



**Figure S2.** GOS scaffolds containing the *g20* portal vertex protein gene. Schematic of the scaffolds containing *g20*, with each gene/ORF (rectangles) colored to match its origin-type, compared to the cultured Cyano-T4 genomes in this region. Those *g20*s which do not represent the majority origin-type on the scaffolds or genomes are marked with an "X". Genes which have been duplicated and/or rearranged are indicated by their respective names within the colored squares. Due to space considerations, some of the genes/ORFs of known origin have been condensed (ex: "x3") and some ORFans have been omitted. The phylogenetic tree inset shows the relationships among the Cyano-T4 gp20 proteins, with T4 as the outgroup and the scale bar indicating 0.1 substitutions per site.