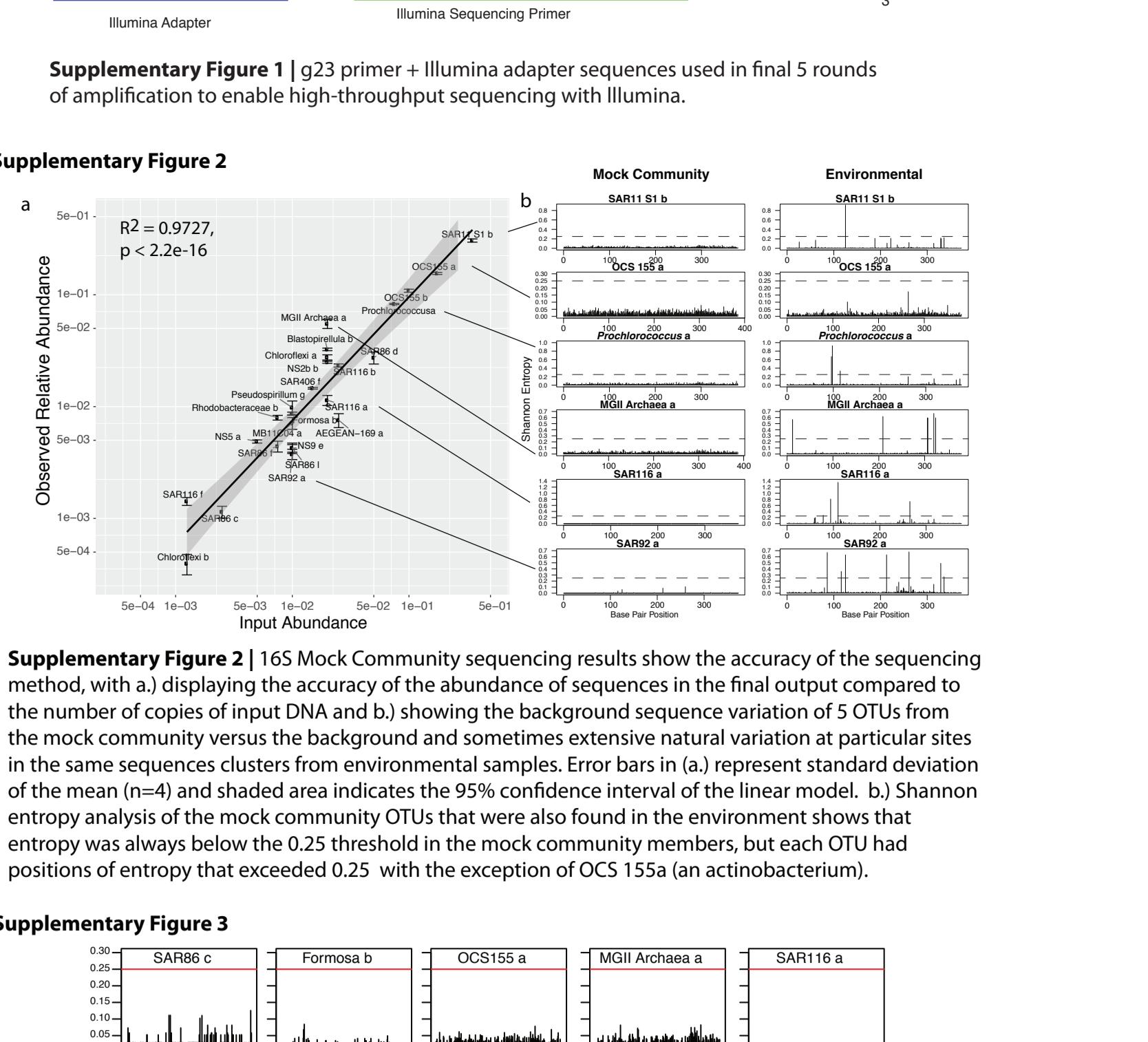
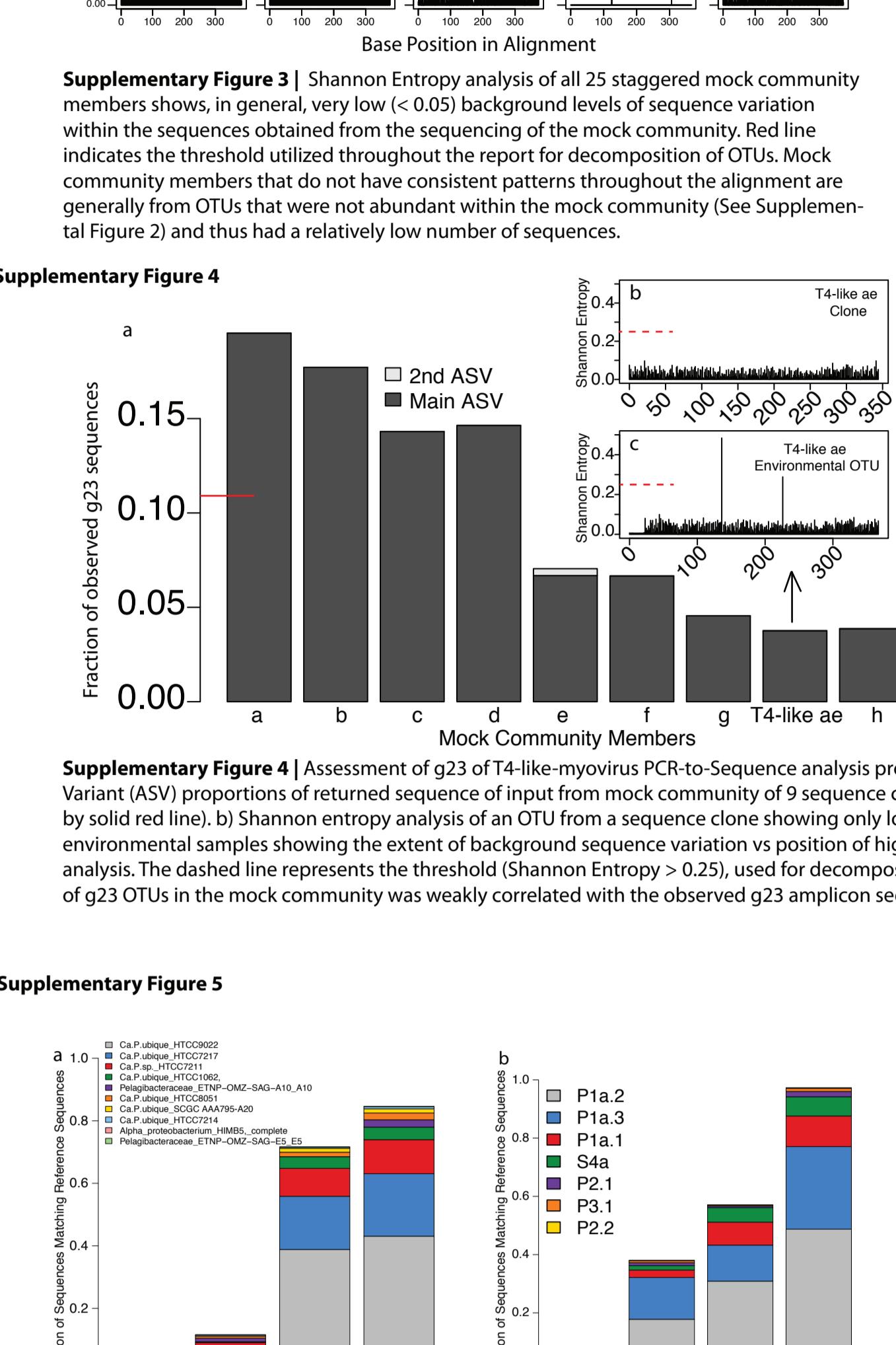


Supplementary Figure 1



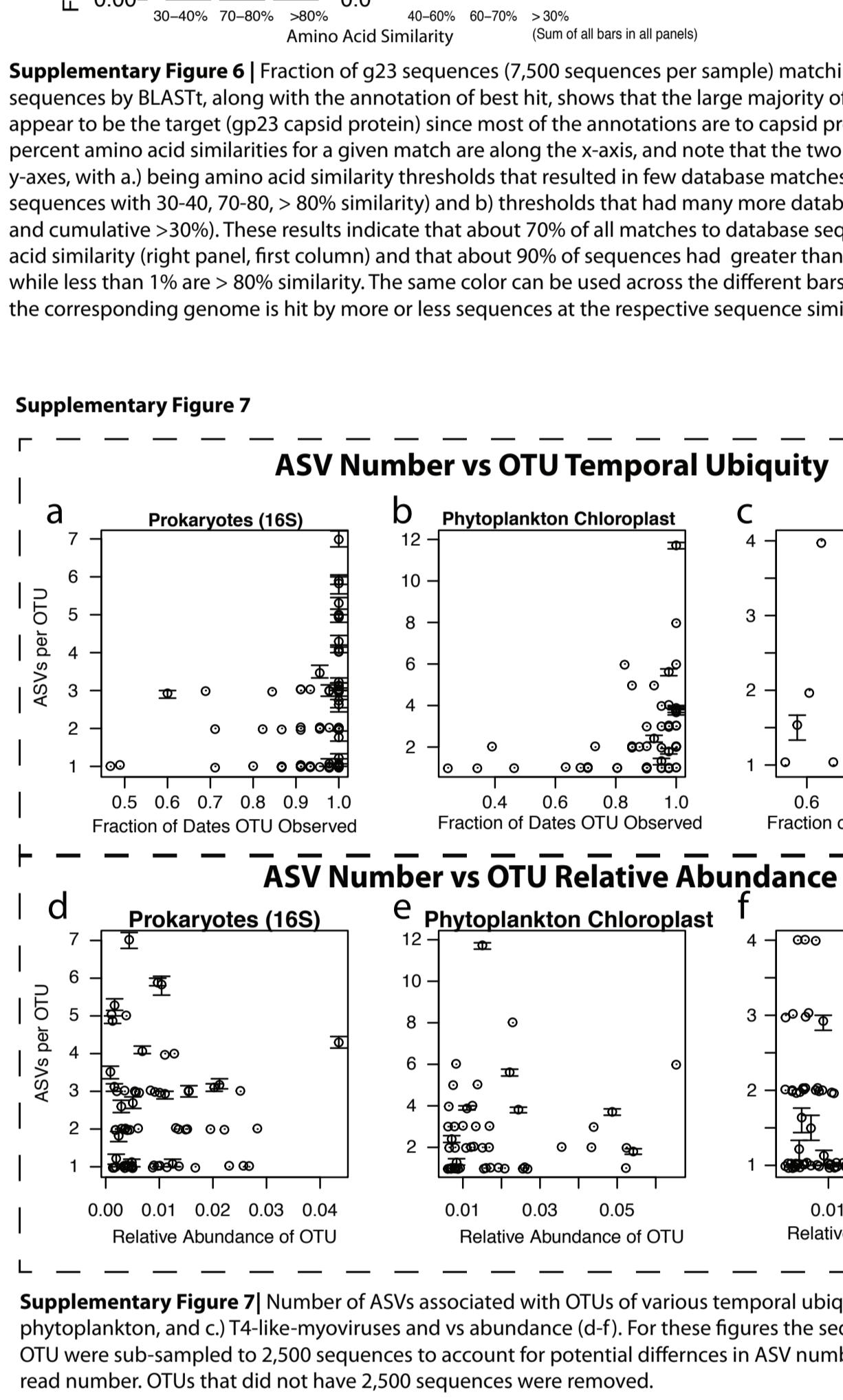
Supplementary Figure 1 | g23 primer + Illumina adapter sequences used in final 5 rounds of amplification to enable high-throughput sequencing with Illumina.

Supplementary Figure 2



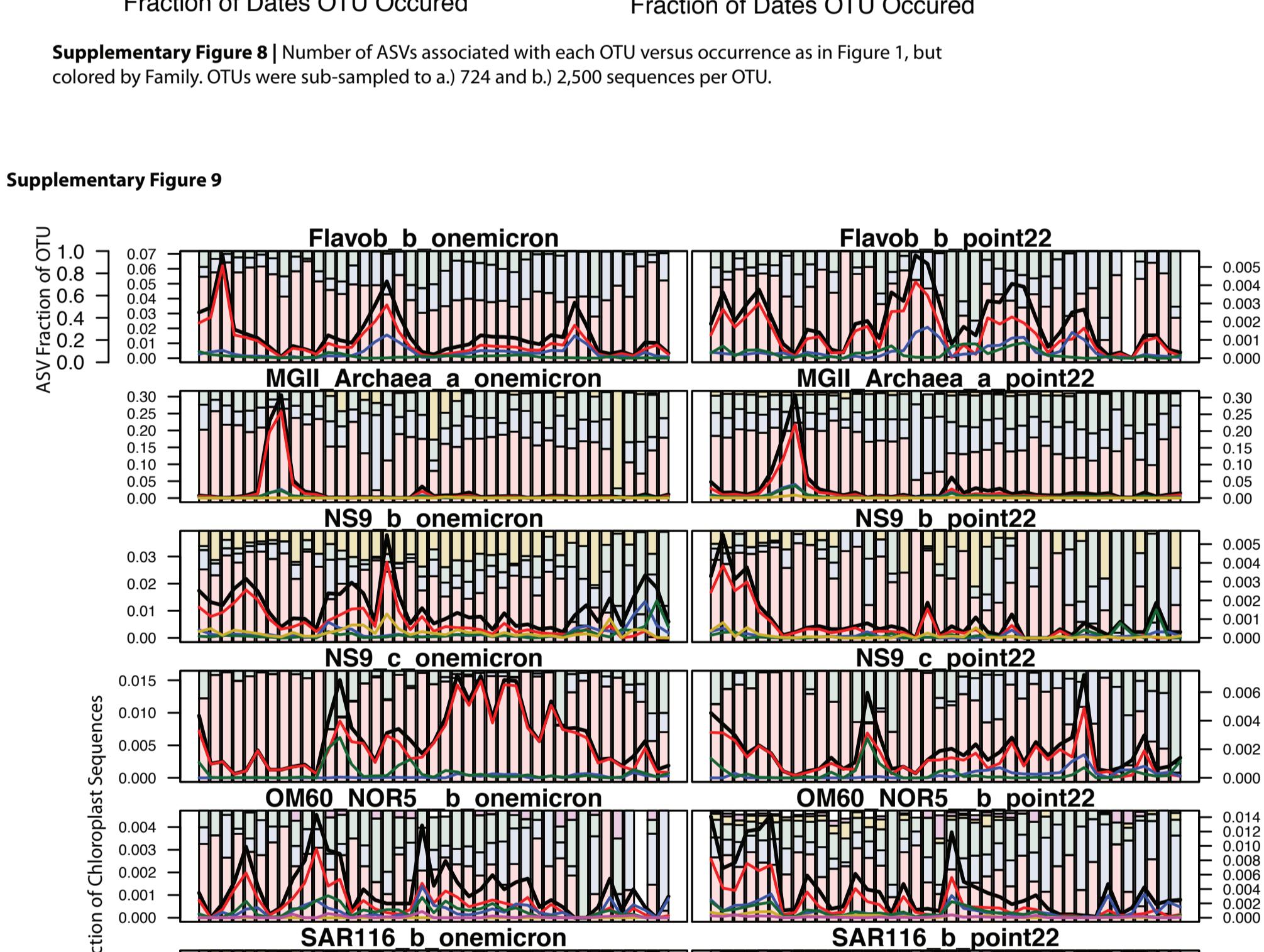
Supplementary Figure 2 | 16S Mock Community sequencing results show the accuracy of the sequencing method, with a.) displaying the accuracy of the abundance of sequences in the final output compared to the number of copies of input DNA and b.) showing the background sequence variation of 5 OTUs from the mock community versus the background and sometimes extensive natural variation at particular sites in the same sequences clusters from environmental samples. Error bars in (a.) represent standard deviation of the mean ($n=4$) and shaded area indicates the 95% confidence interval of the linear model. b.) Shannon entropy analysis of the mock community OTUs that were also found in the environment shows that entropy was always below the 0.25 threshold in the mock community members, but each OTU had positions of entropy that exceeded 0.25 with the exception of OCS155a (an actinobacterium).

Supplementary Figure 3



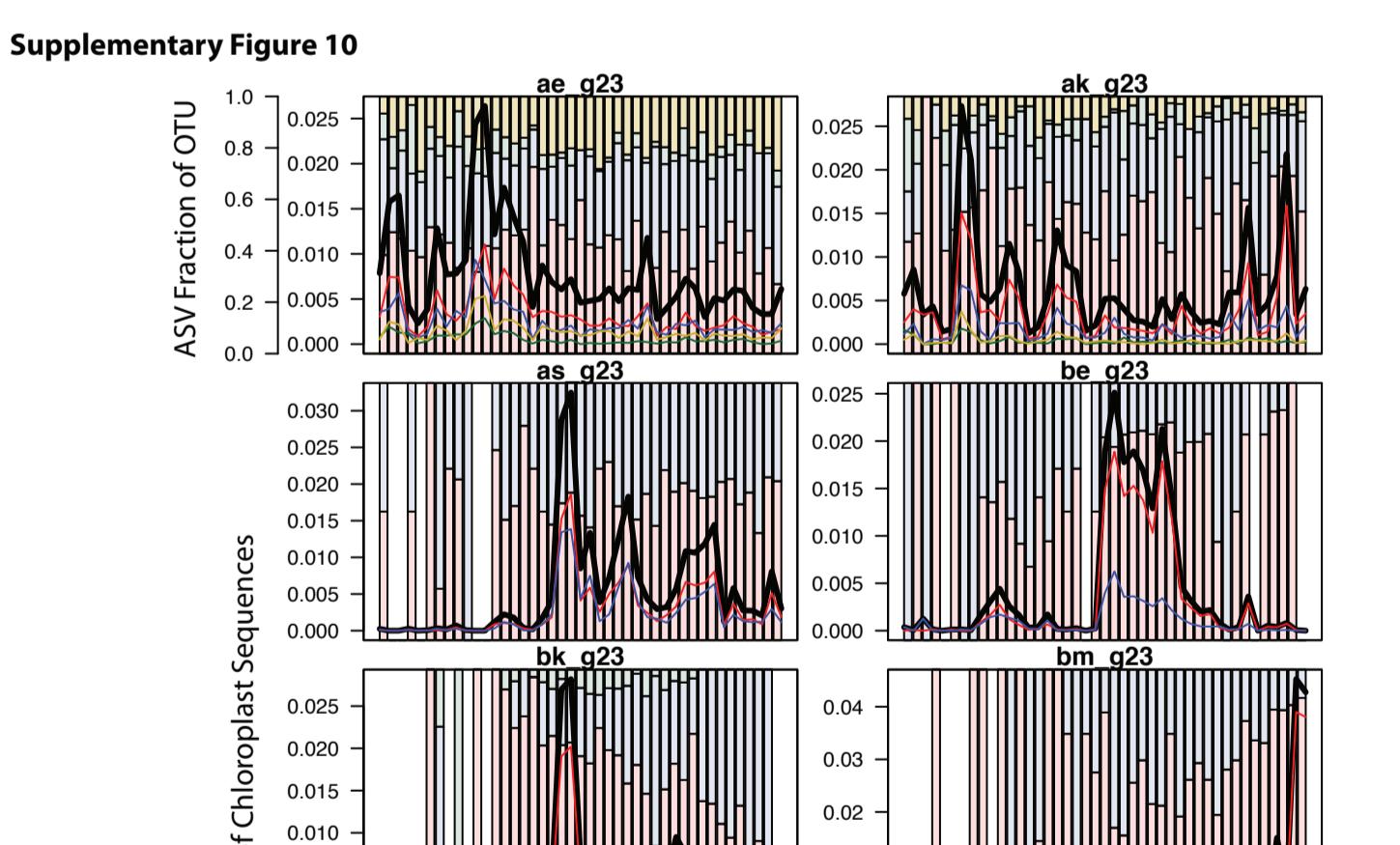
Supplementary Figure 3 | Shannon Entropy analysis of all 51 staggered mock community members shows, in general, very low (< 0.05) background levels of sequence variation within the sequences obtained from the sequencing of the mock community. Red line indicates the threshold utilized throughout the report for decomposition of OTUs. Mock community members that do not have consistent patterns throughout the alignment are generally from OTUs that were not abundant within the mock community (see Supplementary Figure 2) and thus had a relatively low number of sequences.

Supplementary Figure 4



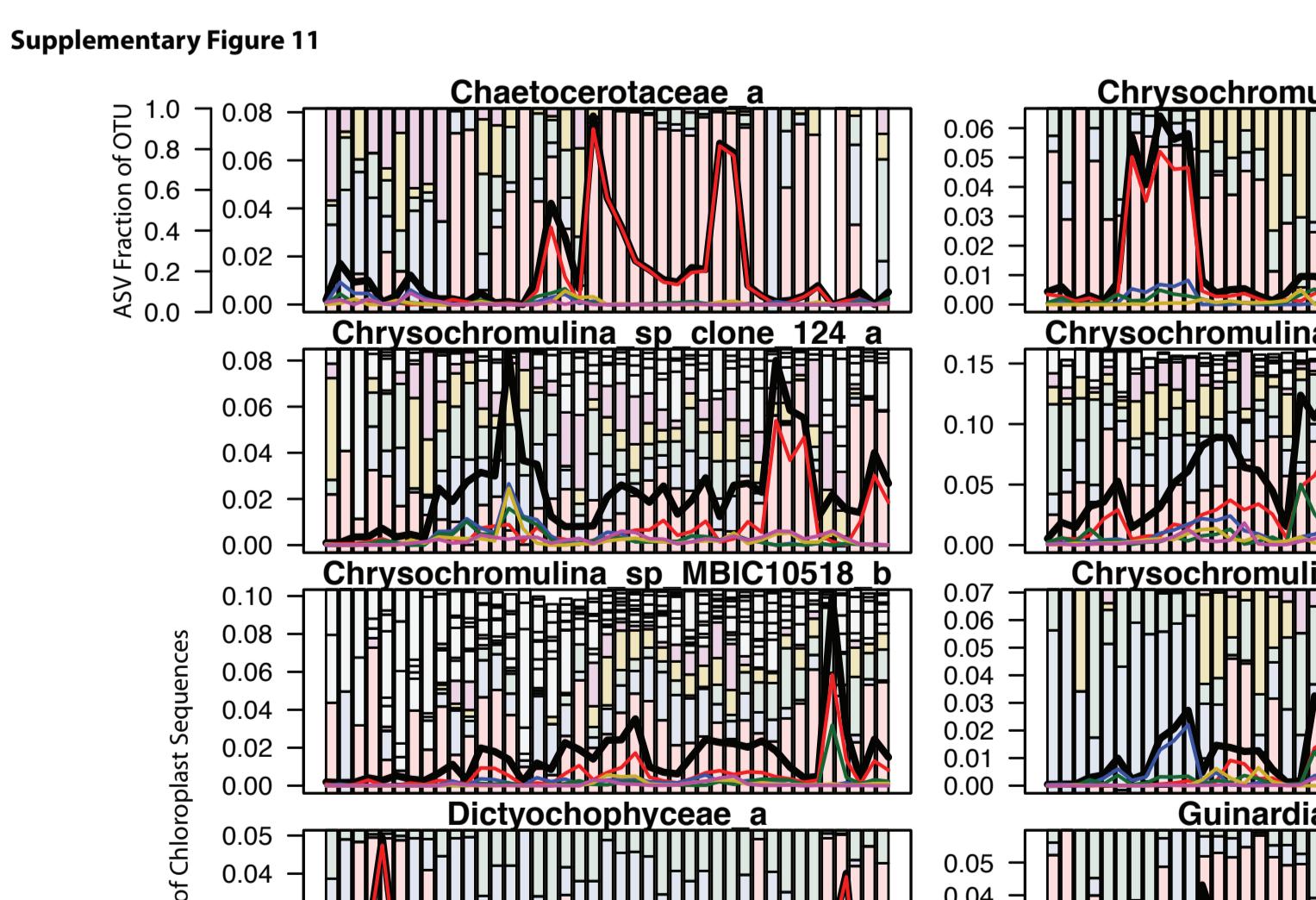
Supplementary Figure 4 | Assessment of g23 of T4-like-myoviruses PCR-to-Sequence analysis procedure via Mock Community. a) Amplicon Sequence Variant (ASV) proportions of returned sequence of input from mock community of 9 sequence clones that input at a theoretical 11% each (indicated by solid red line). b) Shannon entropy analysis of an OTU from a sequence clone showing only low amounts of entropy and c) the same OTU from an environmental sample showing the extent of background sequence variation vs position of high entropy by which the OTU would be split in ASV analysis. The dashed line represents the threshold (Shannon Entropy > 0.25), used for decomposition of sequence clusters. d) Observed abundance of g23 OTUs in the mock community was weakly correlated with the observed g23 amplicon sequence length.

Supplementary Figure 5



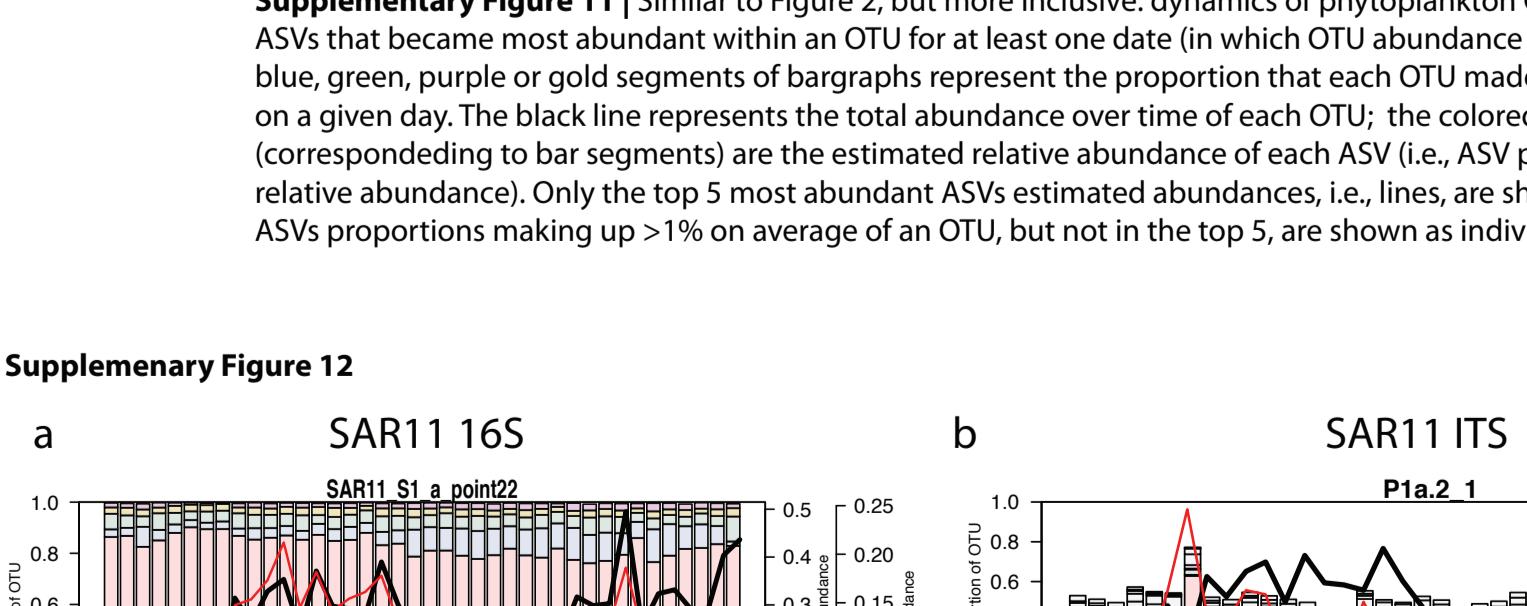
Supplementary Figure 5 | Fraction of SAR11 ITS sequences (7,500 sequences per sample) that match by BLASTn search against (a.) NCBI genomic reference sequences and (b.) a custom SAR11 ITS database from genomic, metagenomic, and environmental cloned sequences. Sequence identifiers and classifications in (b.) are as defined Brown et al. 2012, where "P" stands for "Phylotype", "S" for SAR11 Surface 4. Sequences tend to be very similar to the reference sequences as indicated by sequence similarity thresholds (bottom).

Supplementary Figure 6



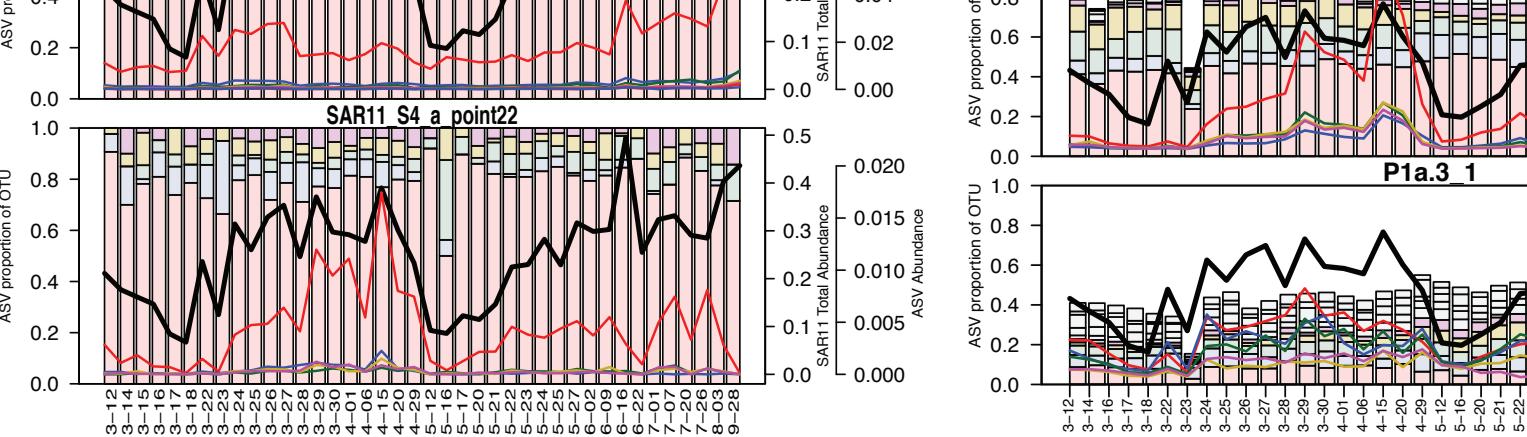
Supplementary Figure 6 | Fraction of g23 sequences (7,500 sequences per sample) matching to viral reference genomic sequences by BLASTn, along with the annotation of best hit, shows that the large majority of g23 sequences obtained appear to be the target (gp23 capsid protein) since most of the annotations are to capsid protein gene sequences. The percent amino acid similarities for a given match are along the x-axis, and note that the two panels have very different y-axes, with a) being amino acid similarity thresholds that resulted in few database matches (which we found to apply to sequences with 30-40, 70-80, >80% similarity) and b) thresholds that had many more database matches (40-60, 60-70, and cumulative >30%). These results indicate that about 70% of all matches to database sequences were at 40-60% amino acid similarity (right panel, first column) and that about 90% of sequences had greater than 30% amino acid similarity, while less than 1% are >80% similarity. The same color can be used across the different bars, and have different heights if the corresponding genome is hit by more or less sequences at the respective sequence similarity thresholds.

Supplementary Figure 7



Supplementary Figure 7 | Number of ASVs associated with OTUs of various temporal ubiquity for a.) prokaryotes, b.) phytoplankton, and c.) T4-like-myoviruses and vs abundance (d-f). For these figures the sequences composing each OTU were sub-sampled to 2,500 sequences to account for potential differences in ASV number that could be due to read number. OTUs that did not have 2,500 sequences were removed.

Supplementary Figure 8



Supplementary Figure 8 | Number of ASVs associated with each OTU versus occurrence as in Figure 1, but colored by Family. OTUs were sub-sampled to a.) 724 and b.) 2,500 sequences per OTU.

Supplementary Figure 9

Supplementary Figure 9 | Similar to Figure 3, but more inclusive: dynamics of bacterial and archaeal OTUs which had > 2 ASVs that became most abundant within an OTU for at least one date in both size fractions (in which OTU abundance was >0.1%). The bar segments and lines are the same as described in Figure 3, where bar segments represent the proportion that each ASV made up of an OTU on a given day. The black line represents the total abundance over time of each OTU; the color-coded lines correspond to bar segments (i.e., ASV proportion of OTU * relative abundance). Only the top 5 most abundant ASVs estimated abundances, i.e., lines, are shown for each OTU. All ASVs proportions making up >1% on average of an OTU, but not in the top 5, are shown as individual gray bar segments.

Supplementary Figure 10



Supplementary Figure 10 | Dynamics of g23 OTUs which had > 1 ASVs become most abundant for at least one date in which OTU abundance was >0.1% in either size fraction March-September 2011. As in Figure 4d, segments of bargraphs represent the proportion that each ASV made up of an OTU over time. The black line represents the total abundance over time of each OTU; the color-coded lines correspond to bar segments (i.e., ASV proportion of OTU * relative abundance).

Supplementary Figure 11

Supplementary Figure 11 | Similar to Figure 2, but more inclusive: dynamics of phytoplankton OTUs which had > 2 ASVs that became most abundant within an OTU for at least one date (in which OTU abundance was >0.1%). The red, blue, green, purple or gold segments of bargraphs represent the proportion that each ASV made up of an OTU on a given day, a green star indicates the top 5 most abundant ASVs estimated abundances, i.e., lines, are shown for each OTU. Only the top 5 most abundant ASVs estimated abundances, i.e., lines, are shown for each OTU. ASVs proportions making up >1% on average of an OTU, but not in the top 5, are shown as individual gray bar segments.

Supplementary Figure 12

Supplementary Figure 12 | Similar to Figure 5, but more inclusive: dynamics of SAR11 a.) 16S and b.) ITS OTUs and ASVs March-September 2011. As in Figure 3, bar segments refer to the proportion of each ASV on an OTU on a given day and the corresponding color coded lines are the ASV proportion of each ASV made up of an OTU on a given day. The black line represents the total abundance over time of each ASV; the top 5 for each ASV are shown as separate lines. For the SAR11 ITS ASVs there is a separate scale for the ASVs and total SAR11 OTU abundance since each ASV tended to make up less than 10% of the OTU.