Iroki: automatic customization for phylogenetic trees

Ryan M. Moore[1*], Amelia O. Harrison[2], Sean M. McAllister[2], Rachel L. Marine[3], Clara Chan[2], and K. Eric Wommack[1]

[1]Center for Bioinformatics & Computational Biology, University of Delaware, Newark, DE, USA

[2]School of Marine Science and Policy, University of Delaware, Newark, DE, USA

[3]Department of Biological Sciences, University of Delaware, Newark, DE, USA

Corresponding author's information

*To whom correspondence should be addressed

Address: Delaware Biotechnology Institute, 15 Innovation Way, Newark, Delaware 19711

(Tel): (302) 831-4362

(Fax): (302) 831-3447

(E-mail): moorer@udel.edu

Abstract

Background

Phylogenetic trees are an important analytical tool for examining species and community diversity, and the evolutionary history of species. In the case of microorganisms,

decreasing sequencing costs have enabled researchers to generate ever-larger sequence datasets, which in turn have begun to fill gaps in the evolutionary history of microbial groups. However, phylogenetic analyses of large sequence datasets present challenges to extracting meaningful trends from complex trees. Scientific inferences made by visual inspection of phylogenetic trees can be simplified and enhanced by customizing various parts of the tree, including label color, branch color, and other features. Yet, manual customization is time-consuming and error prone, and programs designed to assist in batch tree customization often require programming experience. To address these limitations, we developed Iroki, a program for fast, automatic customization of phylogenetic trees. Iroki allows the user to incorporate information on a broad range of metadata for each experimental unit represented in the tree.

Results

Iroki was applied to four existing microbial sequence datasets to demonstrate its utility in data exploration and presentation. Iroki was used to highlight connections between viral phylogeny and host taxonomy, to explore the abundance of microbial groups across samples of cattle hide, to examine short-term temporal dynamics of virioplankton communities, and to search for trends in the biogeography of Zetaproteobacteria.

Conclusions

Iroki is an easy-to-use web app and command line application for fast, automatic customization of phylogenetic trees based on user-provided categorical or continuous

metadata. Iroki allows for rapid hypothesis testing through visualizing custom phylogenetic trees, streamlining the process of phylogenetic data exploration and presentation.

Availability

Iroki can be accessed through a web app or via installation through RubyGems, from source, or through the Iroki Docker image. All source code and documentation is available under the GPLv3 license at https://github.com/mooreryan/iroki. The Iroki web-app is accessible at www.iroki.net or through the Virome portal (http://virome.dbi.udel.edu), and its source code is released under GPLv3 license at https://github.com/mooreryan/iroki_web. The Docker image can be found here: https://hub.docker.com/r/mooreryan/iroki.

Keywords

Phylogeny, visualization, sequence analysis, bioinformatics, metagenomics

Iroki: automatic customization for phylogenetic trees

Background

Studies in microbial ecology often use phylogenetic trees as a means for assessing the diversity and evolutionary history of microorganisms. As the cost of sequencing has declined, researchers have been able to gather ever-larger sequence datasets. While large sequence datasets have begun to fill in the gaps in the evolutionary history of microbial groups [1–5]; they have also posed new analytical challenges as extracting meaningful trends within such highly dimensional datasets can be cumbersome. In particular, scientific inferences made by visual inspection of phylogenetic trees can be simplified and enhanced by customizing various parts of the tree including label and branch color, branch width, and other features. Though many tree visualization packages allow for manual modifications [6–9], the process can be time consuming and error prone especially when the tree contains many nodes. While a handful of existing programs address the issue of tree visualization, most are not capable of batch customization and those that do often require programming experience [10–13].

Iroki, a program for fast, automatic customization of phylogenetic trees, was developed to address these limitations and enable users to incorporate information on a broad range of metadata for each experimental unit represented in the tree. Iroki is available for use through a web interface at www.iroki.net, through the Virome portal (http://virome.dbi.udel.edu), and through a UNIX command line tool. Results are saved in

the widely used Nexus format with color metadata tailored for use with FigTree [8] (a freely available and efficient tree viewer).

Implementation

Iroki enhances visualization of phylogenetic trees by coloring node labels and branches according to categorical metadata criteria or numerical data such as abundance information. Iroki can also rename nodes in a batch process according to user specifications so that node names are more descriptive. A tree file in Newick format containing a phylogenetic tree is always required. Additional required input files depend on the operation(s) desired. Coloring functions require a color map or a biom [14] file. Node renaming functions require a name map. The color map, name map, and biom files are created by the user and, along with the Newick file, form the inputs for Iroki.

Explicit tree coloring

Iroki's principle functionality involves coloring node labels and/or branches based on information provided by the user in the color map. The color map text file contains either two or three tab-delimited columns depending on how branches and labels are to be colored. Two columns, pattern and color, are used when labels and branches are to have the same color. Three columns, pattern, label color, and branch color, are used when branches and labels are to have different colors. Patterns are searched against node labels either as regular expressions or exact string matches.

Entries in the color column can be any of the 657 named colors in the R programming language [15] (e.g., skyblue, tomato, goldenrod2, lightgray, black) or any valid hexadecimal color code (e.g., #FF78F6). In addition, Iroki provides a 19 color palette with complementary colors based on Kelly's color scheme for maximum contrast [16]. Nodes in the tree that are not in the color map will remain black.

Depending on user-specified options, a pattern match to node label(s) will trigger coloring of the label and/or the branch directly connected to that label. Inner branches will be colored to match their descendent branches if all descendants are the same color, allowing quick identification of common ancestors and clades that share common metadata.

Tree coloring based on numerical data

Iroki provides the ability to generate color gradients based on numerical data, such as absolute or relative abundance, from a tab-delimited biom format file. Single-color gradients use color saturation to illustrate numerical differences, with nodes at a higher level being more saturated than those at a lower level. For example, highly abundant nodes will be represented by more highly saturated colors. Two-color gradients show numerical differences through both color mixing and luminosity. Additionally, the biom file may specify numerical information for one group (e.g., abundance in a particular sample) or for two groups (e.g., abundance in the treatment group vs. abundance in the

control group). For biom files with one group, single- or two-color gradients may be used. However, biom files specifying two-group metadata may only use the two-color gradient.

Renaming nodes

Some packages for generating phylogenetic trees restrict the use of special characters and spaces or require node names to be shorter than a specified length or (RAxML [17], PHYLIP [18], etc.). Name restrictions present challenges to scientific interpretation of phylogenetic trees. Iroki's renaming function uses a two-column, tab-delimited name map to associate current node names, exactly matching those in the tree file, with new names. The new name column has no restrictions on name length or character type. Iroki ensures name uniqueness by appending integers to the ends of names, if necessary.

Combining the color map, name map, and biom files

Iroki can be used to make complex combinations of customizations by combining the color map, name map, and biom files. For example, a biom file can be used to apply a color gradient based on numerical data to the labels of a tree, a color map can be used to separately color the branches based on user-specified conditions, and a name map can be used to rename nodes in a single command or web request. Iroki follows a specific order of precedence when applying multiple customizations. First, the color gradient inferred from the biom file is applied. Next, the color map is applied to specified labels or branches, overriding the gradient applied in the previous step if necessary. Finally, the name map is used to map current names to the new names (Fig. 1).

148

149 *Output*

150 Iroki outputs the modified tree in the Nexus format. When building the phylogenetic tree,

151 FigTree uses the Nexus format file and interprets the color metadata output of Iroki.

152

153 **Results & Discussion**

154

155 *Global diversity of bacteriophage*

156 Viruses are the most abundant biological entity on Earth, providing an enormous reservoir

157 of genetic diversity, driving evolution of their hosts, influencing composition of microbial

158 communities, and affecting global biogeochemical cycles [19,20]. The current viral

159 taxonomic system is based on a suite of physical characteristics of the virion rather than

160 on genome sequences. The phage proteomic tree, created to provide a genome-based

161 taxonomic system for bacteriophage classification [21], was recently updated to include

162 hundreds of new phage genomes from the Phage SEED reference database [22], as well as

163 long assembled contigs from viral shotgun metagenomes (viromes) collected from the

164 Chesapeake Bay (SERC) [23] and the Mediterranean Sea [24].

165

166 Taxonomy and host information metadata was collected for the viral genome sequences, a

167 color map was created to assign colors based on viral family and host phyla, and Iroki was

168 used to add color metadata to branches and labels of the phage proteomic tree. Since a

169 large number of colors were required on the tree, Iroki's Kelly color palette was used to
170 provide clear color contrasts. The tree was rendered with FigTree (Fig. 2).
171
172 Adding color to the phage proteomic tree with Iroki shows trends in the data that would
173 be difficult to discern without color. Uncultured phage contigs from the SERC and
174 Mediterranean viromes make up a large portion of all phage sequences shown on the tree,
175 and are widely distributed among known phage. In general, viruses in the same family
176 claded together, e.g., branch coloring highlights large groups of closely related
177 Siphoviridae and Myoviridae. This label-coloring scheme also shows that viruses infecting
178 hosts within same phylum are, in general, phylogenetically similar. For example, viruses
179 within one of the multiple large groups of Siphoviridae across the tree infect almost
180 exclusively host species within the same phylum, e.g., Siphoviridae infecting
181 Actinobacteria clade away from Siphoviridae infecting Firmicutes or Proteobacteria.
182
183 *Bacterial community diversity and prevalence of E. coli in beef cattle*
184 Shiga toxin-producing *Escherichia coli* (STEC) are dangerous human pathogens that
185 colonize the lower gastrointestinal tracts of cattle and other ruminants. STEC-
186 contaminated beef and STEC shed in the feces of these animals are major sources of
187 foodborne illness. To identify possible interactions between STEC populations and the
188 commensal cattle microbiome, a recent study examined the diversity of the bacterial
189 community associated with beef cattle hide [25]. Fecal and hide samples were collected
190 over twelve weeks and SSU rRNA amplicon libraries were constructed and analyzed by

191 Illumina sequencing [26]. The study indicated that the community structure of hide
192 bacterial communities was altered when the hides were positive for STEC contamination.
193
194 Iroki was used to visualize changes in the relative abundance of each cattle hide bacterial
195 OTU according to the presence or absence of STEC. A Mann-Whitney U test comparing
196 OTU abundance between STEC positive and STEC negative samples was performed, and
197 those bacterial OTUs showing a significant change in relative abundance ($p < 0.5$) were
198 placed on a phylogenetic tree according to the 16S rRNA sequence. Branches of the tree
199 were colored based on whether there was a significant change in relative abundance with
200 STEC contamination (red: $p <= 0.05$, blue: $p > 0.05$). Node labels were colored along a
201 blue-green color gradient representing the abundance ratio of OTUs between samples
202 with STEC (blue) and without (green). Additionally, label luminosity was determined
203 based on overall abundance (lighter: less abundant, darker: more abundant) (Fig. 3). Iroki
204 makes it clear that most OTUs on the tree showed a significant difference in abundance
205 (branch coloring) between STEC positive and STEC negative samples (node coloring).
206 Furthermore, we can see that most OTUs are at low abundance with only a few highly
207 abundant OTUs (label luminosity). The color gradient added by Iroki allows us to see that
208 the abundant OTUs were evolutionarily distant from one another and thus spread out
209 across many phylogenetic groups.
210
211 Iroki can be used to quickly test hypotheses without investing a large amount of time
212 annotating trees manually. A UPGMA tree was created based on unweighted UniFrac

213 distance [27] between 356 bacterial community profiles based on SSU rRNA amplicon
214 sequences from cattle hide and feces samples (Fig. 4). Iroki was used to evaluate
215 similarities in sample bacterial communities according to the sampling location. Iroki
216 colored branches based on whether the sample originated from feces (blue) or from hide
217 (red). The coloring added by Iroki shows a clear partitioning of bacterial communities on
218 the tree based on their sampling location (hide or feces). However, four fecal samples
219 claded with hide samples, and two hide samples claded with fecal samples, highlighting
220 the ability of Iroki to easily identify good candidates for more in-depth examination.
221 Additionally, Iroki was used to illustrate a correlation between one of the most abundant
222 bacterial families, Ruminococcaceae, and sampling location. Iroki colored node labels
223 with a color gradient based on Ruminococcaceae family abundance, utilizing both a
224 single color gradient (Fig. 4A) and a two color gradient (Fig. 4B). Custom trees were
225 visualized using FigTree. Iroki's automatic color gradient and ability to label branches
226 and nodes based on different criteria clearly show that Ruminococcaceae is more
227 abundant in fecal samples than in hide samples.
228
229 *Short-term dynamics of virioplankton*
230 The gene encoding Ribonuclotide reductase (RNR) is common within viral genomes and
231 thus can be used as a marker gene for studying viral diversity. Moreover, RNR
232 polymorphism is predictive of some of the biological and ecological features of viral
233 populations [28]. A mesocosm experiment examined the short-term dynamics of phage
234 populations using RNR amplicon sequences, specifically, sequences of class II RNRs of

235 bacteriophages infecting cyanobacterial hosts. A phylogenetic tree was created from the
236 Cyano II RNR amplicon sequences and Iroki was used to color nodes and branches based
237 on the time point (0 h, 6 h, 12 h) at which each amplicon sequence was observed. The
238 customized tree was then visualized using FigTree (Fig. 5). Iroki's coloring showed that no
239 phylogenetic clade was dominated by OTUs observed in any particular time point; rather,
240 time points were spread relatively evenly across clades. This analysis demonstrates Iroki's
241 utility for exploring sequence datasets, allowing the researcher to quickly and easily test
242 hypotheses.
243
244 *Phylogeny of Zetaproteobacteria within a biogeographic context*
245 Biogeographical studies assess the distribution of an organism's biodiversity across space
246 and time. The extent to which microorganisms exhibit biogeography is an open question
247 in microbial ecology. The isolated nature of the microbial communities associated with
248 deep-ocean hydrothermal vents provides an ideal system for studying the biogeography of
249 microbes. In particular, iron-oxidizing bacteria have been shown to thrive in vent fluids,
250 sediments, and iron-rich microbial mats associated with the vents. Globally, iron-
251 oxidizing bacteria make significant contributions to the iron and carbon cycles. A recent
252 study analyzed multiple SSU rRNA clone libraries to investigate the biogeography of
253 Zetaproteobacteria, a phyla containing many iron-oxidizing bacterial species, between
254 three sampling regions of the Pacific Ocean (central Pacific—Loihi seamount, western
255 Pacific—Southern Mariana Trough, and southern Pacific (Vailulu'u Seamount/Tonga
256 Arc/East Lau Spreading Center/Kermadec Arc) [29]. Sequences were aligned and a

257 phylogeny was inferred as described in [29]. Iroki was used to examine the relationship
258 between sampling location and phylotype by adding branch and label color based on
259 geographic location and renaming original node labels with OTU and location metadata.
260 The custom tree was visualized using FigTree (Fig. 6). In some cases, OTUs contained
261 sequences from only one sampling location (e.g., OTUs 12, 15, and 16), whereas other
262 OTUs are distributed among more than one sampling location (e.g., OTUs 1, 2, and 4).
263 Often, sequences sampled from the same geographic location are in the same phylotype
264 despite being members of different OTUs (e.g., OTUs 10 and 19).
265
266 *Availability and requirements*
267 A web-based version of Iroki can be accessed online at www.iroki.net or through the
268 Virome portal (http://virome.dbi.udel.edu/). For users who wish to run Iroki locally, a
269 command line version of the program is installable via RubyGems, from GitHub
270 (https://github.com/mooreryan/iroki). A Docker image is available for users who desire the
271 flexibility of the command line tool, but do not want to install Iroki or manage its
272 dependencies (https://hub.docker.com/r/mooreryan/iroki). Docker is a convenient method
273 for packaging an application with all of its dependencies that is guaranteed to run the
274 same regardless of the user's environment [30,31]. The README provided with the source
275 code provides detailed instructions for setting up and running Iroki. Further
276 documentation and tutorials can be found at the Iroki Wiki
277 (https://github.com/mooreryan/iroki/wiki).
278

License

Iroki and its associated programs are released under the GNU General Public License version 3 [32].

Conclusions

Iroki is a command line program and web app for fast, automatic customization of large phylogenetic trees based on user specified configuration files describing categorical or continuous metadata information. The output files include Nexus tree files with color metadata tailored specifically for use with FigTree. Various example datasets from microbial ecology studies were analyzed to demonstrate Iroki's utility. In each case, Iroki simplified the processes of data exploration, data presentation, and hypothesis testing. Iroki provides a simple and convenient way to rapidly customize phylogenetic trees, especially in cases where the tree in question is too large to annotate manually or in studies with many trees to annotate.

List of Abbreviations

OTU*:* operational taxonomic

RNR*:* Ribonucleotide reductase

STEC: Shiga-toxengenic *Escherichia coli*

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Data and code used to generate figures are available on GitHub at https://github.com/mooreryan/iroki_manuscript_data

Funding

This project was supported by the USDA National Institute of Food and Agriculture award number 2012-68003-30155 and the National Science Foundation Advances in Bioinformatics program (award number DBI_1356374).

Competing Interests

The authors declare that they have no competing interests.

Authors' contributions

RMM and SMM conceived the project. RMM wrote the manuscript and implemented Iroki. AOH and RLM processed and analyzed Cyano II amplicons. All authors read, edited, and approved the final manuscript.

Acknowledgements

We would like to acknowledge Daniel J. Nasko and Jessica M. Chopyk for their work on the phage proteomic tree, and Barbra D. Ferrell for editing the manuscript.

326 **References**

327 1. Lan Y, Rosen G, Hershberg R. Marker genes that are less conserved in their sequences
328 are useful for predicting genome-wide similarity levels between closely related prokaryotic
329 strains. Microbiome. 2016;4:18.
330 2. Larkin A a, Blinebry SK, Howes C, Lin Y, Loftus SE, Schmaus CA, et al. Niche
331 partitioning and biogeography of high light adapted Prochlorococcus across taxonomic
332 ranks in the North Pacific. ISME J. 2016;1–13.
333 3. Simister RL, Deines P, Botté ES, Webster NS, Taylor MW. Sponge-specific clusters
334 revisited: A comprehensive phylogeny of sponge-associated microorganisms. Environ.
335 Microbiol. 2012;14:517–24.
336 4. Wu Z, Yang L, Ren X, He G, Zhang J, Yang J, et al. Deciphering the bat virome catalog
337 to better understand the ecological diversity of bat viruses and the bat origin of emerging
338 infectious diseases. ISME J. 2016;10:609–20.
339 5. Müller AL, Kjeldsen KU, Rattei T, Pester M, Loy A. Phylogenetic and environmental
340 diversity of DsrAB-type dissimilatory (bi)sulfite reductases. ISME J. 2015;9:1152–65.
341 6. University W. Phylogeny Programs.
342 http://evolution.genetics.washington.edu/phylip/software.html#Plotting. Accessed 2016 Jul
343 21.
344 7. Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. EvolView, an online tool for visualizing,
345 annotating and managing phylogenetic trees. Nucleic Acids Res. 2012;40.
346 8. Rambaut A. FigTree. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 2016 Jul 21.
347 9. Zmasek CM. Archaeopteryx.

348 https://sites.google.com/site/cmzmasek/home/software/archaeopteryx. Accessed 2016 Jul
349 21.
350 10. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R
351 language. Bioinformatics. 2004;20:289–90.
352 11. Revell LJ. phytools: An R package for phylogenetic comparative biology (and other
353 things). Methods Ecol. Evol. 2012;3:217–23.
354 12. Huerta-Cepas J, Dopazo J, Gabaldón T. ETE: a python Environment for Tree
355 Exploration. BMC Bioinformatics. 2010;11:24.
356 13. Chen W-H, Lercher MJ, Ganfornina M, Gutierrez G, Bastiani M, Sanchez D, et al.
357 ColorTree: a batch customization tool for phylogenic trees. BMC Res. Notes. BioMed
358 Central; 2009;2:155.
359 14. McDonald D, Clemente JC, Kuczynski J, Rideout J, Stombaugh J, Wendel D, et al. The
360 Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love
361 the ome-ome. Gigascience. 2012;1:7.
362 15. Ripley BD. The R project for statistical computing. 2001. p. 1–3.
363 16. Kelly KL. Twenty-two colors of maximum contrast. Color Eng. 1965. p. 26–7.
364 17. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of
365 large phylogenies. Bioinformatics. 2014;30:1312–3.
366 18. Felsenstein J. PHYLIP. http://evolution.gs.washington.edu/phylip.html. Accessed 2016
367 Jul 21.
368 19. Suttle CA. Marine viruses — major players in the global ecosystem. Nat. Rev.
369 Microbiol. Nature Publishing Group; 2007;5:801–12.

370 20. Rohwer F, Thurber RV. Viruses manipulate the marine environment. Nature. Nature
371 Publishing Group; 2009;459:207–12.
372 21. Rohwer F, Edwards R. The Phage Proteomic Tree: a genome-based taxonomy for
373 phage. J. Bacteriol. 2002;184:4529–35.
374 22. Phage SEED. http://www.phantome.org/PhageSeed/Phage.cgi. Accessed 2016 Jul 21.
375 23. Wommack KE, Nasko DJ, Chopyk J, Sakowski EG. Counts and sequences,
376 observations that continue to change our understanding of viruses in nature. J. Microbiol.
377 2015;53:181–92.
378 24. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere
379 using metagenomics. PLoS Genet. Public Library of Science; 2013;9:e1003987.
380 25. Chopyk J, Moore RM, DiSpirito Z, Stromberg ZR, Lewis GL, Renter DG, et al. Presence
381 of pathogenic Escherichia coli is correlated with bacterial community diversity and
382 composition on pre-harvest cattle hides. Microbiome. BioMed Central; 2016;4:9.
383 26. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved
384 dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina
385 MiSeq platform. Microbiome. 2014;2:6.
386 27. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial
387 Communities. Appl. Environ. Microbiol. American Society for Microbiology;
388 2005;71:8228–35.
389 28. Sakowski EG, Munsell E V., Hyatt M, Kress W, Williamson SJ, Nasko DJ, et al.
390 Ribonucleotide reductases reveal novel viral diversity and predict biological and
391 ecological features of unknown marine viruses. Proc. Natl. Acad. Sci. National Academy

392 of Sciences; 2014;111:15786–91.

393 29. McAllister SM, Davis RE, McBeth JM, Tebo BM, Emerson D, Moyer CL. Biodiversity

394 and emerging biogeography of the neutrophilic iron-oxidizing Zetaproteobacteria. Appl.

395 Environ. Microbiol. American Society for Microbiology (ASM); 2011;77:5445–57.

396 30. Biodocker. http://biodocker.org/. Accessed 2016 Jul 21.

397 31. Merkel D. Docker: lightweight Linux containers for consistent development and

398 deployment. Linux J. Belltown Media; 2014;2014:2.

399 32. GNU Operating System. http://www.gnu.org/licenses/. Accessed 2016 Jul 21.

400

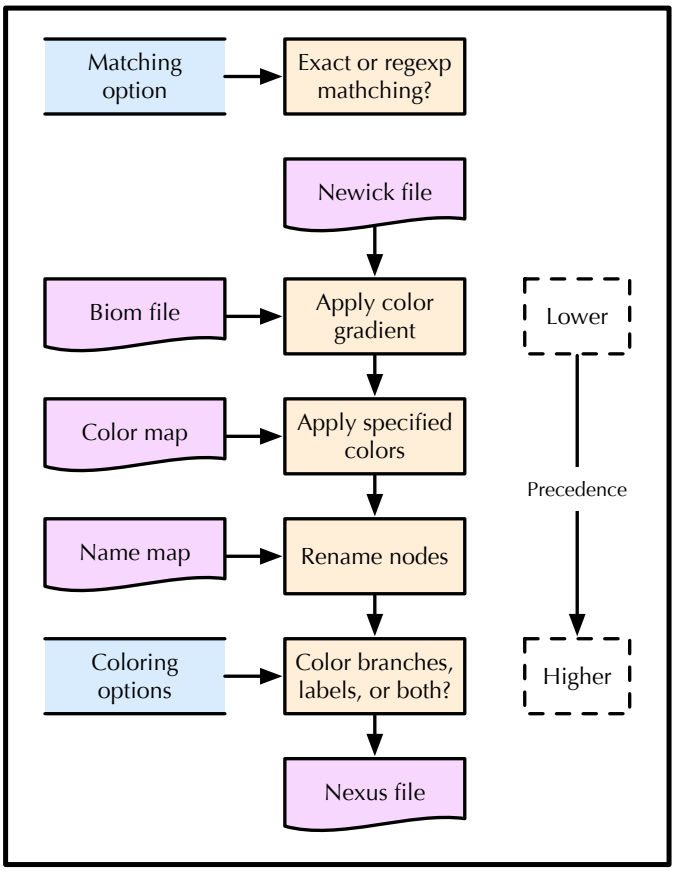

Fig. 1: Precedence of Iroki's customization pipeline

Flowchart illustrating the precedence of steps when performing multiple customizations with Iroki. Input/output files are purple, command line options are in blue, and processes are orange. The choice of exact or regular expression matching guides each subsequent step of the process. Iroki gives higher precedence to processes towards the bottom of the diagram. For example, given that a user selects the options for coloring both labels and branches, and provides both a biom file and color map with the color map specifying colors for the labels only, then the branches will be colored according to the color gradient inferred from the biom file, whereas the labels will be colored according to the rules specified in the color map.

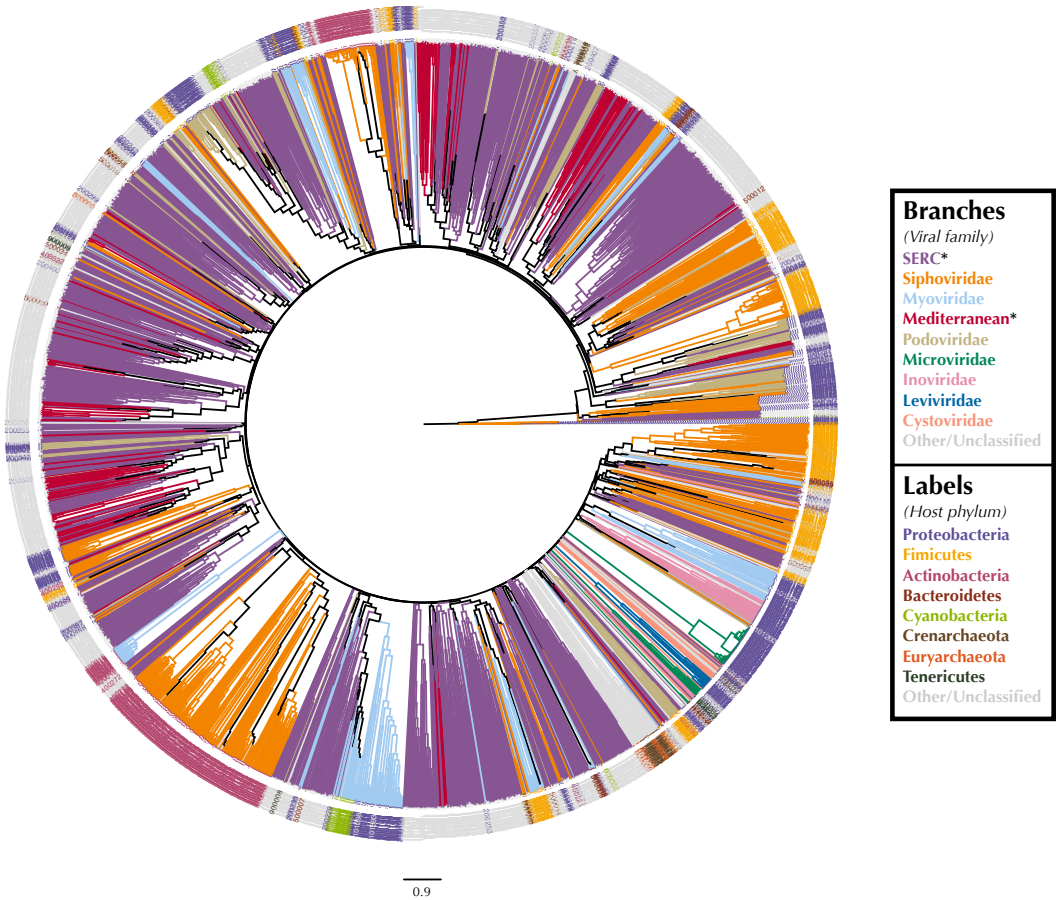

Fig. 2: Comparing phage and their host phyla

All phage genomes from Phage SEED with assembled virome contigs from the Chesapeake Bay and Mediterranean Sea. Iroki highlights phylogenetic trends after coloring branches according to viral family or sampling location in the case of virome contigs (marked with an asterisk in the legend), and coloring node labels according to host phylum of the phage.

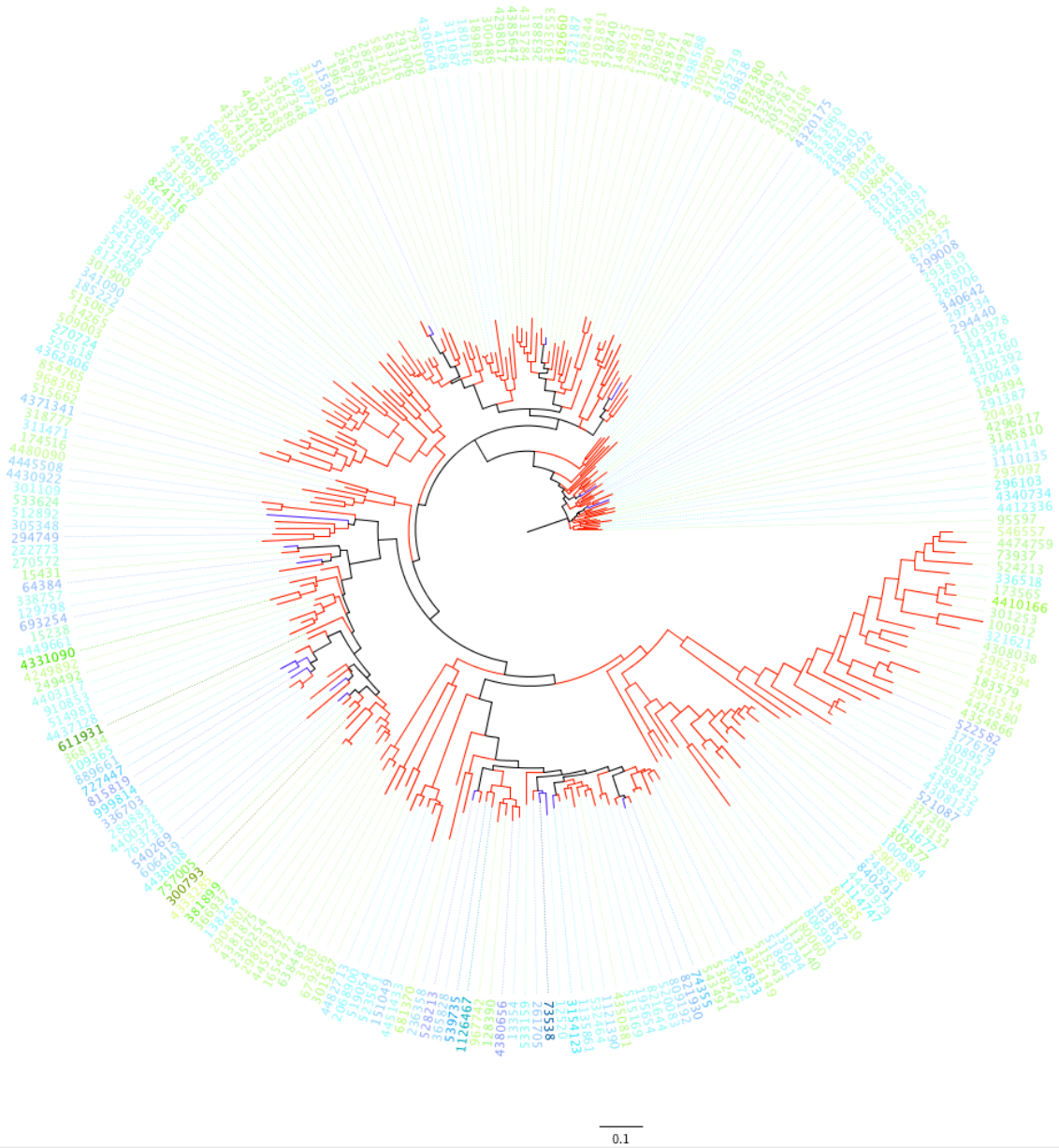

Fig. 3: Changes in OTU abundance in two sample groups

Approximate-maximum likelihood tree of OTUs that showed significant differences in relative abundance between STEC positive and STEC negative cattle hide samples. Branches show significance based on coloring by the p-value of a Mann-Whitney U test examining changes in abundance between samples positive for STEC ($p < 0.05$ – red) and samples negative for STEC, ($p >= 0.05$ – blue). Label color on a blue-green color gradient

428 highlights OTU occurrence based on the abundance ratio between STEC positive samples
429 (blue) and STEC negative samples (green). For example, labels that are darker green had a
430 higher abundance in STEC negative samples. Node luminosity represents overall
431 abundance with lighter nodes being less abundant than darker nodes.

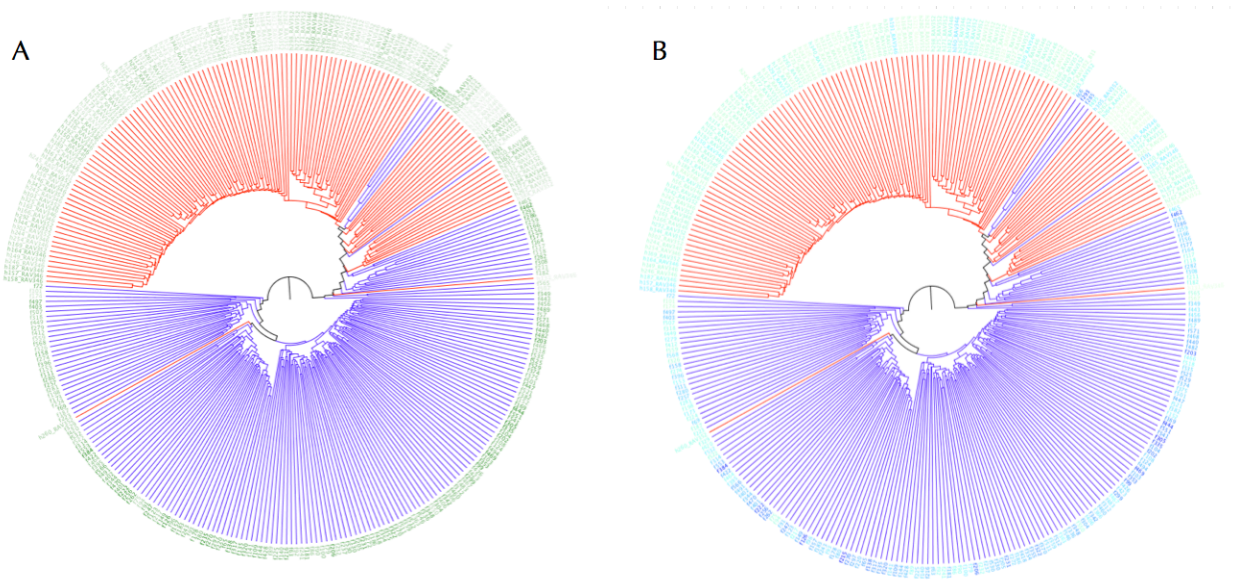

Fig. 4: Comparing cattle fecal and hide samples and the abundance of Ruminococcaceae

Phylogeny based on UPGMA tree of pairwise unweighted UniFrac distance between 356 bacterial community profiles based on SSU rRNA amplicon sequences from cattle hide and feces. Branches are colored by feces (blue) and hide (red). Rapid testing of the hypothesis that the abundance of one of the most abundant families, Ruminococcaceae, and sample origin are correlated is enabled through node label coloring by (A) a green single-color gradient (color saturation increases with increasing abundance of Ruminococcaceae OTUs) and (B) a light green (low abundance of Ruminococcaceae OTUs) to dark blue (high abundance of Ruminococcaceae OTUs) color gradient.

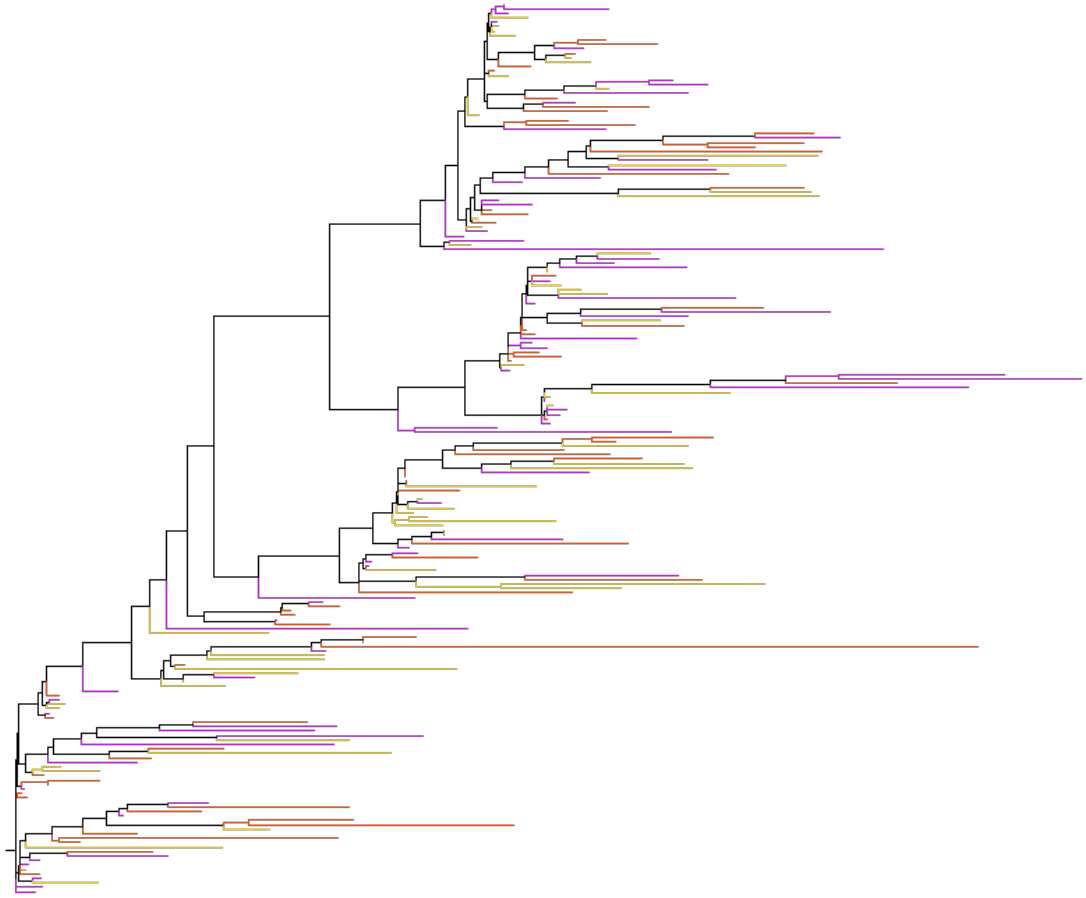

Fig. 5: Temporal dynamics of virioplankton populations according to Cyano II RNR amplicon phylogeny

An approximately-maximum-likelihood phylogenetic tree of 200 randomly selected class II Cyano RNR representative sequences from 98% percent clusters. Iroki was used to color branches by time point: zero hours – yellow, six hours – orange, and twelve hours – purple.

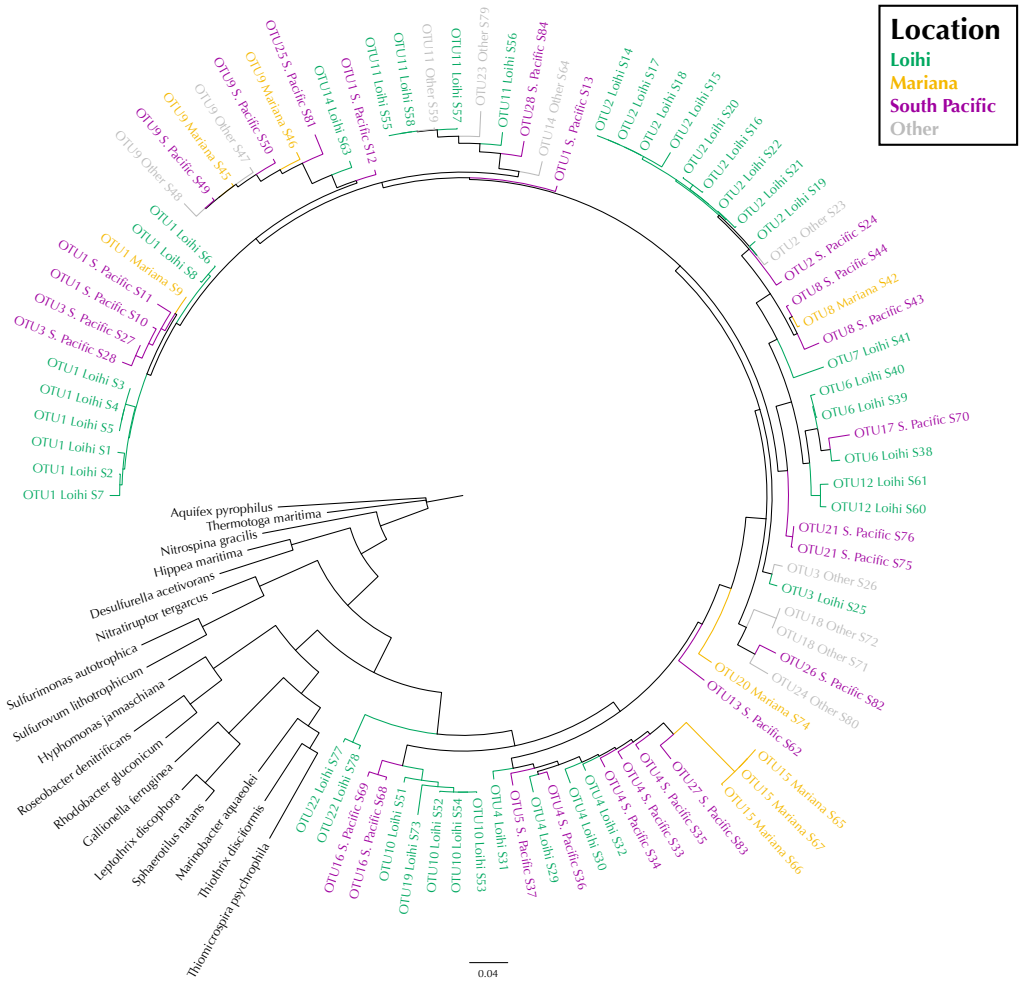

Fig. 6: Zetaproteobacteria show biogeographic partitioning

Phylogenetic tree showing placement of full-length Zetaproteobacteria SSU rRNA sequences with outgroups. Iroki was used to color labels and branches by geographic location of the sampling site (Loihi – green, Mariana – gold, South Pacific – purple, and Other – gray), as well as to rename the nodes with OTU and sampling site metadata.