*Preliminary Summary of Methods and Results for the Taxonomic and Phylogenetic Analysis of University Greenhouses*

Caroline Dolt, Hilary Callahan

A. Receiving and Resolving Original Lists

Emails were sent out to various department faculty and staff, usually the greenhouse manager or a biology professor with botanical research. We received many replies and a total of 30 lists from institutions across the nation. Many of these lists contained old names or best guesses, so it was necessary to go through each of them by hand. The first task was to resolve the names of each plant on the individual lists. This was done going through them by hand and using Google, The Plant List, IPNI, GBIF, and ITIS to determine the most updated names for the specimen. Many hybrids, predominately orchids and succulents, were eliminated from the lists since the programs used to analyze the lists deal best with species binomials. The total came to over 16,000 specimens with 7,366 unique species. Since there is a total of over 350,000 plant species total, this is a small slice of the phylogenetic pie. All the lists were cut into their binomials so they could be used in the RStudio package, taxize, which removed some information concerning subspecies or variation.

B. Finding Higher Level Taxonomic Classifications for all Specimens on Resolved Lists

To get the higher level classification of each specimen, lists were put into RStudio and run through the taxize package.

Each binomial list was put through this until there were a total of 30 profiles, going from the species up to the kingdom level classification. Profiles were corrected manually and new lists with orders, families, and genera were cut to analyze individually. There was both a total and a unique list made for each taxonomic level.

C. Taxonomic analysis using VEGAN with total numbers of orders, families and genera

Using the package, VEGAN, the cut lists made from the individual profiles could be bound and analyzed for similarities and differences.

The orders were all checked against the APG4 list released in 2016, which combined many existing orders. The lists included mostly angiosperms, but mosses and ferns are kept in some greenhouses.

After all the lists were brought in, they were bound to create one long list that had columns classifying the college, listing the order, and listing the frequency. There is a total of 87 orders across all specimens. Then, this bound list was put through ‘matrify’ command, which created a large frequency matrix.

From the frequency matrix, a jaccard index was made, which assigns a number between 0 and 1 to a pairwise comparison between two collections. A heatmap was from the jaccard matrix to visualize how similar and different collections are based on order frequency per collection (Fig 1).

The same process was followed for both the family and genus analyses. There are a total of 326 families and 2091 genera. A frequency and jaccard matrix were generated for both, and these are the resulting heatmaps(Fig 2, Fig 3).

The complete visualization of this data was put into a bar graph to show the complete tally of all orders, families, and genera per each collection(Fig 4).

D. Taxonomic analysis using VEGAN for unique numbers of orders, families and genera

The original cut lists were duplicated and then the duplicates within the list were deleted so that only a single occurrence for each taxon remained. These were loaded into RStudio so that the same process as above would be followed but studying only the presence/absence of a taxon would make some analysis easier.

It was found that there are 8 orders that occur in every greenhouse and 8 orders that occurred only in one greenhouse.

|  |  |
| --- | --- |
| Universal Orders | Endemic Orders |
| Alismatales | Berberidopsidales(UConn) |
| Asparagales | Blasiales(UConn) |
| Caryophyllales | Bruniales(UW Seattle) |
| Gentianales | Hymenophyllales(UW Seattle) |
| Lamiales | Isobryales(UW Seattle) |
| Malpigphiales | Notothyladales(UConn) |
| Poales | Pallaviciniales(UConn) |
| Polypodiales | Polytrichales(UWis Madison) |

For the families, there were 63 that occurred in only one greenhouse and 8 that were in all 28 greenhouses.

Universal(n=8): Amaryllidaceae, Apocynaceae, Araceae, Asparagaceae, Bromeliaceae, Euphorbiaceae

Endemic(n=62):

|  |  |
| --- | --- |
| University of Alaska - Fairbanks | Hydrophyllaceae |
| Southern Illinois University | Dipsacaceae |
| Cornell University | Calophyllaceae |
| Rollins College | Platanaceae, Styracaceae |
| University of Washington - Seattle | Bruniaceae, Fontinalaceae, Hymenophyllaceae |
| Smith College | Asteliaceae, Chrysobalanaceae, Kewaceae, Linnaeaceae, Philesiaceae, Stemonaceae |
| University of Wisconsin – Madison | Alzateaceae, Amblystegiaceae, Bryaceae, Hylocomiaceae, Jungermanniaceae, Polytrichaceae, Putranjivaceae |
| University of Connecticut | Berberidopsidaceae, Blasiaceae, Brachytheciaceae, Butomaceae, Byblidaceae, Cabombaceae, Cephaloziaceae, Clethraceae, Coriariaceae, Corynocarpaceae, Corynocarpaceae, Curtisiaceae, Daphniphyllaceae, Datiscaceae, Diervillaceae, Drosophyllaceae, Griseliniaceae, Grubbiaceae, Hernandiaceae, Hypnaceae, Juncaginaceae, Kirkiaceae, Lanariaceae, Lepidoziaceae, Loganiaceae, Martyniaceae, Montiniaceae, Myrothamnaceae, Notothyladaceae, Pallaviciniaceae, Penaeaceae, Philydraceae, Racopilaceae, Resedaceae, Salvadoraceae, Schoepfiaceae, Staphyleaceae, Stilbaceae, Surianaceae, Tamaricaceae, Trichocoleaceae |

Concerning genera, there are 820 genera that endemic to one greenhouse. No genera occur in all greenhouses, but two appear in 26 of the 28 greenhouses(*Euphorbia* and *Trandescantia*). See Figure 5.

There are no endemic taxa at any level in the collections at Williams College, Wyoming University, the University of Iowa, or Wheaton College. The University of Connecticut has the most endemic taxa at every level.

E. Phylogenetic analysis using picante to determine species richness and phylogenetic diversity statistics

A phylogenetic tree with branch lengths was generated for each collection using a list of species binomials placed into phylomatic(<http://phylodiversity.net/phylomatic/>). Centre College and the University of Connecticut were analyzed only at the genus level. These newick codes were uploaded into RStudio to be analyzed using the ‘picante’ package. The purpose of this package is to analyze the phylogenetic diversity of a plant list and do comparisons between lists. Four statistics were calculated at the order and family level: phylogenetic diversity (PD), species richness (SR), mean nearest taxon distance (MNTD), and mean pairwise distance (MPD).

Phylogenetic diversity is based on the system from Faith 1992, which assigns the number of tree units between taxa. Species richness is the number of unique taxa. Mean nearest taxon distance is the mean distance between each species and its closest relative within a community. The mean pairwise distance is the mean phylogenetic distance between all pairs of species within a community.

i. Results of the Order Phylogenetic Analysis(Figure 9)

The PD ranges from 12,994.01(UConn) to 2,958.12(Centre) with a mean of 6,500.81. The SR ranges from 74(UConn) to 16(ECU) with a mean of 39.17. Many universities shift one to three steps when ranked by highest species richness to highest phylogenetic diversity, but there are no drastic changes.

For MNTD and MPD, a positive z value and high p value(p>0.05) indicate greater phylogenetic distance than expected, or overdispersion. A negative z value and a low p value(p<0.05) indicate clustering and low phylogenetic distance.

For MNTD, 24 collections had a negative z and a low p value, showing significant overdispersion. Two collections, ECU and the University of Connecticut, had positive z values and high p values showing clustering. Four collections, Penn State, Richmond, U Seattle and UWis Madison, had contradictory results. All four had negative z values but high p values, so conclusive results cannot be drawn from these statistical results alone.

For MPD, the results were fairly similar except UWis Madison showed evidence of clustering. More of the collection had contradictory results with this statistic, increasing from 4 to 11, all with negative z values and high p values.

The shortest phylogenetic distance is 211.35 myr between Ericales and Cornales. The longest phylogenetic distances are over 880.58 myr between all species and these outgroups: Blasiales, Lunulariales, and Marchantiales. This makes sense, as these three orders are liverworts and would be much older than the angiosperms that are the majority of species kept in the greenhouses.

ii. Results of the Family Phylogenetic Analysis (Figure 10)

The PD ranges from 26,899.37(UConn) to 3,581.24(ECU) with a mean of 10,392.61. The SR ranges from 255(UConn) to 19(ECU) with a mean of 89.53. Many collections(n=12) rank in the same place for both SR and PD, but the remaining collections change from one(n=14) to four steps(n=1).

Based on MNTD, 4 collections show clustering, 12 show overdispersion, and the remaining 14 have inconclusive results.

For MPD, 19 of the collections show conclusive clustering. 11 collections have contradictory results.

The shortest phylogenetic distance is 112.39 myr between Asteraceae and Goodeniaceae. The longest phylogenetic distances are over 880.58 myr between all species and these outgroups: Blasiaceae, Lunulariaceae, Ricciaceae, and Conocephalaceae. Ricciaceae and Conocephalaceae are both families of liverworts belonging to the order Marchantiales.