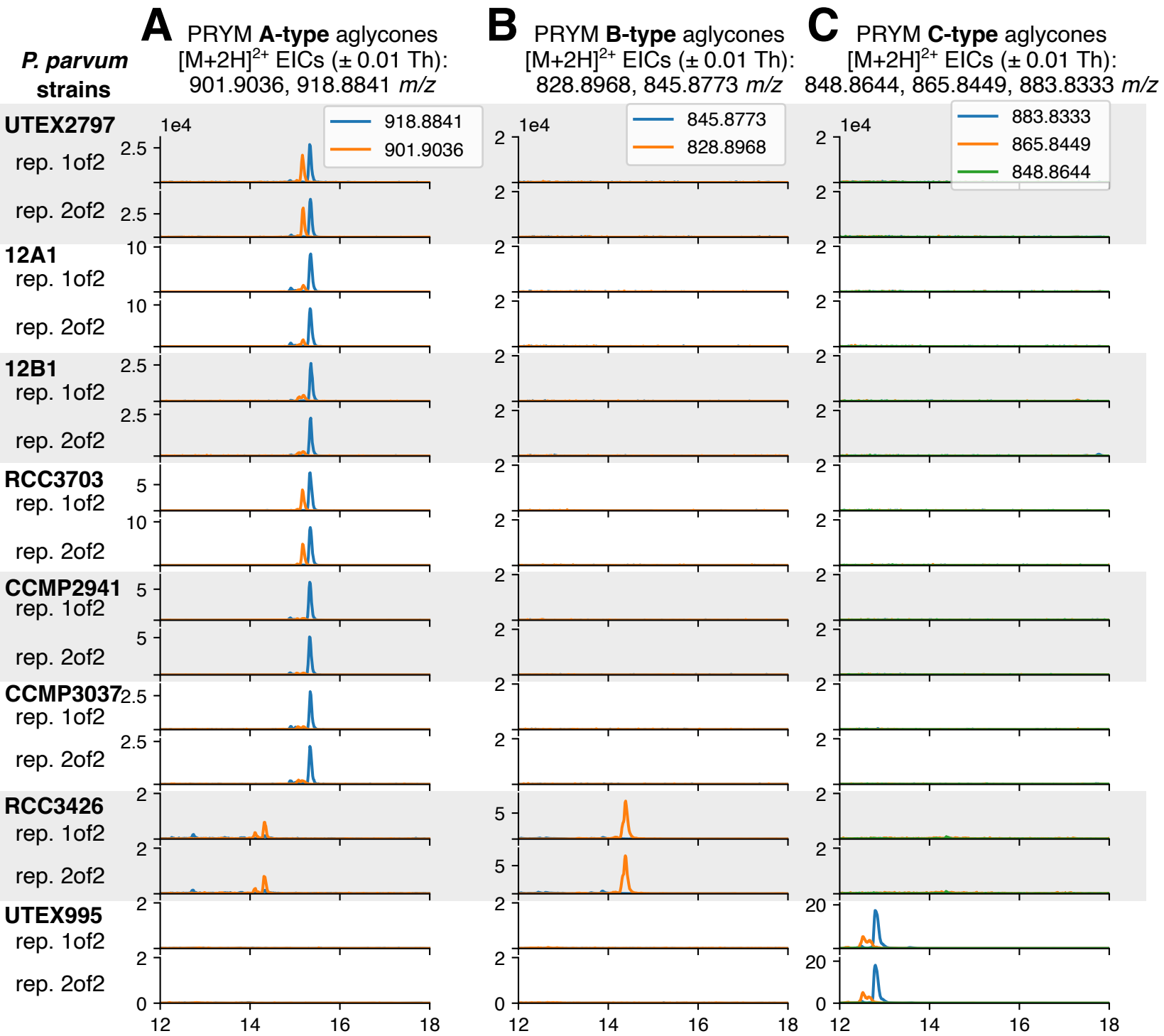


Figure S1. Chemotyping of *P. parvum* strains, Related to Figure 2. Of the 7 strains analyzed, 5 strains (UTEX 2797, 12A1, 12B1, RCC3703, CCMP294) were found to produce A-type prymnesins (PRYM), via LC-HRAM-MS extracted-ion-chromatograms (EICs) of representative [M+2H]2+ aglycone in-source fragment ions. RCC3426 was found to produce B-type prymnesins, and UTEX 995 was found to produce C-type prymnesins. (A) EICs of 2 calculated m/z values for representative PRYM A-type aglycone [M+2H]2+ ions (Cl2: 901.9036 m/z, Cl3: 918.8841 m/z; maintext ref 19). Some strain-to-strain variation in PRYM A-type congener proportions is present. (B) EICs of 2 representative PRYM B-type aglycone [M+2H]2+ ions (Cl2: 828.8968 m/z, Cl3: 845.8773 m/z; from maintext ref 19). (C) EICs of 3 representative PRYM C-type aglycone [M+2H]2+ ions (Cl1: 848.8644 m/z, Cl2: 865.8449 m/z, Cl3: 883.8333 m/z; from maintext ref 19). Rep=Biological replicate. Figure production scripts are available on Figshare 10.6084/m9.figshare.22267066. Raw data is available on Metabolights at study ID MTBLS5893.

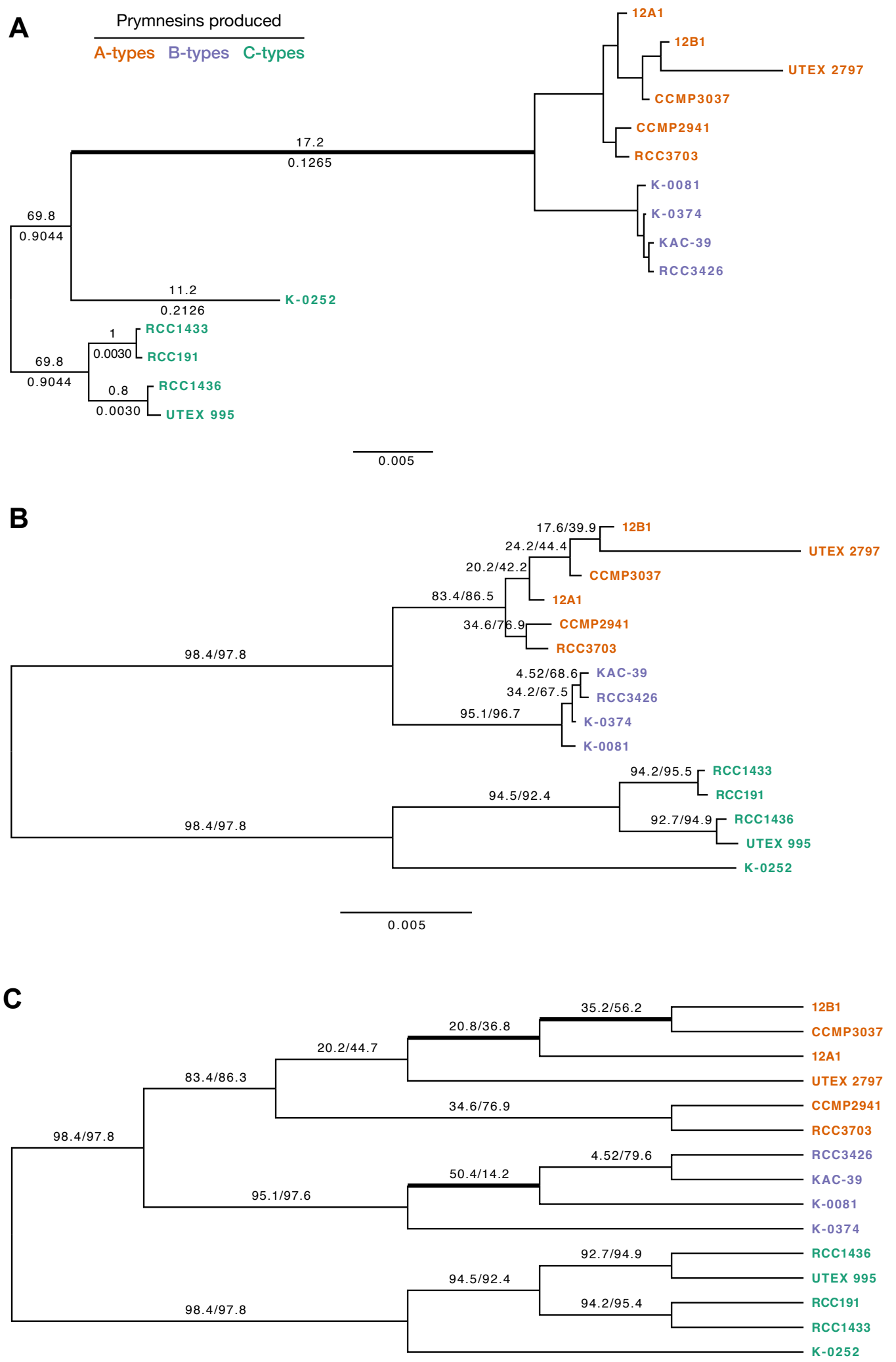


Figure. S2. Phylogenetic relationships of 15 *Prymnesium parvum* strains based on 2699 single-copy genes, Related to Figure 2. A) ML phylogeny constructed based on the concatenated nucleotide data matrix. Numbers above branches indicate IQ-TREE rootstrap support, and numbers below branches indicate p-value of approximately unbiased (AU) test for alternative root locations (branches with p-values < 0.05 were considered a significantly worse root location compared to displayed root). Unlabeled branches had rootstrap support = 0 and pAU < 0.001. Bold branch indicates the branch that was ultimately selected as the root of the species tree for all downstream analyses based on rates of synonymous substitution (see maintext Fig. 2). B) Rerooted ML phylogeny shown in part A; numbers above branches indicate the gene concordance factor/site concordance factors for descendent nodes. C) Species tree inferred from an ASTRAL coalescence-based analysis; numbers above branches indicate the gene concordance factor/site concordance factors for descendent nodes. Thick branches show conflicts between the concatenation-based phylogeny.

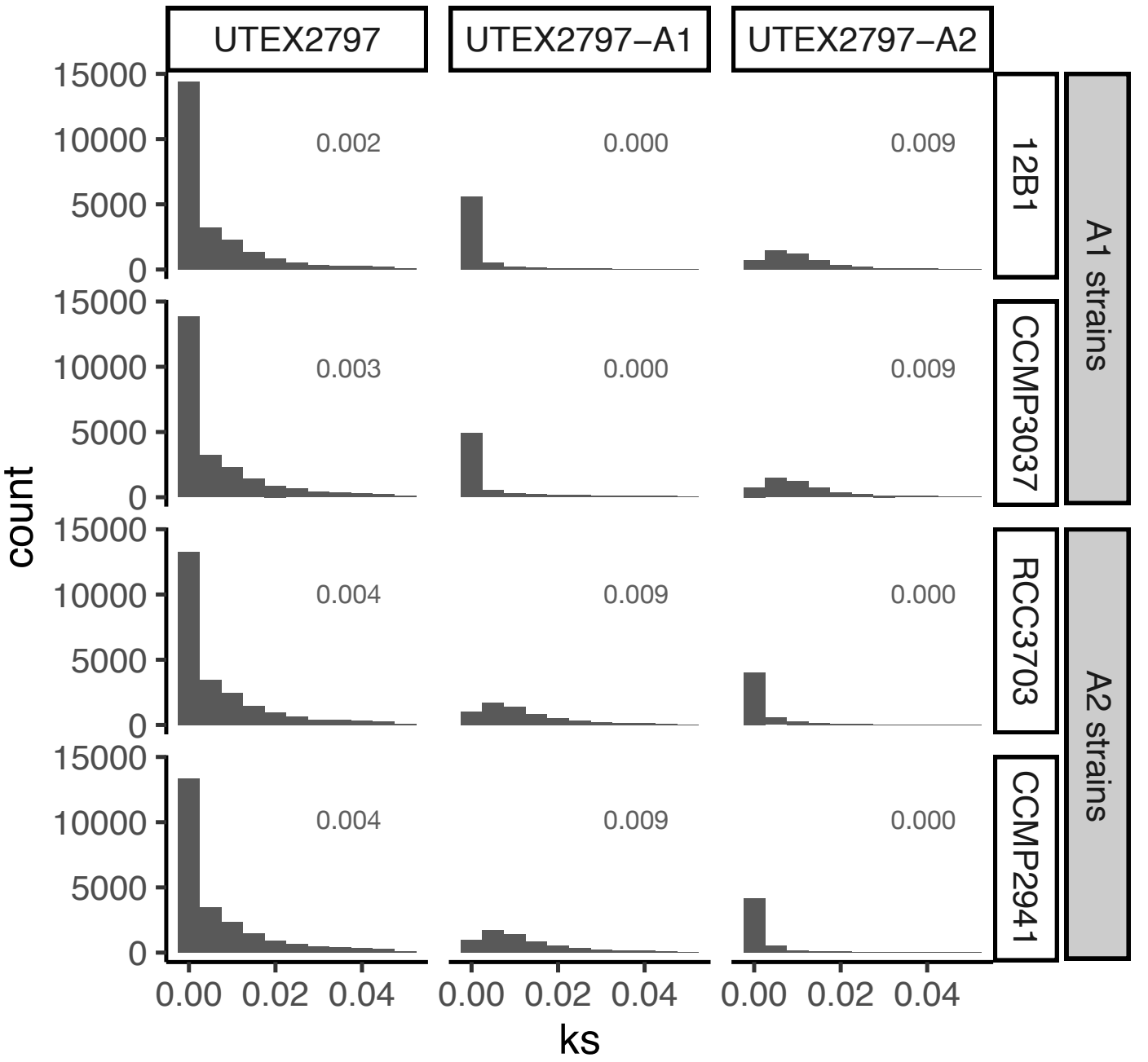
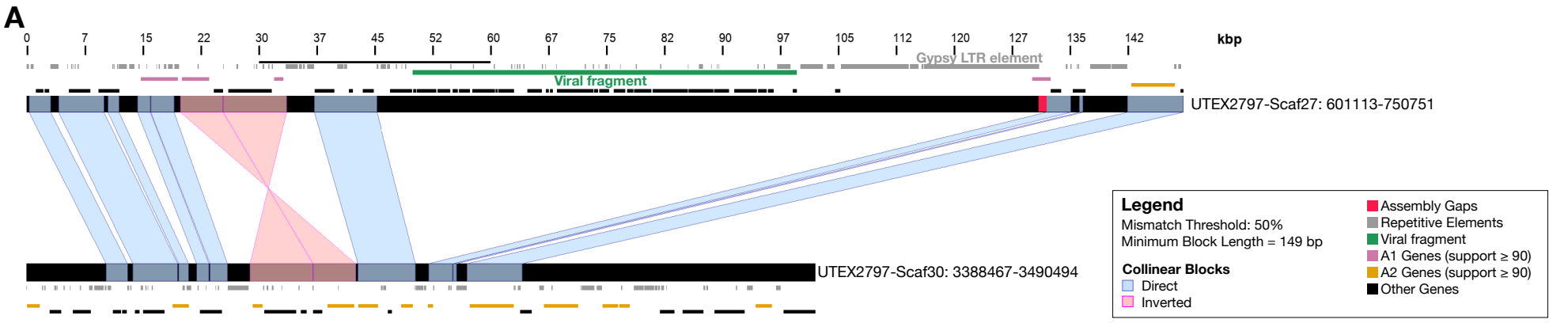


Figure S3. *Ks* plots for UTEX 2797 subgenomes, Related to Figure 3. Distribution of substitutions per synonymous site (*Ks*) between all UTEX 2797 genes compared to those that group with A1 clade strains and those that group with A2 type strains. These gene sets were aligned to homologs in A1 strains (12B1 and CCMP3037) and A2 strains (RCC3703 and CCMP2941). To avoid signal from distant paralogs and/or poor alignments, gene pairs with *Ks* values > 0.05 were excluded. Gray numbers indicate median *Ks* values for each distribution.

Figure S4. Maximum likelihood phylogenies of horizontal gene transfer events in *Prymnesium parvum*, Related to Figure 6. Nodes with IQ-TREE support values > 90 are indicated by numbers on the preceding branch. The branches and outer color bars are color-coded to match the taxonomic classification of each sequence. Trees are midpoint rooted. Triangles indicate collapsed clades that have been reduced for clarity. Triangle color indicates the taxonomy of the collapsed taxa; black triangles indicate that downstream taxa belong to multiple lineages. Leaf IDs hav been truncated for legibility. See FigShare data repository for alignment, newick tree, and full size figure for each HGT: https://doi.org/10.6084/m9.figshare.21376500.

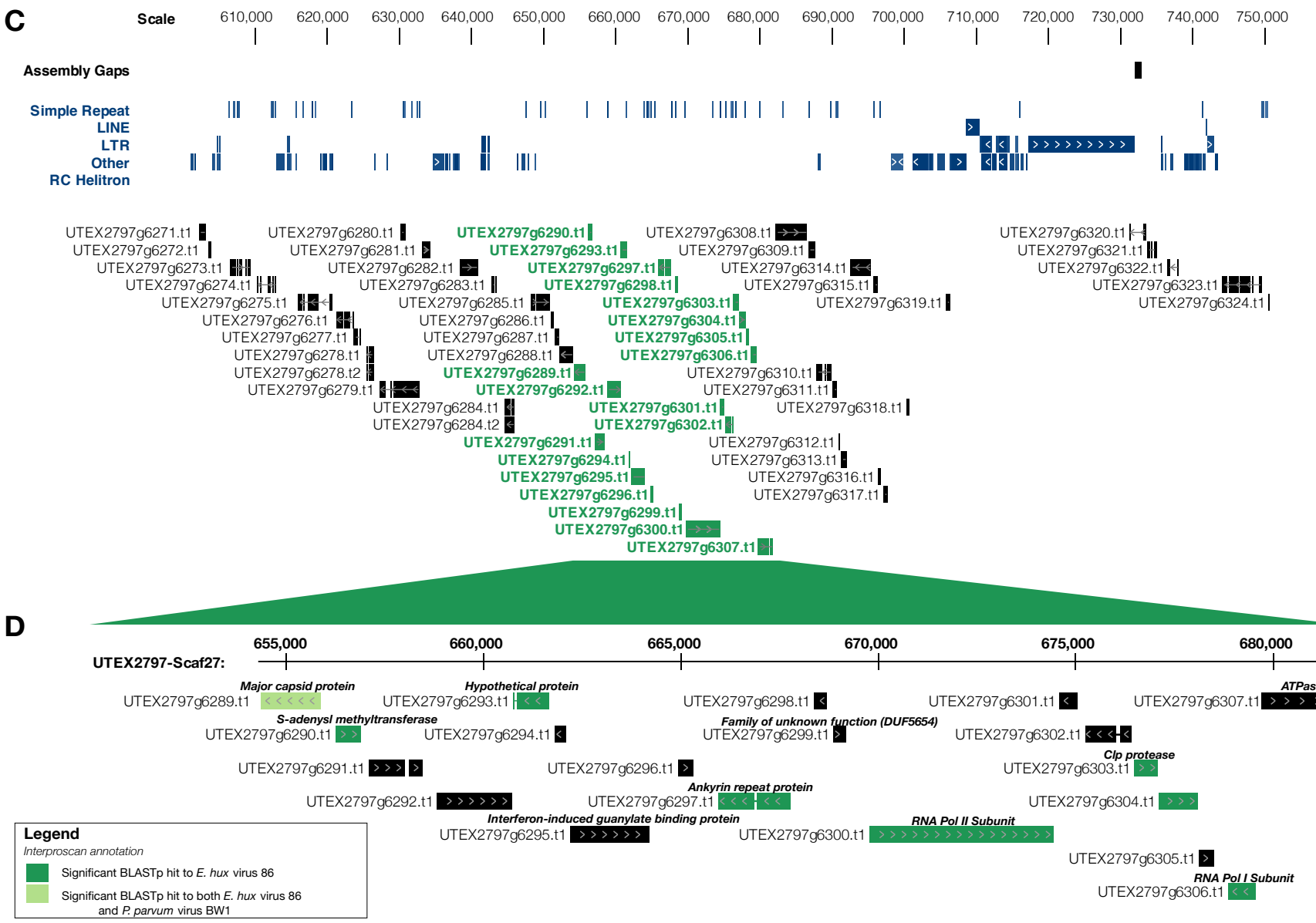
A



B

query	Best hit to Ehux virus 86	Bitscore (Ehux virus 86)	Best hit to Pparvum virus PpDNAV	Bitscore (Pparvum virus PpDNAV)
UTEX2797g6289	YP_293839.2	2.71E-136	AQV04380.1	2.71E-47
UTEX2797g6290	YP_293810.1	1.44E-35	--	--
UTEX2797g6293	YP_293830.1	6.62E-15	--	--
UTEX2797g6297	YP_294212.1	0.00000104	--	--
UTEX2797g6300	YP_294192.1	0	--	--
UTEX2797g6303	YP 293886.1	6.49E-44	--	--
UTEX2797g6306	YP_294157.1	8.67E-17	--	--

C



D

Figure S5. Characterization of the HGT11 viral insertion, Related to Figure 6. (A) Synteny visualization of the HGT11 viral insertion on UTEX2797-Scaf27 (top) with the homeologous region on UTEX2797-Scaf30 (bottom) via XMatchView. Collinear sequences are joined by blue bands if in the same orientation or pink bands if inverted. Repetitive elements are indicated as gray boxes. Genes are differentially colored to indicate whether they group phylogenetically with A1 or A2 strains (pink or orange, respectively) or neither (black). The up- and downstream boundaries of the putative viral fragment is marked by a green bar, which reflect the first and last gene that has BLAST hits to viral genes. (B) Table of best BLASTp hits to the *E. huxleyi* virus 86 and *P. parvum* virus PpDNAV. (C) Genomic view along UTEX2797-Scaf27 surrounding the insertion. Gaps (black) and repeats (blue) are denoted as separate tracks. Genes that are putatively of viral origin are green, while genes that are not are black. (D) Zoomed in view of the viral-like genes (green genes in part C). Genes with a significant BLASTp hit to the *E. huxleyi* virus 86 (NC_007346.1) are dark green, while genes that had a significant hit to both *E. huxleyi* virus 86 and *P. parvum* virus PpDNAV (AQV04380.1) are light green. If a gene was functionally annotated through InterProScan, the annotation is provided above the gene model.

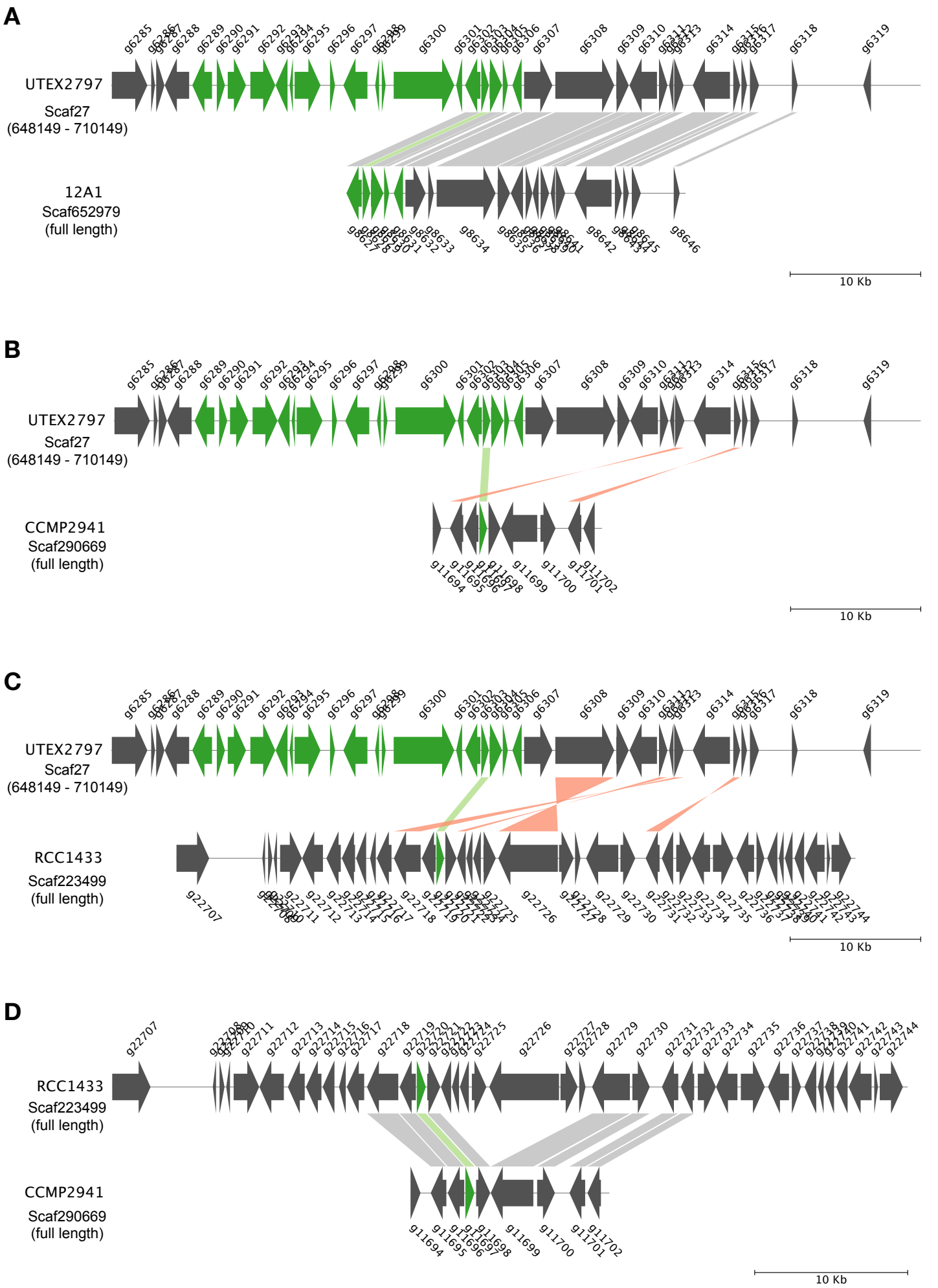


Figure S6. Comparison of shared synteny in region of HGT11 viral insertion, Related to Figure 6. Genome regions are visualized with pyGenomeViz. HGT11 consists of g6303 in UTEX 2797; g8628 in 12A1; g11697 in CCMP2941; and g22720 in RCC1433 and is absent in all other eleven strains in the analysis. Light green bands indicate the location of HGT11 genes; all other collinear sequences are joined by grey bands if in the same orientation or orange bands if inverted. Green genes in UTEX 2797 indicate the genes putatively acquired via a single HGT event (See Fig. S5). In 12A1, Scaf652979 is syntenic across its entire length with UTEX 2797, including five genes of the putative viral insertion (part A), suggesting that 12A1 and UTEX 2797 likely share the same insertion event. In CCMP2941, Scaf290669 is syntenic across its entire length with Scaf223499 in RCC1433, but neither regions are syntentic with UTEX 2797 despite sharing a common viral gene (parts B-D). This suggests that the HGT event giving rise to the viral gene in RCC1433 and CCMP2941 is likely separate from that of UTEX2797 and 12A1. However, long-read assemblies are required for 12A1, CCMP2941, and RCC1433 to rule out viral contamination in Illumina-only assemblies.