

Supplementary Figure Legends for Rocap et al. 2003

Supp. Figure 1. Circular representation of the *Prochlorococcus* genomes. **a**, MED4. **b**, MIT 9313. For both genomes outermost circles (1 and 2) are predicted protein coding regions on the plus and minus strands, respectively. Color coding is as in Supplementary Figure 2. The next two circles show genes not present in the other *Prochlorococcus* genome on the plus (circle 3) and minus (circle 4) strands. Circles 5 and 6 show genes on the plus and minus strands, respectively that contain transmembrane domains. Circle 7 is % G+C content (deviation from average). Innermost circle (8) represents the GC skew curve.

Supp. Figure 2. Functional categorization of predicted open reading frames in the *Prochlorococcus* genomes, following the classification scheme used by CyanoBase. **a**, MED4, entire genome. **b**, MIT 9313, entire genome. **C**, Genes present in both MED4 and MIT 9313. **d**, Genes in MED4 not present in MIT 9313. **e**, Genes in MIT 9313 not present in MED4.

Supp. Figure 3. Comparison of *Prochlorococcus* MED4 and MIT 9313 open reading frames with those of other complete prokaryotic genomes. The predicted coding sequences of each gene in both genomes were aligned with the coding sequences of 90 bacterial genomes using BLASTP. Significant alignments were defined as having an e-value less than 10^{-6} . The bacterial genomes comprised the 89 completed bacterial genomes available from <ftp.ncbi.nih.gov/genbank/genomes/Bacteria> on 30 October 2002 and *Synechococcus* WH 8102⁸. **a**, MED4, entire genome. **B**, MIT 9313, entire genome. **c**, MED4 genes present in MIT 9313 **c**, MIT 9313 genes present in MED4 **e**, Genes in MED4 not present in MIT 9313 **f**, Genes in MIT 9313 not present in MED4.

Supp. Figure 4 Alignment of the putative nitrite transporter in *Prochlorococcus* MIT9313 (PMT2240) with its most significant matches in the NR database (all proteobacteria) and with cyanobacterial nitrate/nitrite transporters. The MIT 9313 gene has a formate/nitrite transporter domain (Pfam PF01226) in contrast to the cyanobacterial nitrate transporters which are permeases of the major facilitator superfamily (Pfam PF00083). Furthermore, the MIT 9313 gene has no significant matches (BLASTP e -value $< e^{-2}$) in the genomes of *Prochlorococcus* MED4, *Synechococcus* WH8102, *Synechocystis* sp. PCC 6803, *Thermosynechococcus elongatus* BP-1, or *Anabaena* sp. PCC 7120 suggesting it may have been acquired via lateral gene transfer. Alignment generated using ClustalW. Shaded residues indicate $>50\%$ similarity. Abbreviations and accession numbers as follows: Rhodopseud., *Rhodopseudomonas palustris* (ZP_00012718.1); Bradyrhiz., *Bradyrhizobium japonicum* (NP_769441); Vibrio, *Vibrio vulnificus* (NP_762336.1); Nitros., *Nitrosomonas europaea* (NP_840759); WH 7803, *Synechococcus* WH 7803 napA (AAG45172); PCC 7002, *Synechococcus* PCC 7002 nrtP (AAD45941); WH9601, *Trichodesmium* WH 9601 napA (AAF00917); PCC 73102, *Nostoc punctiforme* PCC 73102 (ZP_00107423).

Supp. Figure 5 Phylogenetic tree showing the relationship of a possible alkaline phosphatase like gene in *Prochlorococcus* MED4 (PMM0708) with the most significant matches in the NR database, which include several proteobacterial sequences, and with the atypical alkaline phosphatase of *Synechococcus* PCC 7942 and related cyanobacterial genes. Accession numbers as follows: *Brucella melitensis* (NP_541633.1), *Agrobacterium tumefaciens* str. C58 (NP_531956.1); *Sinorhizobium meliloti*, (NP_385365.1); *Vibrio vulnificus* (NP_762849.1), *Streptomyces coelicolor* A3(2) (NP_624650.1),

Shewanella oneidensis MR-1 (NP_717877.1) *Anabaena* PCC 7102 (NP_489331.1), *Synechocystis* sp. PCC 6803 (NP_440276); *Synechococcus* sp. PCC 7942 (A47026).

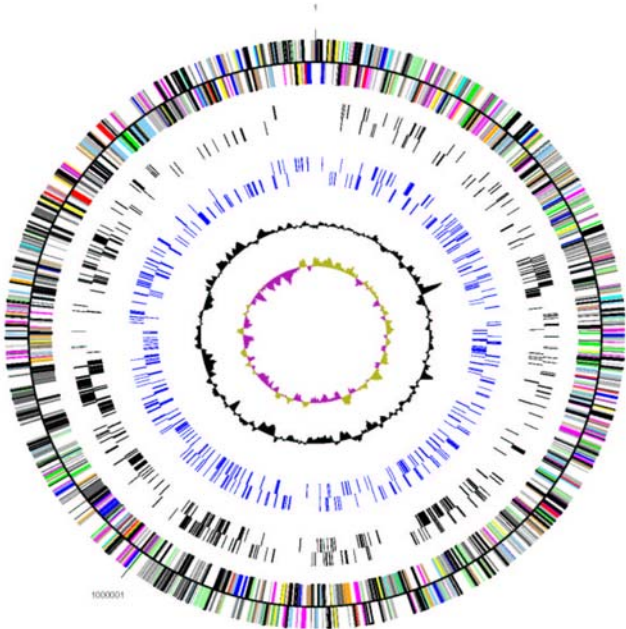
Supp. Figure 6 Insertions, deletions and rearrangements of genes involved in lipopolysaccharide biosynthesis (LPS clusters) in MED4. Color coding is as follows: blue, orthologous genes present in all three genomes; pink, genes hypothesized to be part of lateral transfer events, many have roles in LPS biosynthesis; red, tRNAs; green, orthologous genes present in two genomes, many have roles in LPS biosynthesis; white, other genes. Length in bp represents the size of the region shown for each genome. **a**, Insertion of a 74.5 kbp cluster of LPS genes in MED4, roughly between two tRNAs. The 67 potential surface polysaccharide genes in this cluster include sugar transferases, sugar epimerases, and modifying enzymes such as aminotransferases, methyltransferases, carbamoyltransferases, and acetyltransferases. In MIT 9313 and WH 8102 the genes that flank this insertion are rearranged to other parts of the genome. **b**, Deletion of LPS biosynthesis genes in MED4. LPS related genes present in MIT 9313 and WH 8102, several of which have homologs in the acquired genes shown in part a, have been deleted. In this region a selenophosphate synthase (*selD*) and a tRNA nucleotidyl-transferase in the center of the cluster have been retained suggesting that they are essential genes and separate deletion events have occurred on either side of them.

Supp. Table 1 Number of predicted signal transduction and transcription factors suggests reduced regulatory capacity in *Prochlorococcus*

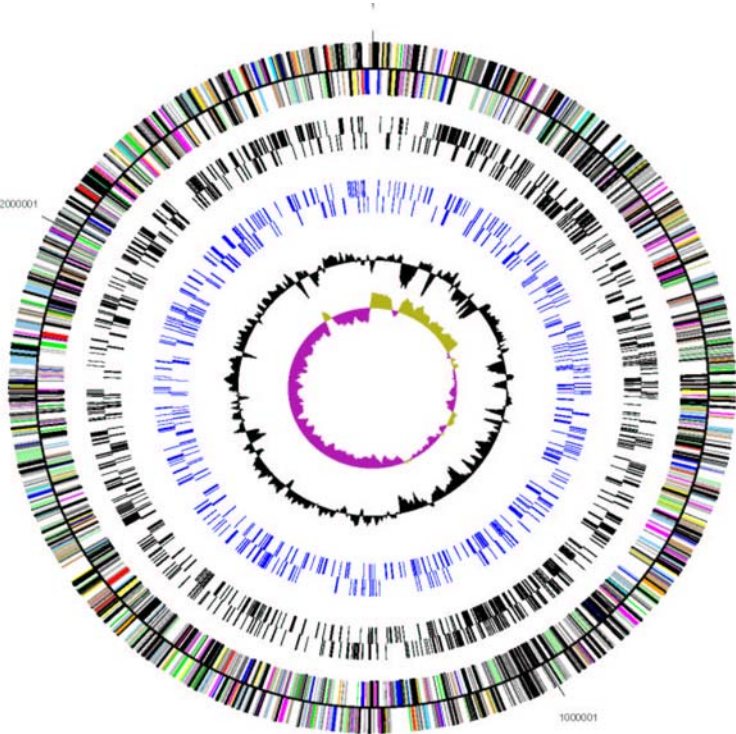
	MED4	MIT 9313	<i>T. elongatus</i>
Sigma Factors	5	8	8
Two Component systems			
Histidine Kinases	4	5	17
Response regulators	6	8	27
Ser/Thr protein Kinases	0	1	11
Transcription Factors			
LuxR family	2	5	4
LysR family	1	1	3
CRP family	3	4	3
ArsR family	1	2	2
FUR family	2	3	3
Other	2	3	3
Light sensors/transducers			
Cryptochrome	2	0	2
Bacteriophytochrome	0	0	5
Phototropin	0	0	1

Rocap et al. 2003 Supplementary Figure 1

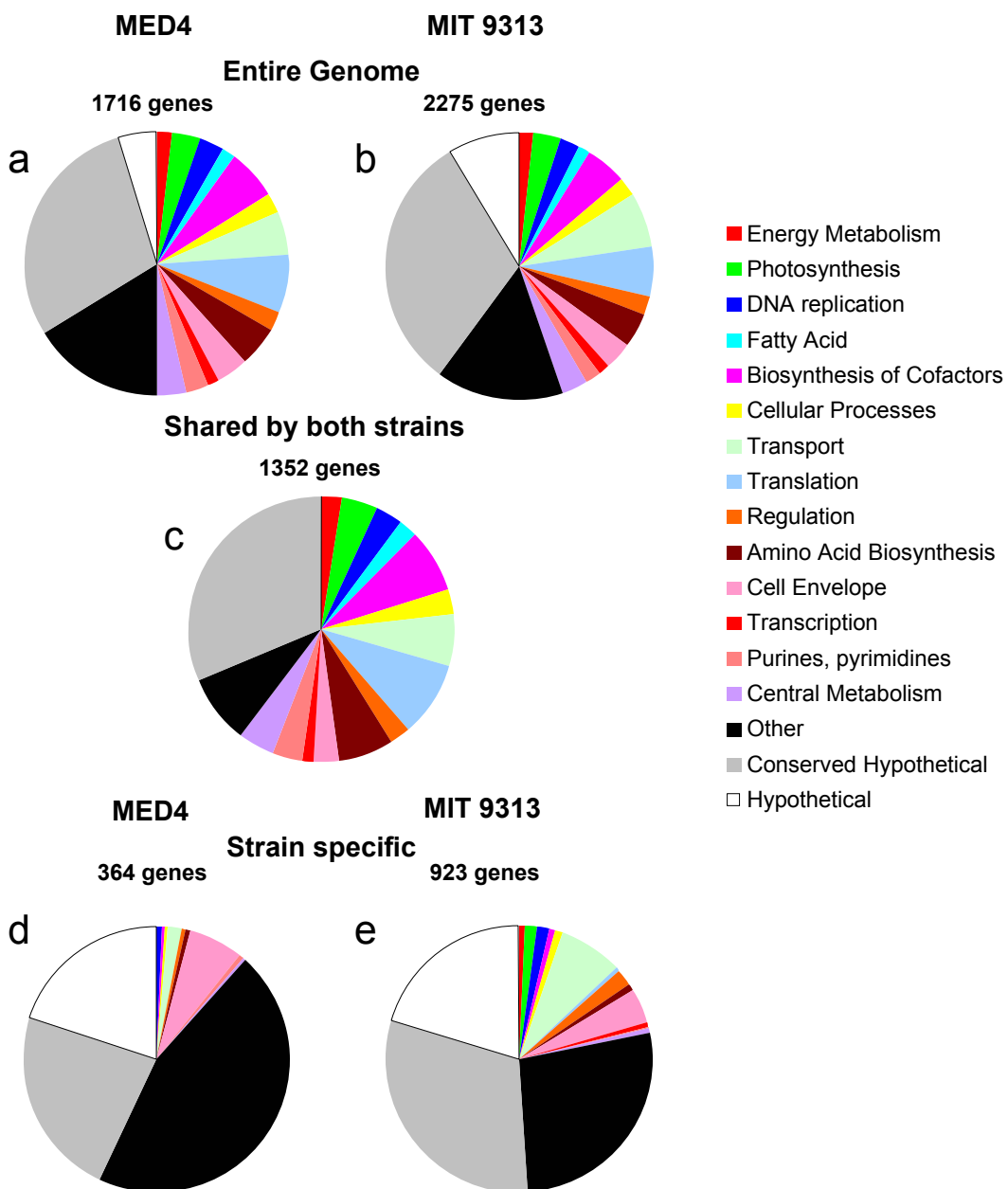
a



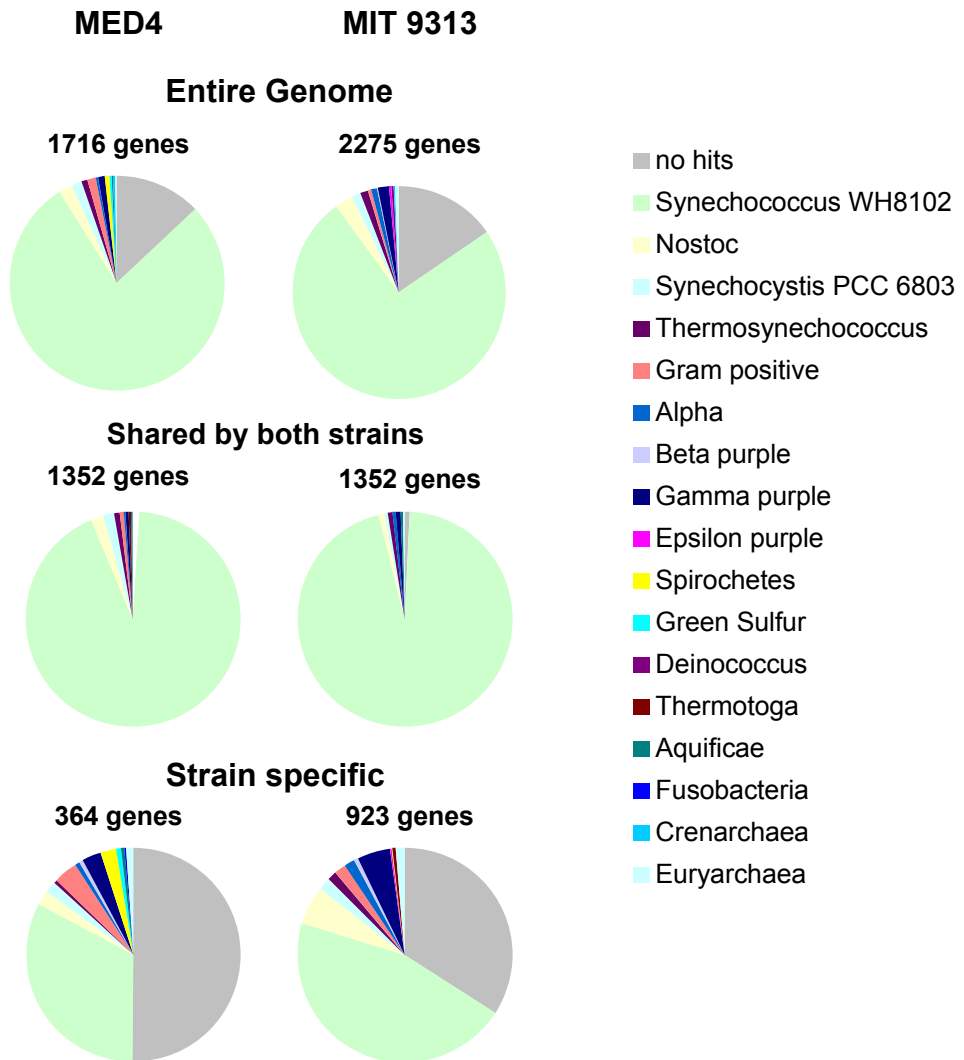
b



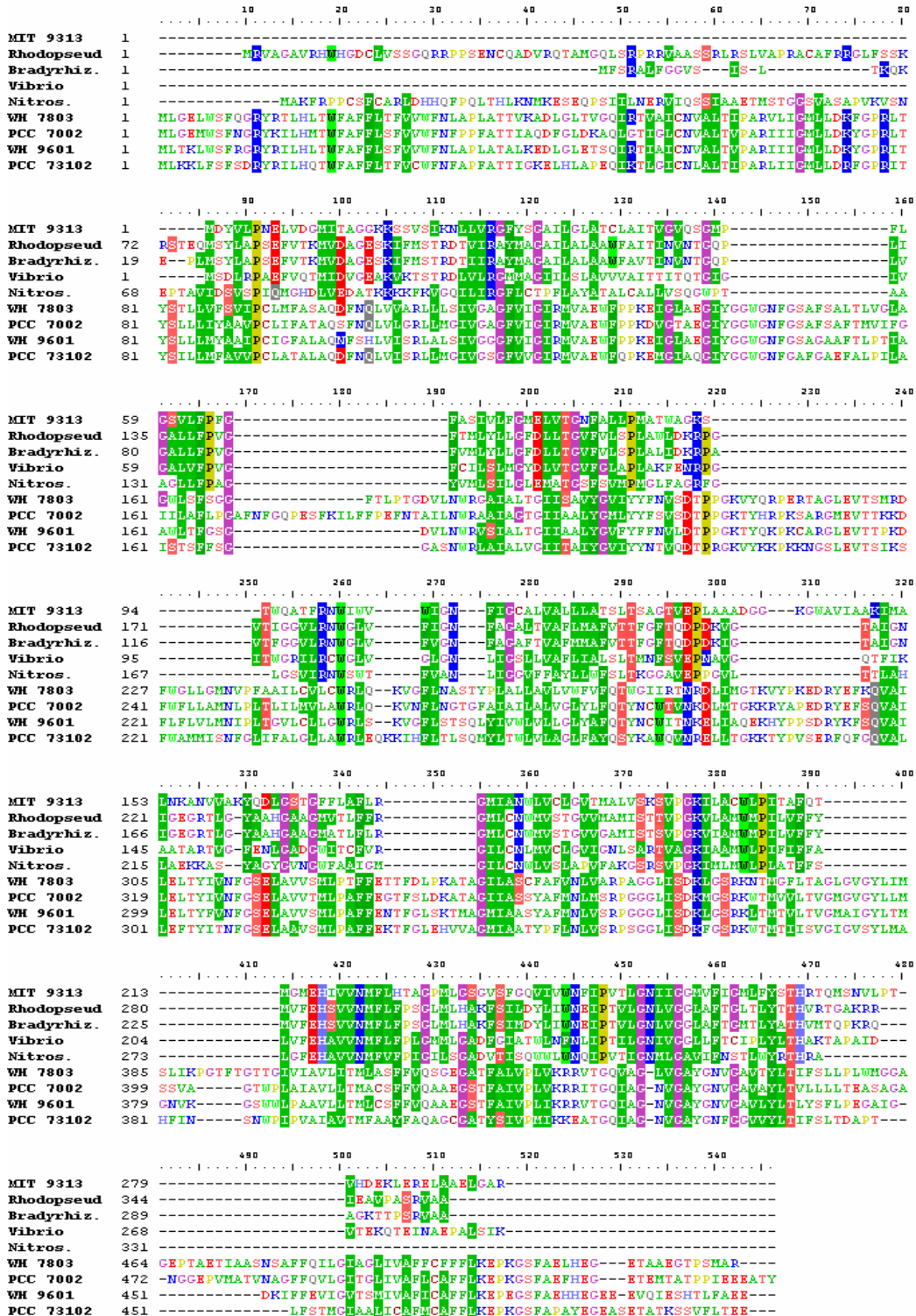
Rocap et al. 2003 Supplementary Figure 2



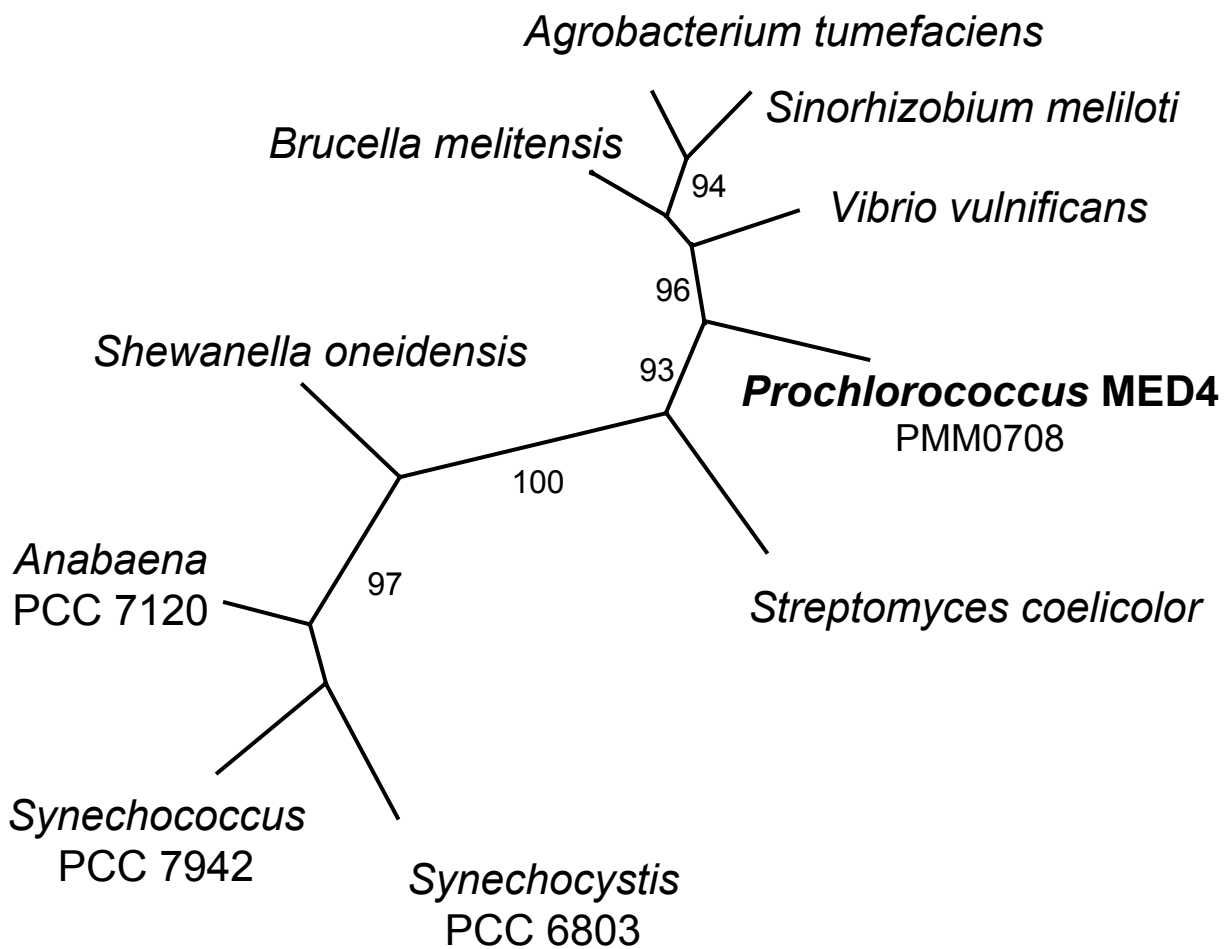
Rocap et al. 2003 Supplementary Figure 3



Rocap et al. 2003 Supplementary Figure 4



Rocap et al. 2003 Supplementary Figure 5



Rocap et al. 2003 Supplementary Figure 6

