Variation in Chemical Production among Cryptic Species of the Soft Coral *Sarcophyton glaucum*

Sydney Davis[1], Samantha Gromek[2], Marcy Balunas[2], and Katherine Maloney[1]
[1]Department of Chemistry, Point Loma Nazarene University, San Diego, CA
[2]Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT

Abstract:

Marine organisms commonly produce an abundance of natural products, some of which have been found to have important medicinal, cosmetic, and pharmaceutical applications. *Sarcophyton glaucum* is a species of soft coral known to produce a large number of natural products varying in type and amount. *S. glaucum* was recently determined to be a cryptic species complex of seven distinct clades, such that the large variety in the natural products might possibly be explained by these new subdivisions. In this study, we used liquid chromatography tandem mass spectrometry to convert the chemical inventory of various *S. glaucum* specimens into a molecular network. The results revealed the overall diversity of the *Sarcophyton* natural products, as well as a pattern of division in some of the chemical production between different species and clades. This technique can be further used to better evaluate its diversity and discover new natural products.

Introduction:

Background

Sarcophyton glaucum is a species of soft coral found in the Pacific Ocean region that synthesizes more than 90 natural products. Some of the secondary metabolites for this species of soft coral are shown to have exciting medicinal applications. One example is sarcophytol A, which is shown to inhibit tumor promotion.[1] Unfortunately, the re-isolation of specific metabolites from *S. glaucum* is difficult to accomplish. Variation in the types and amounts of

metabolites produced by individual soft corals within this species makes isolating a specific metabolite inconsistent. Originally, to identify if an individual soft coral even belonged to the *Sarcophyton* genus, morphological features like a mushroom shape and the appearance of tiny protrusions called polyps were used. To distinguish the classification of species, the size and shape of microscopic calcium carbonate spines called sclerites were compared and organized, allowing for the division of *Sarcophyton* into dozens of species. However, when DNA sequencing analysis emerged, the previously identified members of *S. glaucum* were found to be further divided into seven genetically distinct clades.[2] We hypothesize that the chemical variation being observed in *S. glaucum* and causing the inconsistent isolation of metabolites is due to the fact that it is a cryptic species divided into seven distinct clades.

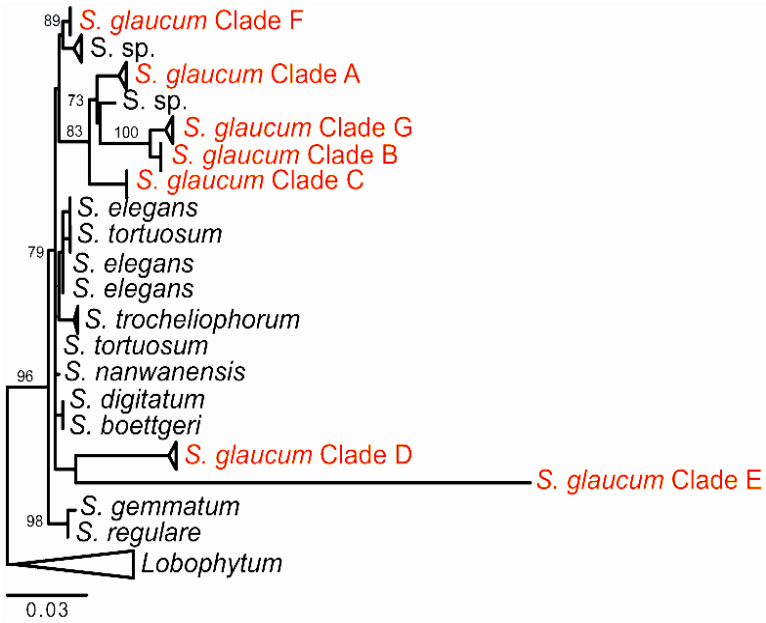

Figure 1: Maximum likelihood reconstruction of the phylogenetic relationships among species of *Sarcophyton* based on 735 base pairs of the mitochondrial *msh1* gene.[2] The specimens previously considered to be *S. glaucum* fall into 7 genetically distinct clades (red).

Molecular networking is a computational process that allows for the rapid comparison of various mass spectrometry profiles built from extracts of different *Sarcophyton* soft corals.

Molecular networking uses mass spectrometry ion fragmentation data to connect similarities between the natural products. These fragmentation patterns can be obtained using liquid chromatography tandem mass spectrometry, which first separates the compounds found in a sample using liquid chromatography and then fragments each of the compounds using tandem mass spectrometry. The fragmentation pattern for each compound is reproducible, a characteristic of mass spectrometry, where a given compound will fragment the same way every time. This allows for a mass spectrometry fragmentation pattern of a compound to act as a fingerprint for it, where that fragmentation pattern can then be compared to other fragmentation patters, or fingerprints, to look for similarities or differences. Each node of a molecular network represents a peak in the chromatogram corresponding to a compound and displays the corresponding parent mass. After fragmentation, compounds that result in the same fragmentation patterns are computed in GNPS as a cosine score of 1 to reflect the identical fragmentation spectra.[3] On the other hand, compounds resulting in completely different fragmentation patterns and completely different parent ions are computed in GNPS as a cosine score of 0.[3] Similar fragmentation patterns will result in a cosine score between 0 and 1 depending on how similar they are. The edges, or lines, connecting nodes in a molecular network are then constructed based on the computed cosine score. If there were no structural similarities between compounds, the cosine score was zero and they were not connected into the same molecular family. Compounds that are structurally similar are grouped together into molecular clusters.

Past Work

Specimen collection of *Sarcophyton glaucum* occurred off the coast of the Republic of Palau. Using SCUBA, over 200 specimens were collected while the depth and dive site location

for each of the specimens was recorded. An individual soft coral was first identified morphologically as *Sarcophyton* and later identified by clade as *Sarcophyton glaucum* by using DNA sequencing analysis of the *msh1* gene. Specimens belonging to *S. glaucum* clades D and F, as well as to the control group *S. trocheliophorum,* were extracted and fractionated.

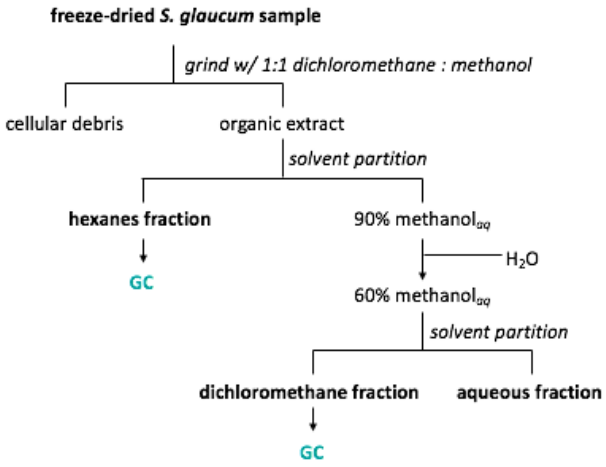

Figure 2: Extraction and fractionation of *Sarcophyton glaucum* specimens.

The hexane and dichloromethane fractions from *S. trocheliophorum* and the *S. glaucum* clades D and F were analyzed using gas chromatography. The resulting chromatograms contained promising patterns. For example, the *S. glaucum* clade D chromatograms looked similar across the samples and, while the *S. glaucum* clade F chromatograms had a few different variations in their patterns, each looked distinct from the *S. glaucum* clade D chromatograms. Visually a difference between the chemical profiles of the *S. glaucum* clades was obvious, but further research and analysis was needed to better understand the diversity and if differences in the chemical production was indeed observed between the different clades of *S. glaucum.*

Results and Discussion:

The molecular network for the liquid chromatography mass spectrometry profiles for all of the Sarcophyton samples from the DCM fraction contained 10,484 nodes, reflecting the general abundance of natural products being observed in just two of the cryptic species of *Sarcophyton glaucum*. The 10,484 nodes in the molecular network were clustered into 716 molecular families and 7,270 individual nodes. The molecular network contained natural product fragmentation patterns appearing either uniquely in *S. glaucum* clade D, *S. glaucum* clade F, *S. trocheliophroum* or as some combination of two or more of the *Sarcophyton* species. The color of the node in the molecular network represented the phylogenetic classification of the species that each respective fragmentation pattern stemmed from.

Table 1: Generated node colors for the molecular network of the natural products from *S. glaucum* clades D (*Sg* D) and F (*Sg* F) and from *S. trocheliophorum* (*St*).

Red	***Sg* D**
Yellow	***Sg* F**
Blue	***St***
Teal	***Sg* D, *Sg* F, and *St***
Orange	***Sg* D and *Sg* F**
Purple	***Sg* D and *St***
Green	***Sg* F and *St***

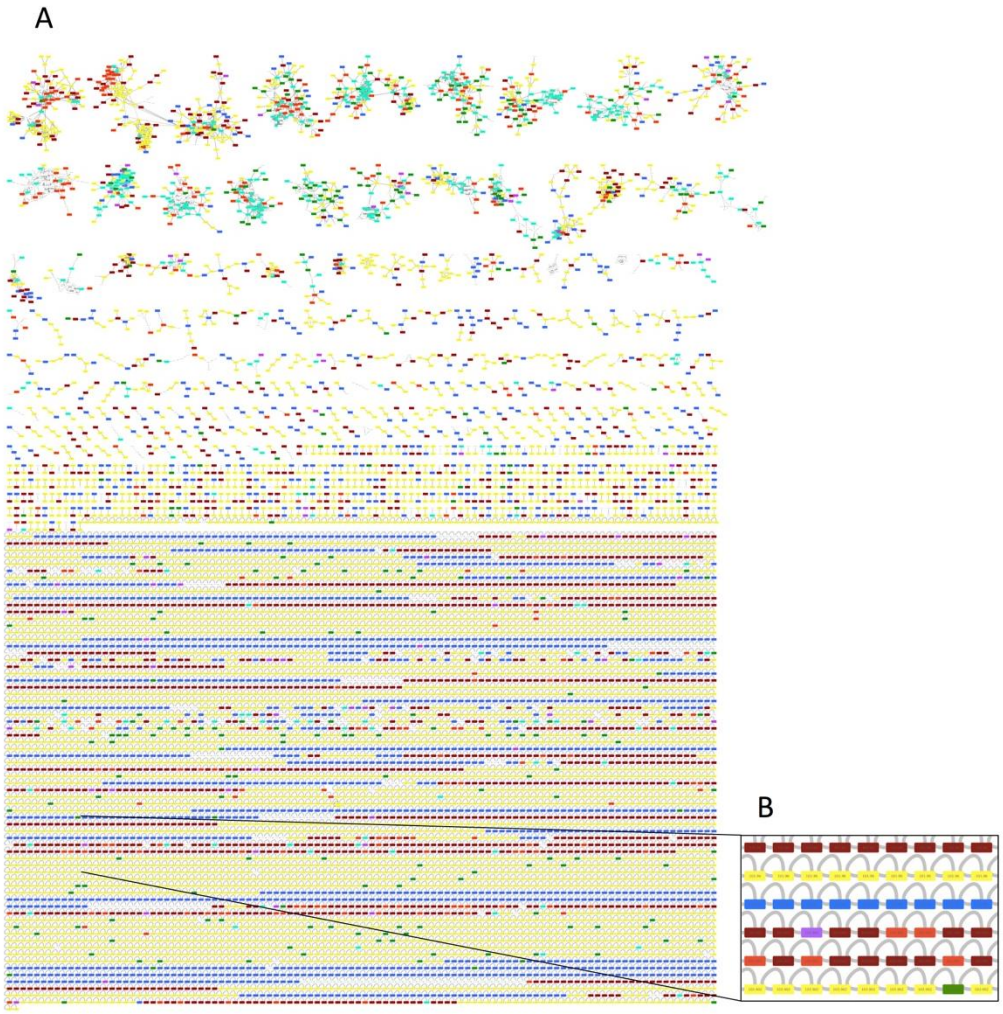

Figure 3: (**A**) Molecular network of natural products from *Sarcophyton glaucum* clades D and F and from *Sarcophyton trocheliophorum*. The color of the node represents the phylogenetic classification. (**B**) Enlarged self-loop clusters representing fragmentation patterns of compounds that loop only to themselves.

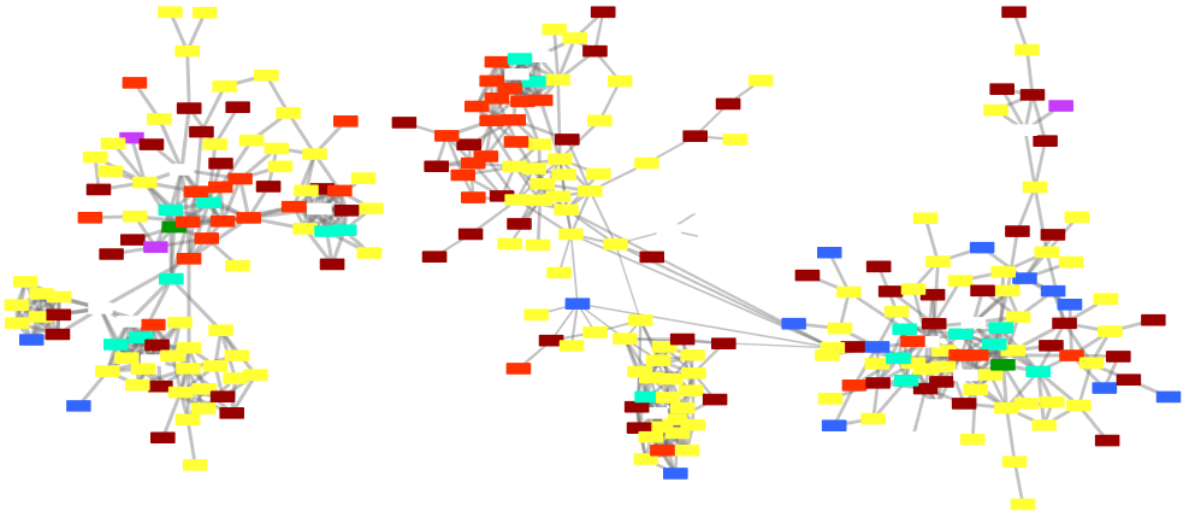

Figure 4: The largest molecular network clusters made from the natural products of various S*arcophtyon* soft corals with each having over 90 nodes.

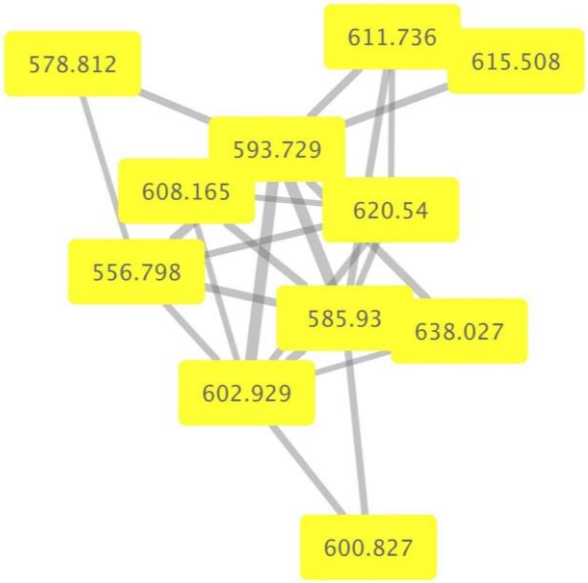

Figure 5: A smaller cluster of the molecular network, where each node is only from one type of soft coral, as is seen with the colors red, yellow, and blue. This trend is seen in most of the smaller or self-looping clusters.

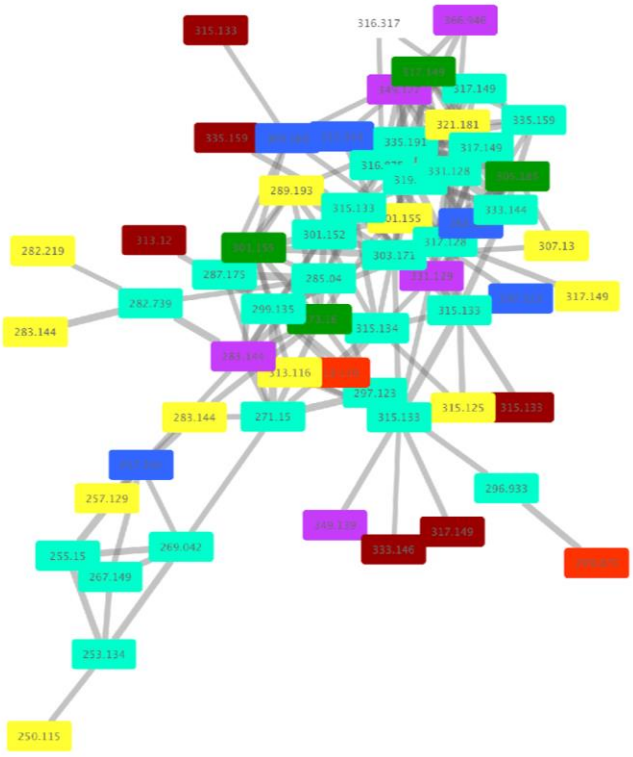

Figure 6: One of the larger molecular network clusters made from the natural products of various *Sarcophtyon* soft corals, where the majority of nodes appear to come from a combination of different soft corals, as seen with the colors teal, orange, and green. Several of the masses shown are consistent with cembranoids from *Sarcophyton*. This trend is seen in most of the larger clusters in the molecular network.

The majority of compounds found in all three of the different soft coral types (teal color) are commonly observed in the larger molecular network clusters (Figure 3). Compounds being produced only from *S. glaucum* clade F, represented by yellow colored nodes, seem to appear the most frequently in the molecular network, especially in the self-loops where unique fragmentation patterns are being observed in a compound (Figure 3). The self-looping

compounds appear to reflect the differences in natural product production being observed between *S. glaucum* clades D and F as well as between *S. trocheliophorum* and *S. glaucum*. Each of these self-looping compounds has a distinct fragmentation pattern causing it to connect only to itself and each is found in only one type of soft coral (red, green, or blue), as seen in the self-loops in Figure 3. These compounds represent the unique chemical production being found in each of the *Sarcophyton* soft corals used in this study. The complexity of the diversity can be seen throughout the molecular network, with clusters containing a variety of compounds and connections. Some compounds formed clusters with other compounds that were also only found in the same *Sarcophtyon* soft coral, like with the molecular network cluster of only *Sarcophyton glaucum* clade F (yellow) compounds seen in Figure 5. Other compounds formed clusters with compounds found in other individual *Sarcophyton* soft corals and with compounds found in multiple of the *Sarcophyton* soft corals, as seen in Figure 6 with the colors in the cluster representing the source. The cluster in Figure 5 represents an example of where compounds that are structurally similar to each other also stem from the same clade. This would suggest that some compounds are indeed specific to certain clades and reflect the chemical distinction between the natural products of different phylogenetic classifications. Figure 5 also shows the differences in edge width observed in the molecular network, where there were varying edge thicknesses between nodes depending on how similar the fragmentation patterns were, as edge width correlates to the cosine score for chemical similarity.

Experimental:

Liquid Chromatography Mass Spectrometry Profiles

The dichloromethane fractions from *S. trocheliophorum* and from the *S. glaucum* clades D and F were prepared for LC-MSMS using around 1mL of dichloromethane from Sigma Aldrich to dissolve each fraction. A small amount of the solution was then removed and placed into a new vial. The solution was completely dried using a rotovap and weighed to the nearest 0.0001g. More of the original solution was added to increase the weight of the sample or the sample was dissolved in 1mL of dichloromethane and a small amount was removed to decrease the weight of the sample. The sample was again dried using a rotovap and weighed to the nearest 0.0001g. This process was repeated until between 1 and 2 micrograms of each fraction was removed and placed into a new vial. In total, 54 of these samples were prepared and the vials were sent to the University of Connecticut. Extracts were analyzed via LC-MS/MS using a Xevo G2-XS QTof coupled to an Acquity UPLC system with an Acquity UPLC HSS T3 (2.1 x 50 mm, 1.8 μM) at a flow rate of 0.45 mL/min. Samples were prepared at 1 μg/mL in methanol and gradient elution was performed with mobile phases A (water with 0.1% formic acid) and B (acetonitrile with 0.1% formic acid). The gradient began with 5% mobile phase B, isocratic for 0.5 min; increased to 60% B over 3.5 min; increased to 98% B over 4 min; remained at 98% B for 1.2 min; then back to 5% B; isocratic for 1.8 min. Parameters for electrospray ionization (ESI) were a voltage of 3.0 kV, m/z range of 50-2000 was calibrated with sodium iodide. The capillary and sampling cone were set to 2 kV and 30 kV, respectively. The desolvation gas was used at a flow rate of 800 L/h and the source and desolvation temperature were set to 100 ^{o}C and 200 ^{o}C, respectively. Data was obtained using a data dependent acquisition (DDA) method with a positive MS survey scan, ranging from *m/z* 50-2000, followed by 5 data-dependent MS/MS scans of ions that had greater intensity than 1000 from the first scan event.

Molecular Networking

The chemical profiles generated via LC-MSMS were converted to mzML format using ProteoWizard package and processed in GNPS: Global Natural Products Social Molecular Networking. The following parameters were selected in GNPS for the soft coral network: minimum cosine similarity score of 0.65 with at least four matching peaks, parent mass tolerance of 1.0 Da, fragment ion tolerance 0.3 Da, and the filter peaks in a 50Da window was turned off. Opening Cytoscape 3.5.1, select start new network with empty network and name network. Then go to file →import→network→file and select the folder netowrkedges_selfloop to generate a network containing the clusters of both molecular families and the self-looping compounds. When a window pops up about Importing Network From Table, use the arrows pointing left next to the names to select the specific symbols shown in images 1 and 2 of the Appendix. The green circle denotes selecting that data as the source, the orange circle denotes selecting that data as your target, and the black slash means that it is not included. Next go to Advanced Options and make sure that TAB has a check mark and Use first line as column names has a check mark. All that is portrayed now is how the various masses are arranged in the network, so the parameters for the clades also need to be uploaded. Go to File→import→table→file and select the folder clusterinfosummarygroup_attributes_withIDs. Edit the columns again using the arrows in the window Import Columns From Table by following image 3 in the Appendix to change the data type to ab (or string) for all of the columns (not just for cluster index). Then go to Advanced Options and make sure that again TAB and Use first line as column names are selected and then click OK. Now the clades are imported. After uploading with representative parameters, the molecular network layout was adjusted to clearly portray the diversity of the natural products found in *Sarcophyton* as well as the differences between clades and species. In

the control panel of Cytoscape, going to Style is where changes can be made to the colors, shapes, sizes, etc. of the nodes. For example, to change the color of different nodes select the drop down for Fill Color and select Default Groups for Columns and Discrete Mapping for Mapping Type. Selecting in the white box next to the different combinations of groups allows you to select a color for each type of node. Here it is important to note that G1 is clade D, G2 is clade F, and G3 is *S. trocheliophorum*. The edge layout was adjusted so that edges with larger cosine scores had wider lines connecting the nodes and then the edges with smaller cosine scores had thinner lines connecting the nodes. If there were no structural similarities between compounds, the cosine score was zero and they were not connected into the same molecular family. This can be done in the control panel under style by changing to edge rather than node at the bottom. Then select the dropdown arrow by the width category and select Cosine for column and Continuous Mapping for Mapping type.

Conclusion:

The development of molecular networks from a large array of mass spectrometry samples has proven to be an efficient method in exploring the diversity of thousands of natural products. The results accentuated the importance for chemical organization within *Sarcophyton glaucum* and between the seven distinct clades. This technique can be further explored in the search for novel compounds derived from *Sarcophyton glaucum* and can be fine-tuned to create molecular networks with different chemical relationships. Furthermore, the molecular network can be used for dereplication of the MS-MS spectrum against the GNPS's spectral libraries for further identification of present ions.

Acknowledgements:

This project was supported by the Point Loma Nazarene University Chemistry Department. Thank you Dr. Marcy Balunas and Samantha Gromek for obtaining the soft coral LC-MSMS data. Special thanks to Dr. Katherine Maloney for guiding and supporting this project and for giving me the tools to pursue a career in both research and medicine.

Appendix:

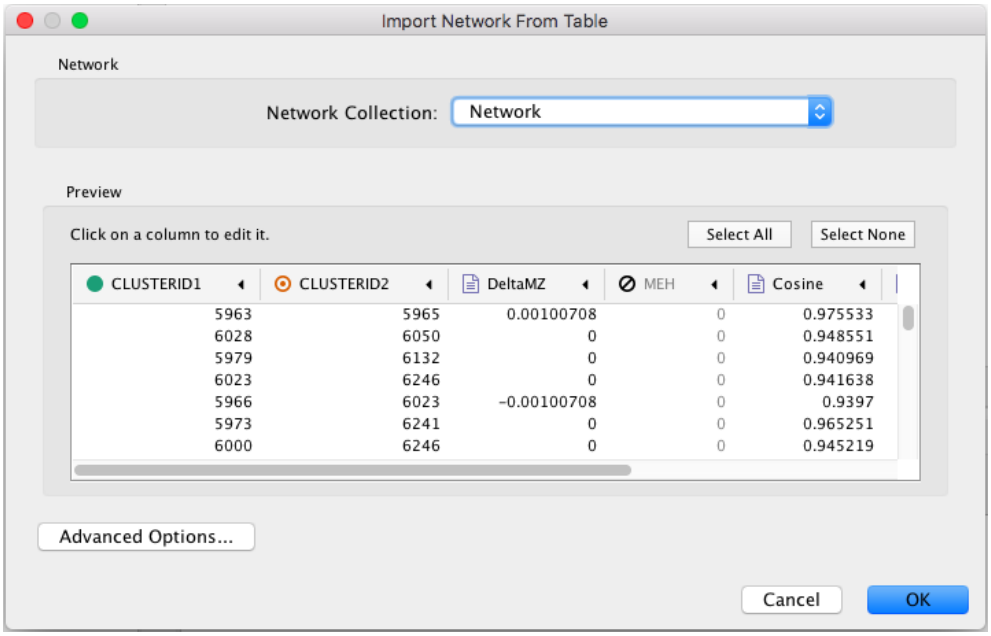

Image 1: Parameters for uploading *Sarcophtyon* data into Cytoscape.

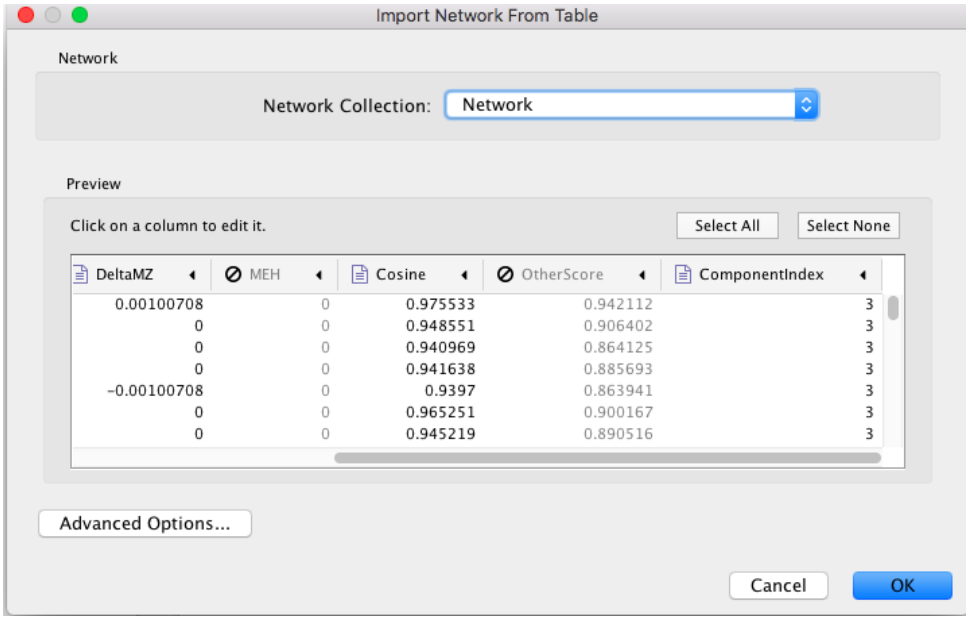

Image 2: More parameters for uploading *Sarcophtyon* data into Cytoscape.

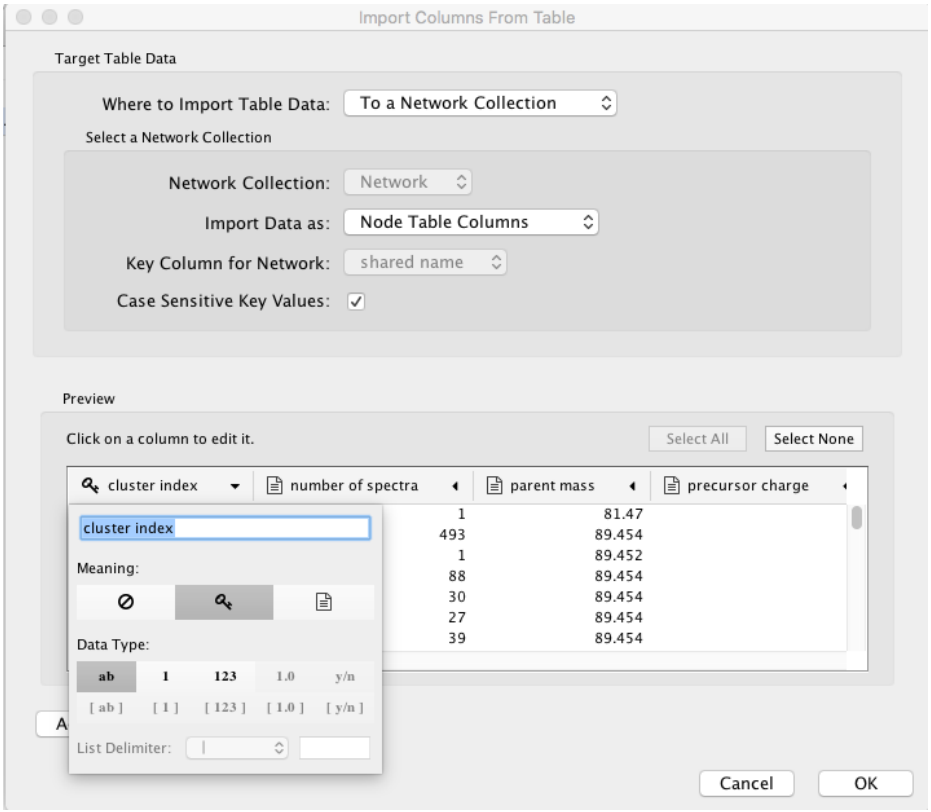

Image 3: More parameters for uploading *Sarcophtyon* data into Cytoscape.

References:

1) H. Fujiki and T. Sugimura, Cancer Surveys, 2, 540 (1983).

2) Mcfadden, Catherine S., et al. “Phylogenetic Relationships within the Tropical Soft Coral Genera Sarcophyton and Lobophytum (Anthozoa, Octocorallia).” *Invertebrate Biology*, vol. 125, no. 4, 2006, pp. 288–305., doi:10.1111/j.1744-7410.2006.00070.x.

3) Purves, Kevin, et al. “Using Molecular Networking for Microbial Secondary Metabolite Bioprospecting.” *Metabolites*, vol. 6, no. 1, 8 Jan. 2016, p. 2., doi:10.3390/metabo6010002.