TvA screenshot of a computer screen

AI-generated content may be incorrect.

Figure 2.1: A phylogenomic tree constructed using 38 aligned and concatenated ribosomal protein sequences from the 52 cyano-mOTUs. Ribosomal proteins were annotated using the Anvi’o “anvi-run-hmms” pipeline for the Bacteria\_71 ribosomal dataset [(Eren et al., 2021; Hyatt et al., 2010; Lee, 2019)](https://www.zotero.org/google-docs/?c1CJNj). The final alignment contained 6,583 aligned positions. The tree was created using the FastTree default option within Anvi’o [(Eren et al., 2021; Price et al., 2009)](https://www.zotero.org/google-docs/?VaQV92) and visualized using iTOL [(Letunic & Bork, 2021)](https://www.zotero.org/google-docs/?lLLjfR). The tree was re-rooted with the Vampirovibrionia (Obscuribacteraceae) genomes, within the iTOL interface. The tree tips are annotated with symbols to denote the taxon’s inferred morphology (filamentous - green square, colonial - orange star, stalked - pink triangle, and solitary - purple single). Calculated bootstrap values between 0.1 and 0.9 are shown. The color strip on the right denotes clades that we defined for the current study (Pseudanabaena, Nodosilinea, Filamentous-1, Snowella, Microcystis, Cyanobium, Solitary-1, and Vulcanococcus). These assigned colors will be used in the proceeding figures.

A screenshot of a graph

AI-generated content may be incorrect.Figure 2.2: Comparison between mapped metagenomes filtered at 92% ANI or 95% ANI during inStrain processing. **A**-**B.** Boxplots showing the difference in total number of mapped reads (A) or cumulative cyano-mOTU coverage (B) within a metagenome that passed the inStrain minimum read ANI threshold. **C-D.** Boxplots of coverage unevenness (C) and nucleotide diversity (D) for each cyano-mOTU between the different read ANI filtering levels. Coverage unevenness is calculated by subtracting the measured breadth from the calculated expected breadth. The x axis tip labels are arranged by tree order (figure 3.1) and colored by assigned clade ID. For **A-D**, boxplots were drawn centered around the median. Upper and lower bounds of the box represent the 25th and 75th quartile range. Lines extend to data points within 1.5x the interquartile range (IQR) and outliers were hidden. Non-parametric Wilcoxon Rank Sum Test (α = 0.05) was used to compare the averages between the tested pair. Significance levels in C-D are shown as stars (\* = p-value < 0.05, \*\* = p-value < 0.01, \*\*\* = p-value < 0.001, \*\*\*\* = p-value < 0.0001) or “ns” (p-value > 0.05).

A graph of different colored lines

AI-generated content may be incorrect.

Figure 2.3: Normalized coverage for the cyano-mOTUs at an mOTU and clade level across the TYMEFLIES dataset. Mean genome coverage was obtained through inStrain [(Olm et al., 2021)](https://www.zotero.org/google-docs/?0mCCHn) with a 95% ANI read-mapping cutoff and normalized using a ratio between average reads mapped across the time series to reads mapped within a sample. **A.** Boxplot of normalized coverage for each individual cyano-mOTU with the number of samples in which they were detected. Average relative abundance (RA) is shown by the difference in height of the dark grey bars to the right. Cyano-mOTUs are arranged by tip order from the phylogenomic tree in Figure 2.1. Boxplots are drawn with the line drawn through the median, box upper and lower bounds cover the 25th and 75th quartile range (IQR), lines extend to 1.5x the IQR, and outliers have been hidden. **B.** Phenological trends in normalized coverage averaged across all years. Area under the line is shaded by clade groups contribution to the total. A moving average with a window size of a week was used to smooth the time series and account for different sample dates within each week, across years. The x axis tip labels in A and shaded lines in B are colored by our defined clades (light green - Pseudanabaena, green - Nodosilinea, dark green - Filamentous-1, pink - Snowella,  orange - Microcystis, light blue - Cyanobium,  blue - Solitary-1, purple - Vulcanococcus).

A collage of graphs and charts

AI-generated content may be incorrect.

Figure 2.4: Nucleotide diversity (π) for cyano-mOTUs organized by morphology (A) and clade (B). Individual data points are a single metagenome - cyano-mOTU pair, with pi calculated by the pipeline inStrain [(Olm et al., 2021)](https://www.zotero.org/google-docs/?PHxUP4), after applying a threshold requirement of at least 10x coverage and 80% breadth. **A.** Differences in π among different morphological groups (filamentous, colonial, stalked, and solitary). Statistics between morphological groups were performed with Kruskal-Wallis multiple comparisons using the post-hoc Dunn method [(1964)](https://www.zotero.org/google-docs/?DHq6Nv) via the R package “ggpubr” [(Kassambara, 2023)](https://www.zotero.org/google-docs/?AOf0Il). P values adjusted with the Holm method [(Holm, 1979)](https://www.zotero.org/google-docs/?WaQStR) and comparisons are identified by a line and significance level through stars (\*\*\*; p value < 0.001, \*\*\*\*; p value < 0.0001). **B.** Differences in π among the cyano-mOTU clades **C.** Relationship between the cyano-mOTUs average normalized coverage and average pi, across the timeseries. Boxplots in A and B are drawn with the line drawn through the median, box upper and lower bounds cover the 25th and 75th quartile range (IQR) and lines extend to 1.5x the IQR with outliers beyond the IQR are shown. All plots, including C, are colored by the predefined themes in Figure 2.1 (light green - *Pseudanabaena*, green - *Nodosilinea*, dark green - Filamentous-1, pink - *Snowella*,  orange - *Microcystis*, light blue - *Cyanobium*,  blue - Solitary-1, purple - *Vulcanococcus*).

A graph of different colored lines

AI-generated content may be incorrect.

Figure 2.5: Interannual patterns of change in nucleotide diversity (π) for cyano-mOTUs that passed a “minimum-presence” threshold; cyano-mOTUs must be present in over 20 samples with at least 10x coverage and 80% breadth. Plots show the average annual pi with each line representing one mOTU, across the time series. Cyano-mOTUs are divided into separate panels depending on their clade designation from Figure 3.1. Time series for each individual cyano-mOTU can be found in supplemental information (Figure S2.1). The table to the right is a brief summary of the linear regression output for pi versus year since the first time point (Table S2.2).

A close-up of a graph

AI-generated content may be incorrect.

Figure 2.6: Compilation of nucleotide diversity, coverage, and STRONG weighted coverage for Microcystis aeruginosa (MCYST\_2) and Aphanizomenon flos-aquae (APHAN\_134). **A-B.** Top panels are normalized coverage (solid black line) and nucleotide diversity (colored, dash line) in 2015 for MCYST\_2 (A) and APHAN\_134 (B). Bottom panels are theSTRONG weighted coverage for the 12 unique haplotypes in the MCYST\_2 (C) and APHAN\_134 (D) populations.

A graph of a number of different numbers

AI-generated content may be incorrect.

Figure 2.7: Linkage disequilibrium (LD) using raw and normalized D’ and r2 across 18 cyano-mOTUs. **A.** Single nucleotide polymorphism (SNP) linkage metrics calculated between SNP pairs that are from 5 to 250 bp distance from each other for MCYST\_2 and APHAN\_134. All identified SNP pairs with the same distance apart were averaged across metagenomes and multiples of 5 are plotted. The number of metagenomes that went into this average is shown in the bottom left of the individual plots. LD for all cyano-mOTUs included in this analysis can be found in Figure S3.3. **B.** Linear relationship between mean π and mean r2 for all mOTUs with point size representing mean D’. Linear regression: y = 0.69 - 33x, R2 = 0.49, p < 0.05. Points are labeled by the numeric text in their mOTU identifier and colored by defined clades (light green - *Pseudanabaena*, green - *Nodosilinea*, dark green - Filamentous-1,  orange - *Microcystis*, light blue - *Cyanobium*,  blue - Solitary-1, purple - *Vulcanococcus*).

A screenshot of a graph

AI-generated content may be incorrect.

Figure 2.8: Relationship between the index of replication (iRep) and nucleotide diversity (π) follow distinct patterns. Cyano-mOTUs selected were in the previous linkage disequilibrium analysis and only include metagenomes with 10x coverage and 80% breadth. Since iRep could not be calculated for every sample that passed this threshold, mOTUs with less than nine data points were removed from the analysis. **A.** Hypothetical diagram for two observed patterns defining the relationship between iRep and pi: asymptotic, a signal for “diversity maintenance”, and exponential, a signal for “sup-population selection.” **B-C.** In B we have highlighted MCYST\_31 and APHAN\_134 as representatives of the hypothetical relationship patterns between iRep and pi. The remaining cyano-mOTUs are shown in panel C. The log of iRep was used in order to reduce the spread of data points on the x-axis. Axes are not set to the same limits. Data point opacity is controlled by raw coverage. The highest opacity includes all coverage values above 100x.