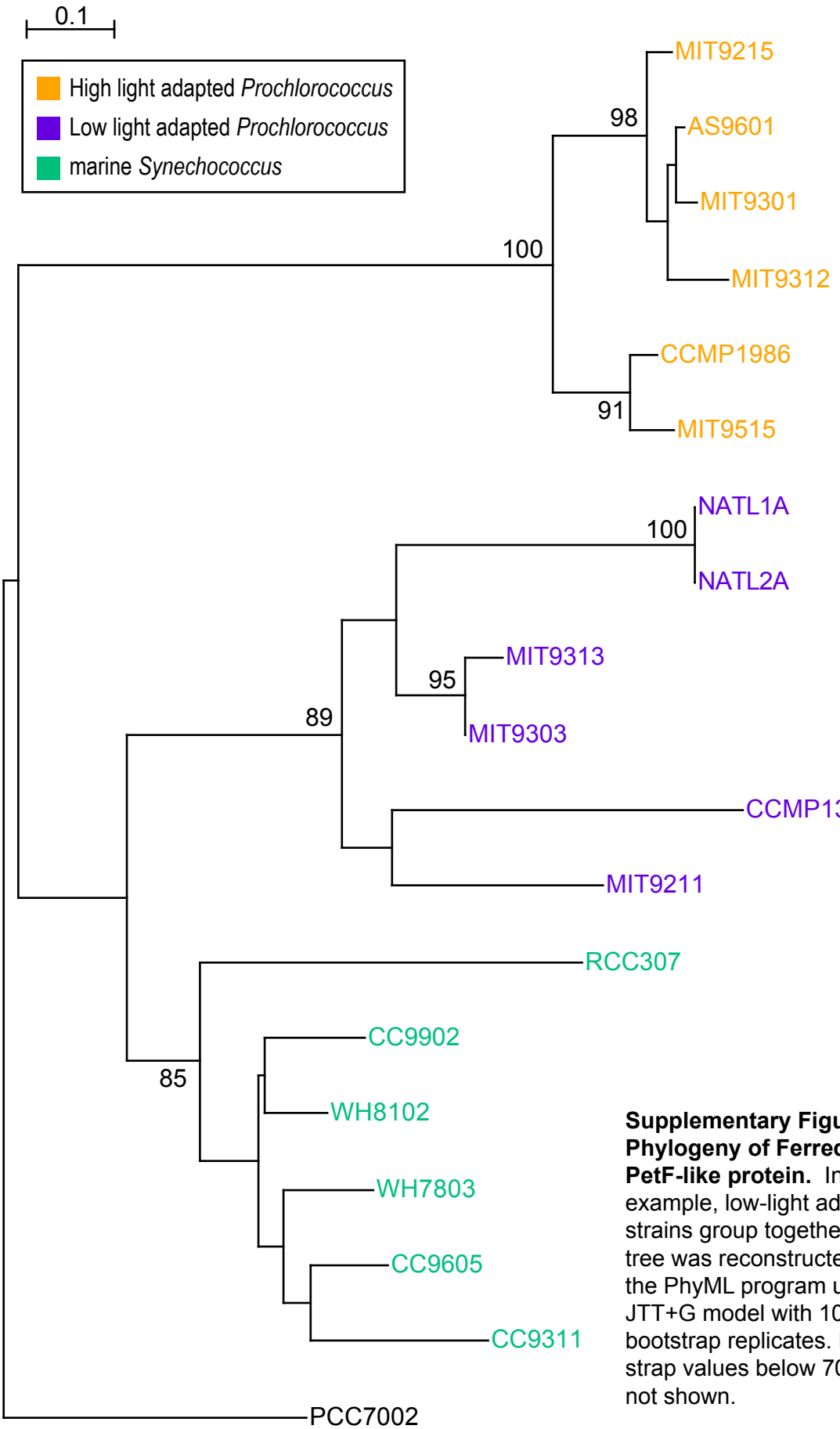
Supplementary Figure 11. Scatterplots for four independently generated random partitions of data.



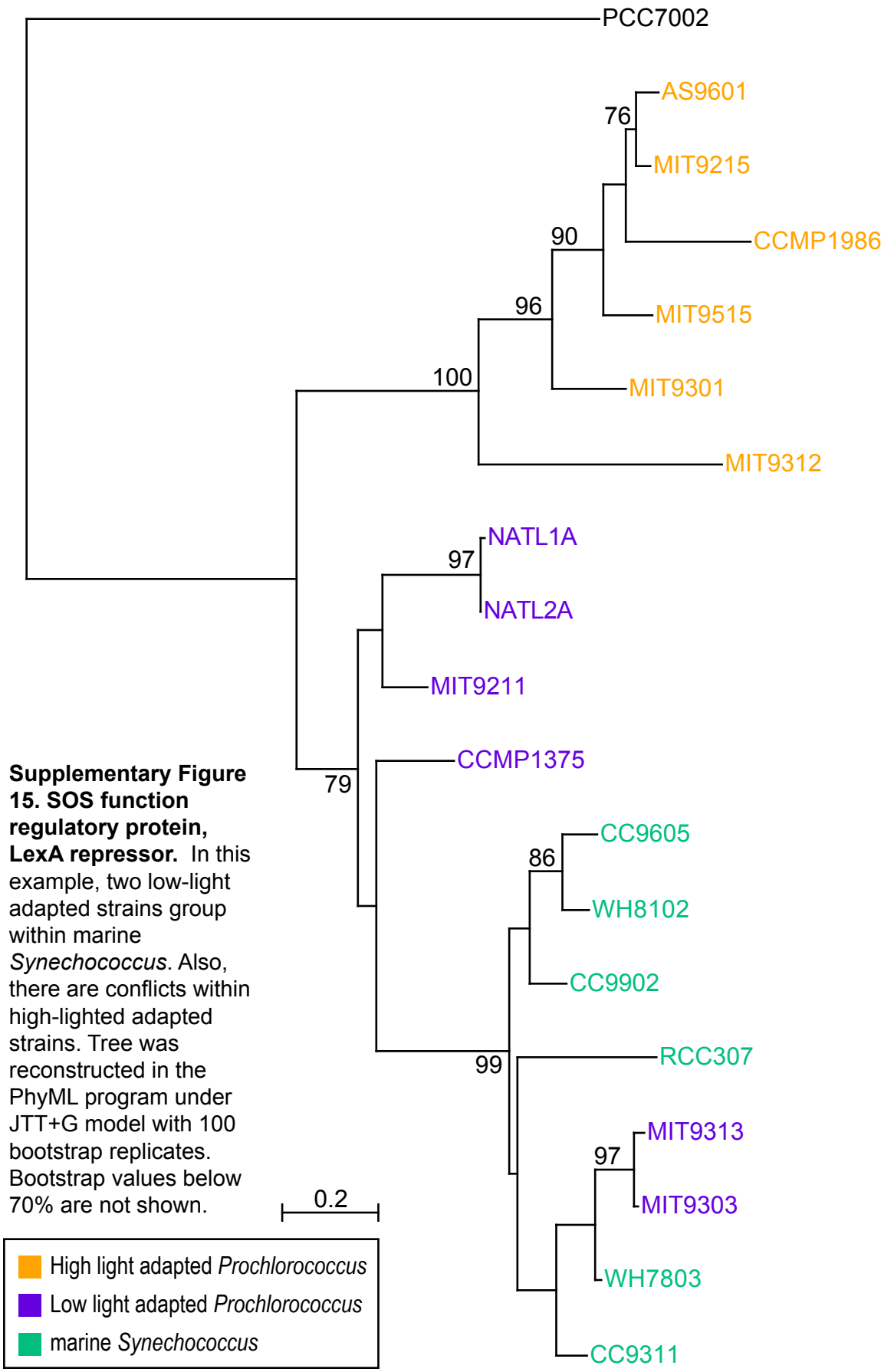
Supplementary Figure 12. Hydrolase of the metallo-beta-lactamase superfamily. In this example, high-light adapted strains do not form a monophyletic group and show conflict to plurality within the two subgroups. The tree was claculated in the PhyML program under JTT+G model, 100 bootstrap replicates. Bootstrap values below 70% are not shown.



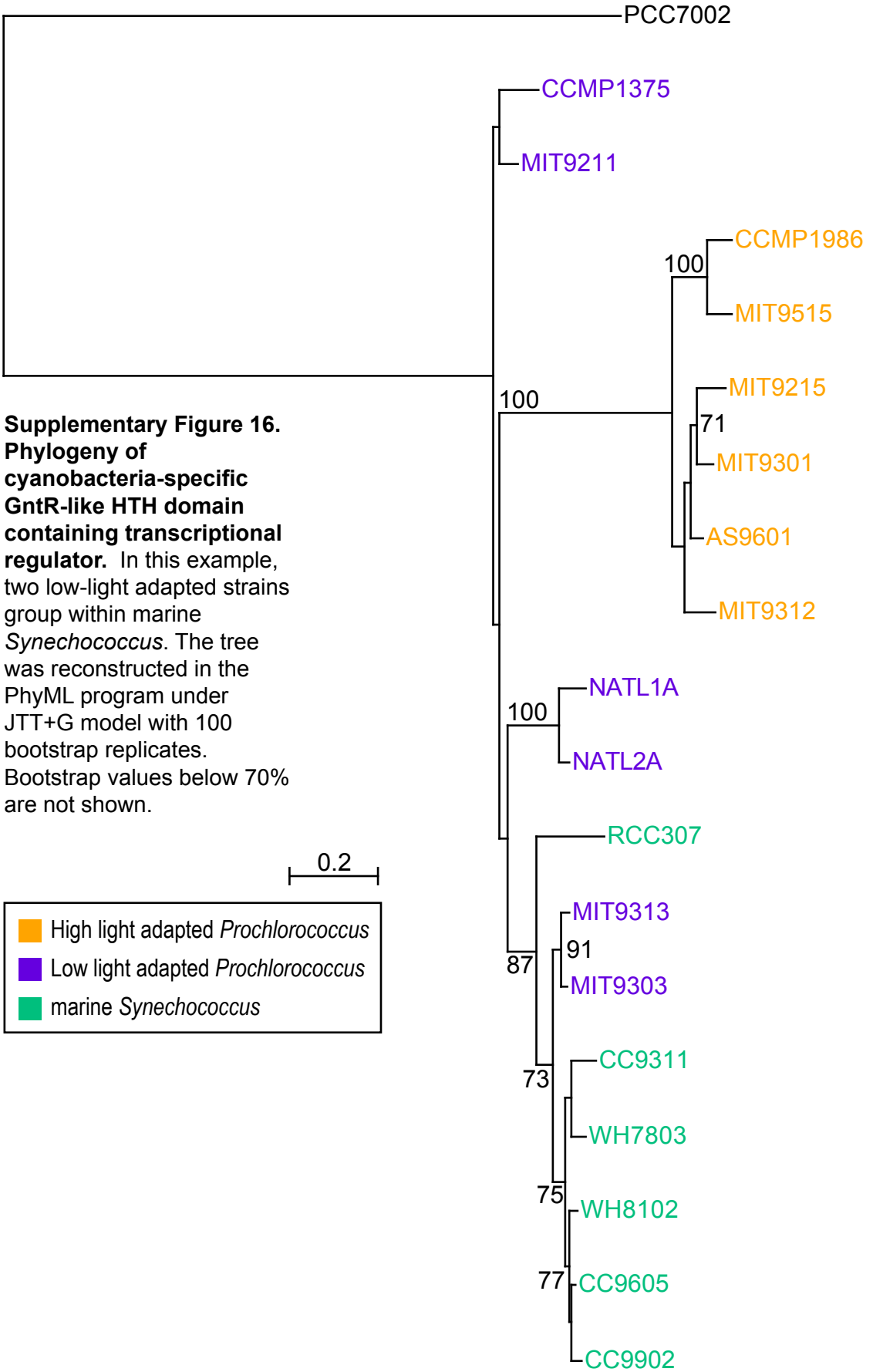
Supplementary Figure 13. Phylogeny of aromatic-ring hydroxylase. In this example, high-light adapted strains do not form a monophyletic group and show conflict to plurality within the two subgroups. The tree was reconstructed in the PhyML program under JTT+G model with 100 bootstrap replicates. Bootstrap values below 70% are not shown.



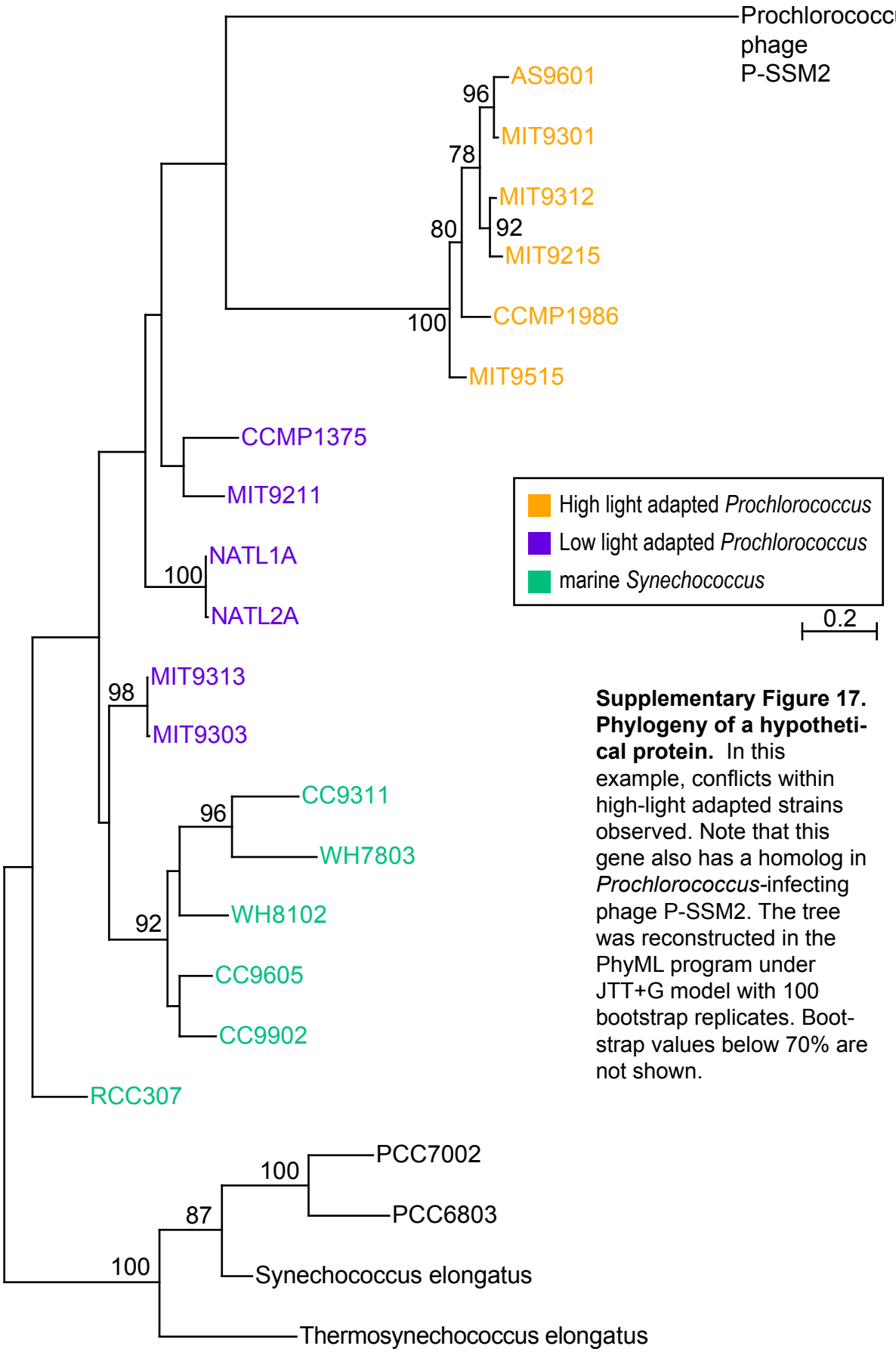
Supplementary Figure 14. Phylogeny of Ferredoxin, PetF-like protein. In this example, low-light adapted strains group together. The tree was reconstructed in the PhyML program under JTT+G model with 100 bootstrap replicates. Bootstrap values below 70% are not shown.



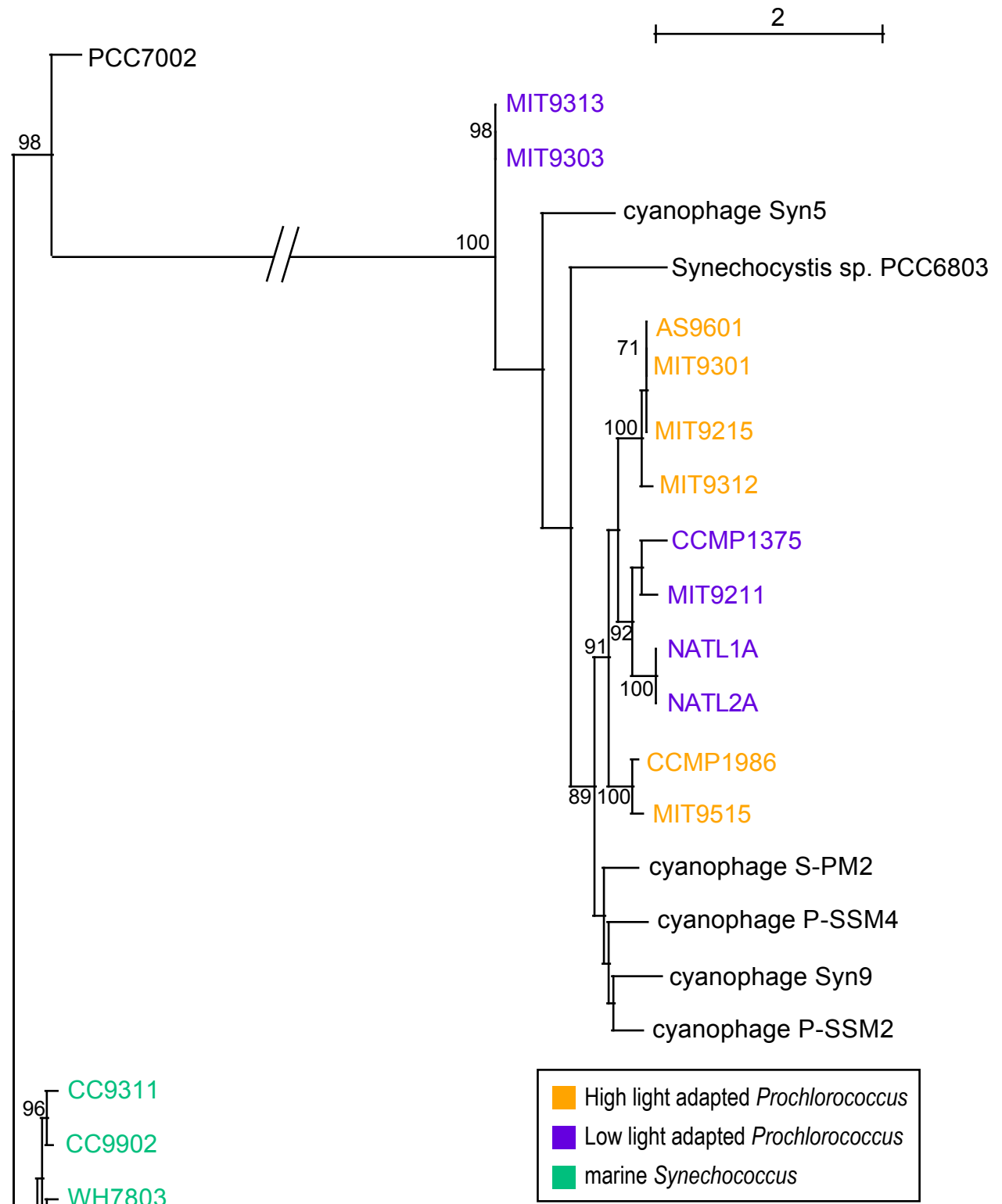
Supplementary Figure 15. SOS function regulatory protein, LexA repressor. In this example, two low-light adapted strains group within marine *Synechococcus*. Also, there are conflicts within high-lighted adapted strains. Tree was reconstructed in the PhyML program under JTT+G model with 100 bootstrap replicates. Bootstrap values below 70% are not shown.



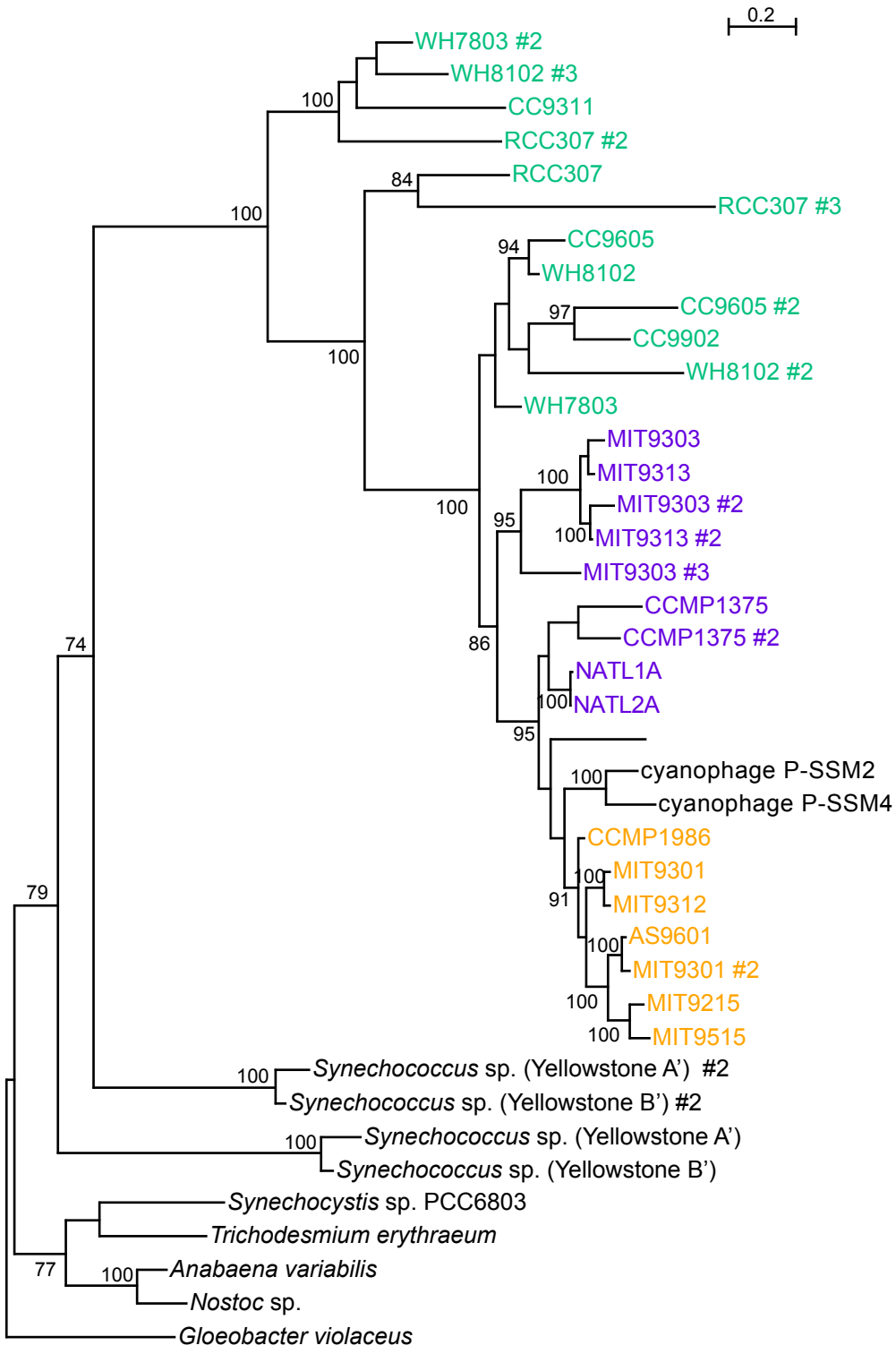
Supplementary Figure 16. Phylogeny of cyanobacteria-specific GntR-like HTH domain containing transcriptional regulator. In this example, two low-light adapted strains group within marine *Synechococcus*. The tree was reconstructed in the PhyML program under JTT+G model with 100 bootstrap replicates. Bootstrap values below 70% are not shown.



Supplementary Figure 17. Phylogeny of a hypothetical protein. In this example, conflicts within high-light adapted strains observed. Note that this gene also has a homolog in *Prochlorococcus*-infecting phage P-SSM2. The tree was reconstructed in the PhyML program under JTT+G model with 100 bootstrap replicates. Bootstrap values below 70% are not shown.



Supplementary Figure 18. Phylogeny of thymidilate synthase. In this example of a fast evolving gene family, cyanophage homologs group with high-light and low-light adapted strains, while marine *Synechococcus* homologs are quite distantly related to them. Note that high-light adated strains do not form a monophyletic group. The tree was reconstructed in the PhyML program under JTT+G model with 100 bootstrap replicates. Bootstrap values below 70% are not shown.



High light adapted *Prochlorococcus*

Low light adapted *Prochlorococcus*

marine *Synechococcus*

Supplementary Figure 19. Phylogenetic tree of *pstS* gene. Multiple homologs from the same genome are numbered sequentially. Cyanobacteria outside of *Prochlorococcus*/marine *Synechococcus* group are used as an outgroup. Numerous relationships within each subgroup are in conflict with plurality signal and suggest that this gene is frequently exchanged, possibly mediated by phages.