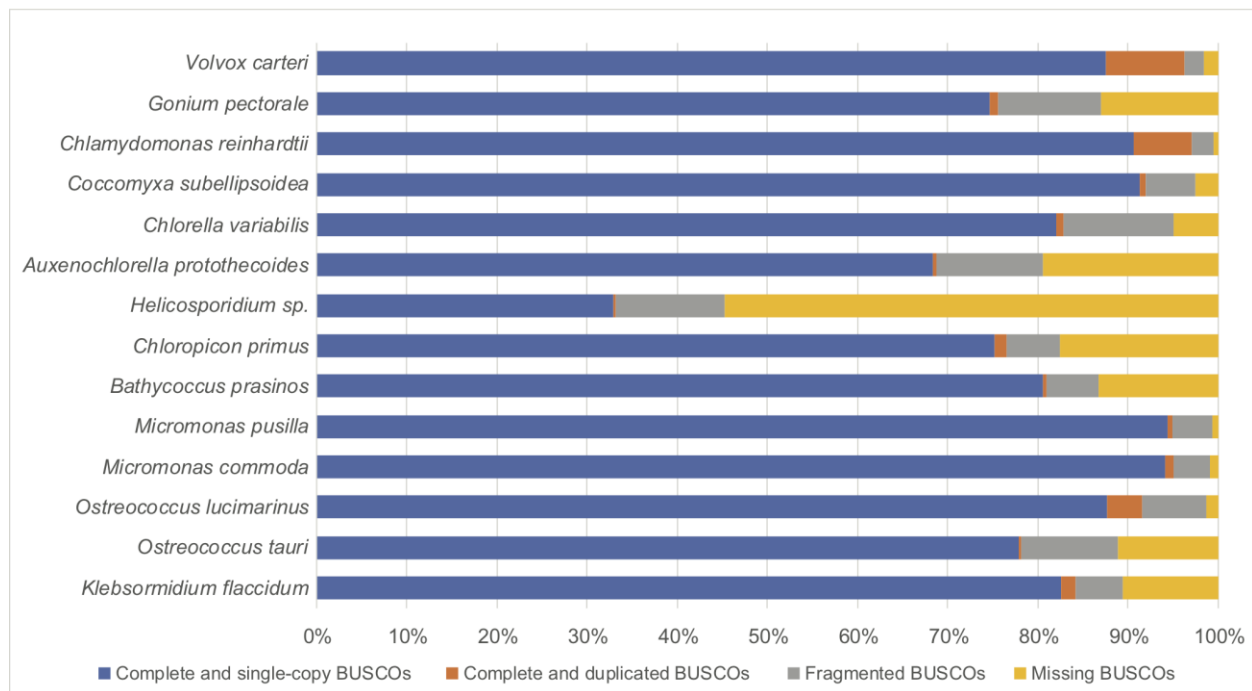
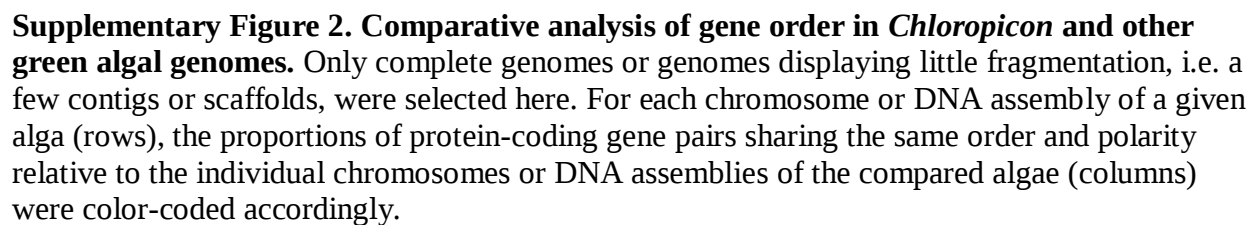


**A streamlined and predominantly diploid genome in the tiny marine green
alga *Chloropicon primus***

Lemieux *et al.*

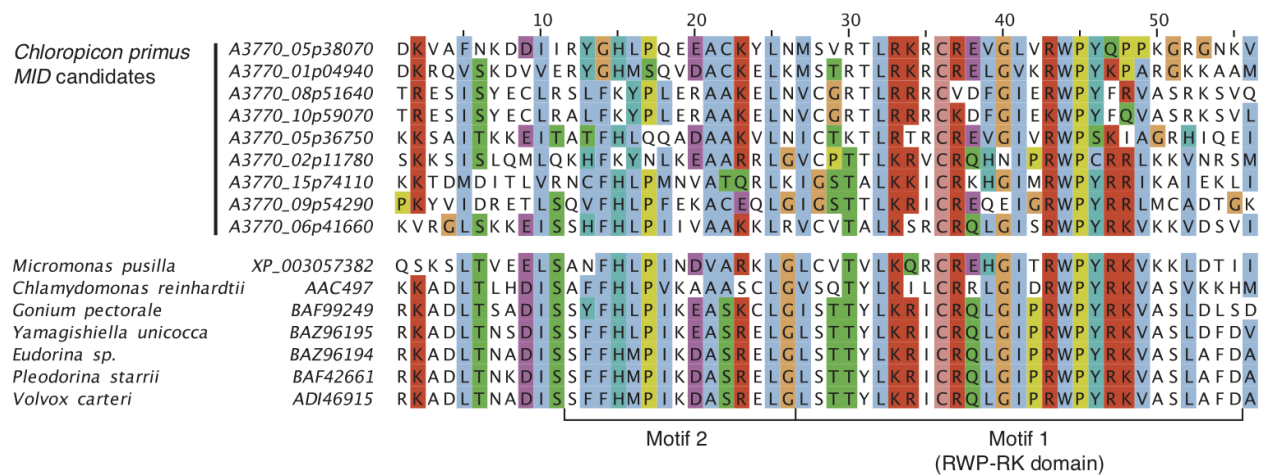


Supplementary Figure 1. BUSCO assessment of sequence assembly completeness of the *Chloropicon* and 13 other green algal genomes.

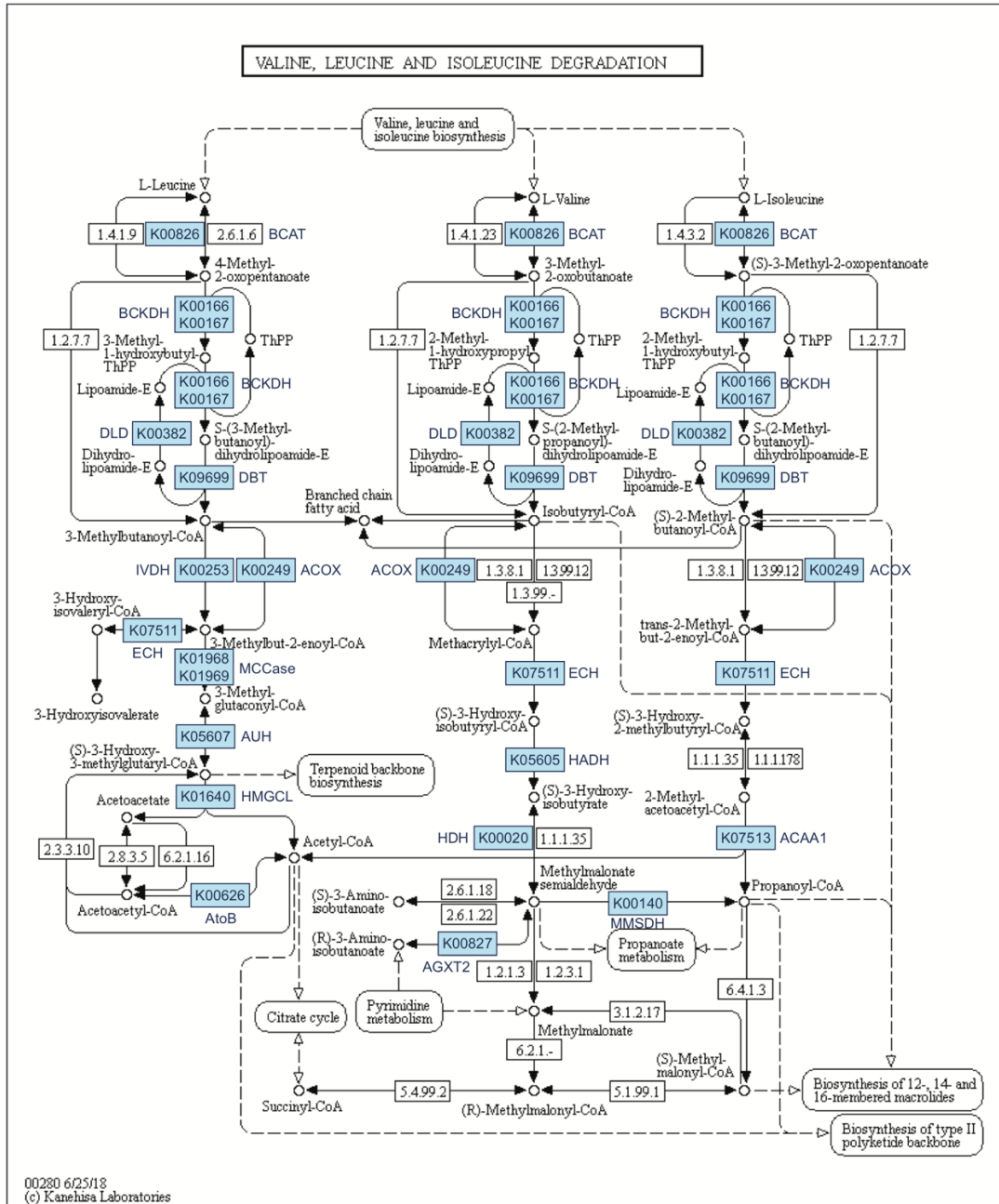




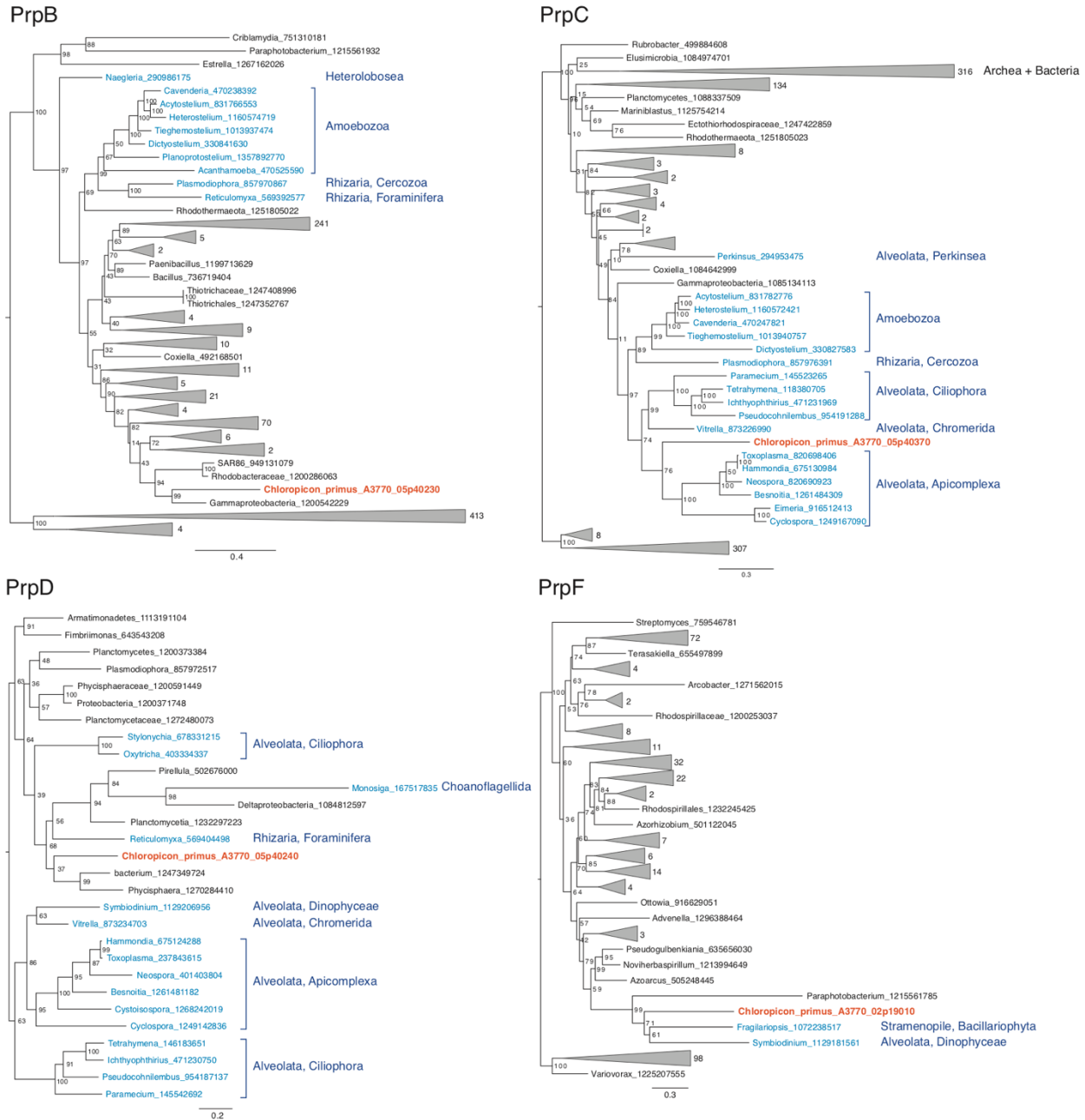
Supplementary Figure 3. Chromosomal locations of polymorphisms belonging to two categories of allelic frequencies. Polymorphisms types include SNP and small indels. 40-60%: blue; 20-39% + 61-80%: orange.



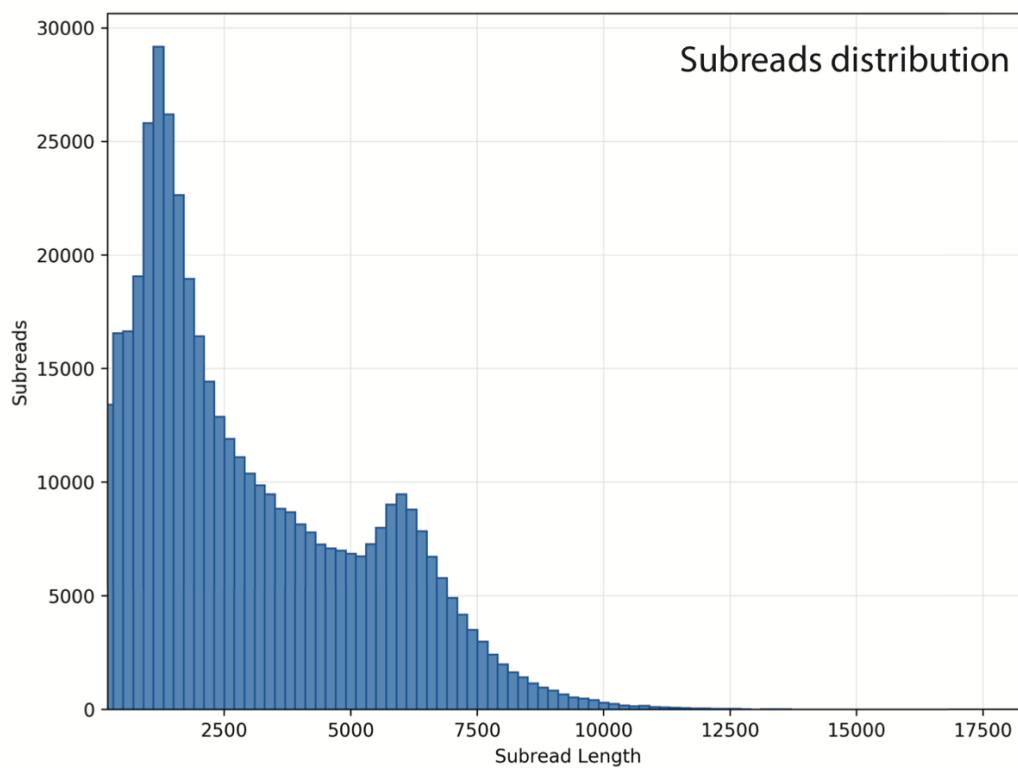
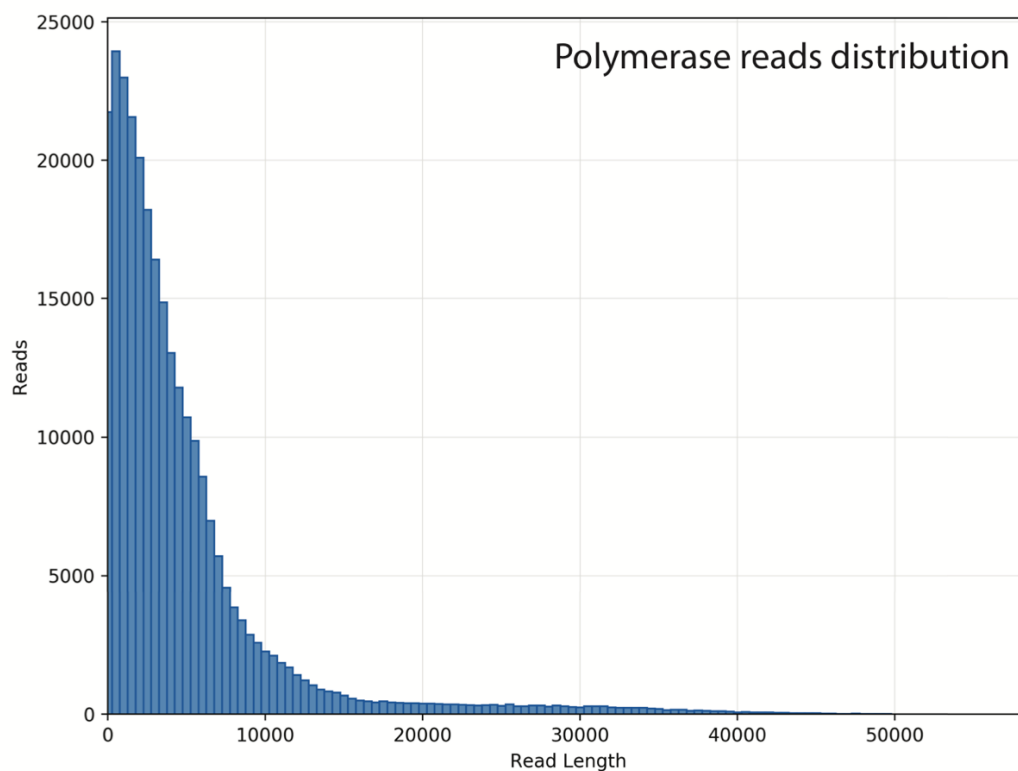
Supplementary Figure 4. Sequence conservation of motifs 1 and 2 among the nine putative RWP-RK transcription factors identified in *Chloropicon* and the *Mid* proteins of *Micromonas* and volvocine algae.



Supplementary Figure 7. Mapping of *Chloropicon* predicted proteins (blue boxes) on the KEGG pathways for BCAA catabolism (ko00280). Note that presence of the enzyme catalysing the conversion of 3-hydroxy-2-methylbutyryl-CoA to 2-methyl-acetoacetyl-CoA [1.1.1.35/1.1.1.178] in the isoleucine pathway could not be inferred because no protein with this activity has been experimentally validated and annotated in any member of the Chloroplastida. Supplementary Data 14 reports the positions of the encoding genes, protein names, presence/absence of homologs in compared green algae as well as other information, including expression data and predicted cellular localization.



Supplementary Figure 8. Phylogenetic relationships of the *Chloropicon* PrpB, PrpC, PrpD and PrpF proteins with their prokaryotic and eukaryotic homologs. To infer these relationships, maximum likelihood trees were constructed using IQ-TREE; bootstrap support values are reported on the nodes of the consensus trees. Prokaryotic clades with more than 90% bootstrap support were collapsed (grey triangles) to simplify the topologies; the numbers of taxa in these clades are indicated on the right of the triangles. Eukaryotic taxa are shown in blue.



Supplementary Figure 9. PacBio polymerase read and subread length distributions.