Journal of
Emerging Investigators

A phylogenetic study of conifers describes their evolutionary relationships and reveals potential explanations for current distribution patterns

David Seong and Dr. Susan Offner

Lexington High School, Lexington, MA

Summary

A phylogenetic tree is a diagram showing the evolutionary relationships between organisms. Gymnosperms produce cones to house their seeds and are commonly assumed to have originated during the Carboniferous era and conifers are a major division of gymnosperms. This study used the matK gene to construct a phylogenetic tree of conifer species (junipers, sequoias, Cupressacae, spruces, pines) with an outgroup of cycads. The *matK* gene was aligned using the program Seaview and the tree was created using PAUP. Analysis revealed that pines, spruces, sequoias, and Cupressacae have similar phenotypes while junipers and cycads appear to be distinct. The Bootstrap values of the tree show that the overarching clades are strongly supported and, therefore, are reliable. The purpose of this study was to generate hypotheses regarding the evolution of certain traits as well as the current distribution patterns of these species. Firstly, leaf type variability seems to have been determined by environmental selection during Pangea. Secondly, the winged seed phenotype dates back to the common ancestor of all conifers and is accompanied by a loss of winged seeds in the juniper clade. Finally, the distributions of junipers and pines have both been affected by other organisms including humans. The phylogenetic tree constructed in this study lends support to the hypothesis that North American pines originated from Asian pines. This phylogenetic study of conifers provides additional information regarding the evolutionary history of conifers. The hypotheses generated as a result of this study also provide additional ideas regarding the relative status of conifers in the history of evolution and new directions for future field research.

Received: Oct 1, 2012; **Accepted:** Aug 28, 2013; **Published:** Oct 2, 2013

Introduction

A phylogenetic tree is a diagram that shows the evolutionary relationships between different organisms based on their genetic information. These trees are used to map the evolution of traits, to show the evolutionary relationships between organisms, and to explain the origin of certain species. To construct a phylogenetic tree, genetic sequences of a gene common to all the species must be aligned and then analyzed by a program. The accuracy of the tree can be verified using Bootstrap values generated by the program.

A phylogenetic tree of conifers was generated in this study because conifers are incredibly well-adapted organisms that can and have survived under a variety of conditions. Conifers are the biggest group of gymnosperms, non-flowering plants with seeds protected in cones (6). A phylogenetic tree of conifers allows the mapping of certain phenotypes that make conifers so successful. Using a phylogenetic tree, inferences such as when the species arose, how they arose, and how they are individually affected by their environment can be made. The tree, along with distribution maps, provides possible explanations for current distribution patterns as well. However, an exact tree showing the phylogeny of these particular conifer species of interest has not been previously generated. The gymnosperms selected for this study were the following: **spruce**: *Pinaceae Piceoideae Picea* (family, subfamily, genus), **junipers**: *Cupressaceae Juniperus*, **pine**: *Pinaceae Pinoideae Pinus* (family, subfamily, genus), **sequoias**: *Cupressaceae Sequoioideae Sequoia* (family, subfamily, genus), **conifers** from the *Cupressaceae* family, and **cycads** (outgroup): *Cycadaceae Cycas*. The cycads served as my outgroup in this study. An outgroup is a group of organisms that are closely related to the organisms of interest, but not in the same species. Outgroups are critical for rooting trees and allows identification of the loss or gain of certain traits. To construct a tree of these organisms, I used the *matK* gene. The *matK* (maturase K) gene is found in the chloroplast of plants and codes for a maturase protein that is involved in splicing introns (8). *matK* genes are highly conserved across plants and are commonly used for professional phylogenetic studies (11). The results of this project may produce clues about the traits and evolutionary history of a group of enormously old and successful plants. This study has generated numerous hypotheses about the evolution of certain phenotypes and the effect of the environment on the distribution and selection of these species.

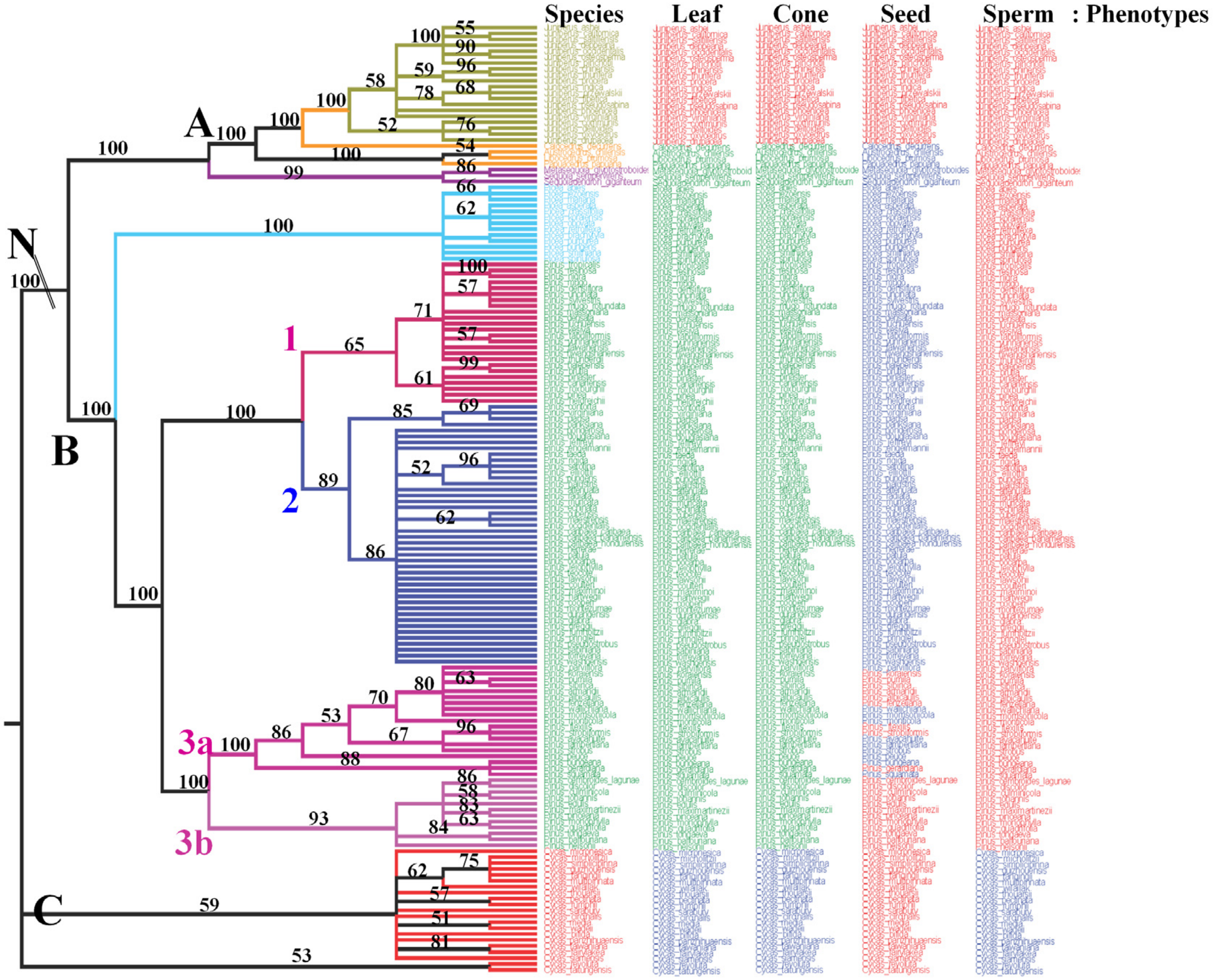

Figure 1: Bootstrap Tree. A phylogenetic tree of gymnosperms with Bootstrap values and multiple phenotypes coded by color. Bootstrap values indicate the accuracy of clades. Multiple phenotypes have been mapped for each species (indicated by color coding). Pines are divided into three clades labeled 1-3. Letters A, B, and C represent the three overarching clades of junipers/others, spruces/pines, and cycads respectively. The trait for needle-like leaves is at least as old as the clade marked "N". Species: Junipers (light green), Others (orange), Sequoias (purple), Spruces (light blue), Pines (green), and Cycads (red). Leaf: Needles/Scales (red), Needles (green), and Palm Fronds (blue). Cone: Berries (red), Cones (green), and "Leaf" Cones (blue). Seed: Wingless (red) and Winged (blue). Sperm: Nonmotile (red) and Motile (blue).

Results

Overview of Phylogenetic Tree of gymnosperms

Figure 1 is the Bootstrap tree constructed by using the *matK* gene and parsimony in PAUP. Parsimony is a statistical method of generating phylogenetic trees. Parsimony involves scoring all possible combinations of a tree that can be produced with the given species. Scoring is based on the amount of evolutionary changes required given the evolutionary relationship a tree depicts. Parsimony tries to select the tree that minimizes the necessary amount evolutionary changes in the given gene sequences because it is thought to be the most likely. The tree with the most favorable score is the resulting tree. This tree shows the evolutionary relationships between organisms and can be largely divided into 3 clades: A, B, C. The numbers shown are Bootstrap values, which signify how well the nodes are supported (100=very well supported). Consistent with the modern classification of species, junipers, cedars, sequoias, spruces, pines, and cycads are individually grouped with all other members of their species in strongly supported clades. The pine clade, spruce clade, and clade A have bootstrap values of 100 indicating that the evolutionary relationships between those clades indicated by this phylogenetic tree are very well supported. In general, the bootstrap values remain above 80 until the clades separating individual species. Pine clades can be subdivided into 3 smaller clades labeled 1,2,3a, and 3b. Interestingly, the phenotypes map cleanly and consistently by species. **Figure 2**, a phylogram, explains the indeterminate section of clade 2. A phylogram shows the number of changes in the base pairs of the gene between species.

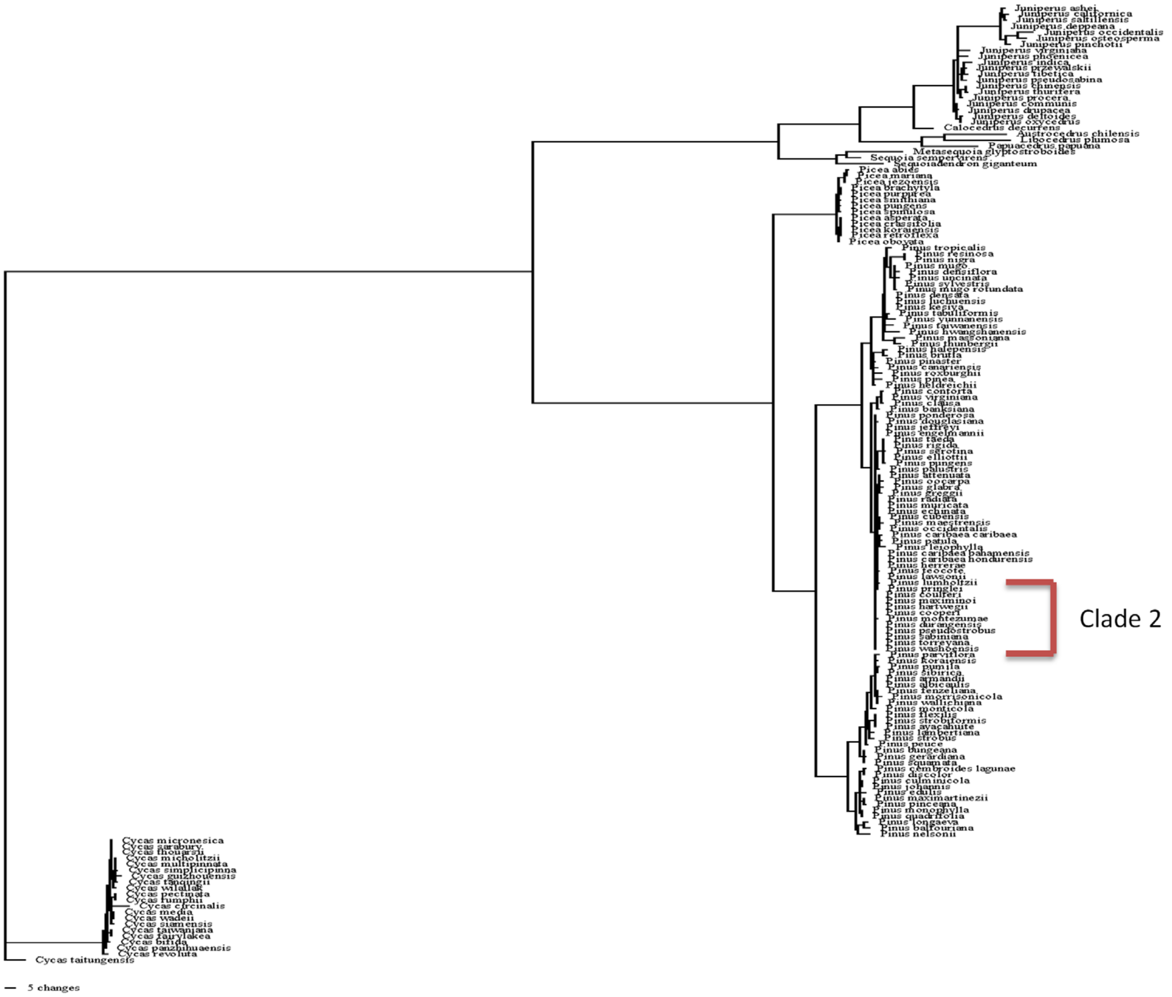

Figure 2: Phylogram. A phylogram shows the number of changes in DNA sequence between species by the length of lines connecting the species. This particular phylogram is a phylogram of the entire tree as shown in **Figure 1**. The almost non-existent lines in a part of clade 2 reveals that there is not enough difference in the genetic code of the matK gene to accurately determine evolutionary relationships between those species which corresponds to the non-conclusive evolutionary relationships of the bottom half of clade 2 in **Figure 1**.

It shows almost no differences in pine clade 2 which is why the clade is largely indeterminate in **Figure 1**.

Evolution of Traits Including Leaf Type, Seeds, and Cones

Although it is difficult to infer whether the common ancestor to conifers and cycads had needles or not due to the lack of additional outgroups, **Figure 1** reveals that the trait (N) coding for needle-like leaves is at least as old as the common ancestor to all conifers since all of the species except cycads have needle-like leaves. This analysis provides a hypothesis regarding when the needle-like leaves first evolved. **Figure 3** shows that cycads generally inhabit warmer equatorial regions. In contrast, conifers generally inhabit a wide variety of areas ranging from taiga to equatorial biomes suggesting an interesting relationship between leaves and habitat. **Figure 4** maps the phenotypes of winged and wingless seeds as well as the shape of the cones. Another outgroup species, ginkgos, not shown on this tree, have wingless seeds. It is evident that the common ancestor of at least all of the conifers in this study had winged seeds (WS). However since cycads and ginkgos are wingless, it is possible that seeds were wingless (Wls) before these species and winged species evolved separately in the conifer clade. The junipers mutated back into wingless seeds since they branched out from cedars, which branched out from sequoias, both of which have winged seeds. It is likely that junipers evolved independently of conifers since both their cones and seeds appear to have mutated just before the common ancestor of all Junipers (JC and JS). Additional phylogenetic studies with a larger numbers of species will be necessary to validate the hypotheses generated from this data.

Distribution Patterns of Junipers and Pines

Junipers are widely distributed, making them difficult

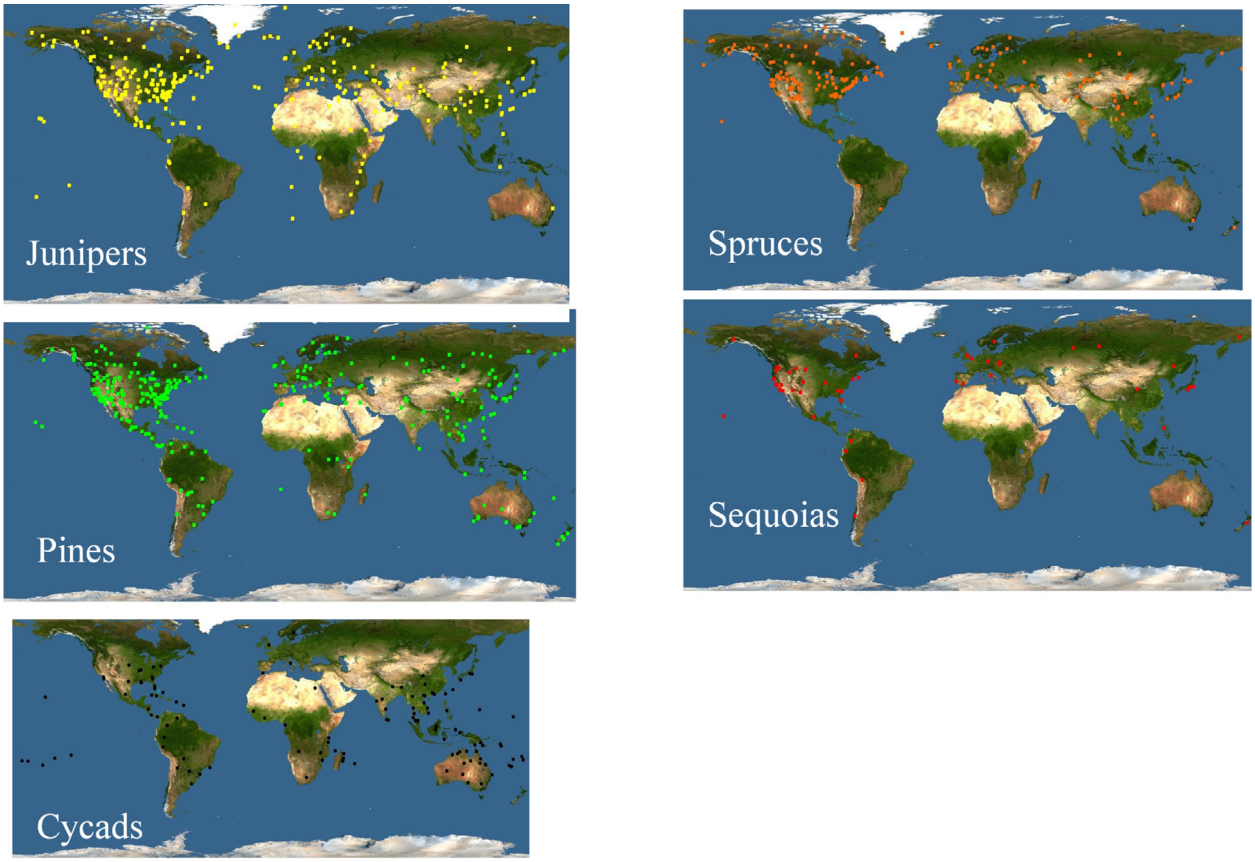

Figure 3 : Distribution map of species. The 5 maps shows current global distribution of all 5 major groups of gymnosperms studied. Note that cycads generally inhabit warmer regions while conifers inhabit a greater variety of climates.

to group them in clusters. Interestingly, **Figure 5** reveals there are series of juniper populations following the Swahili Coast (East Coast) of Africa. There are many small juniper populations on islands, which may have been used by ships on oceanic trade routes as well. This global distribution pattern is especially significant given that junipers have wingless seeds (**Figure 4**). Therefore, seed dispersal is not necessarily related to wind dispersal. A hypothesis that links the distribution of junipers to other animals can be generated given their proximity to trade locations and the overall widespread dispersal across the world despite their wingless seeds. Juniper distribution may have been influenced largely by humans, as well as birds, which could explain their widespread dispersal and small cluster patterns. This hypothesis provides a base for field studies that may potentially illuminate the distribution of junipers as well as other organisms.

Pine distribution is slightly easier to group into patterns. **Figure 6** shows concentrated pine populations in Eastern North America, Europe, and Asia. A combined analysis of **Figure 7** and **Figure 8** suggests hypotheses regarding pine distribution. **Figure 7** shows discrepancies in clade 3 between pines with winged and wingless seeds. Every species in clade 3b of **Figure 7** has wingless seeds which reveal that the wingless seed trait is at least as old as that clade (PWL). **Figure 8** shows world distribution by clade of pines from **Figure 1**. Despite the differences in seed type, **Figure 8A** reveals that the clade 3 (wingless) map is approximately the same as the maps of clade 1 & 2 (winged), showing large clusters of pines in Eastern Asia and the two coasts of North America. This suggests the influence of external factors similar to that of junipers. **Figure 8A** also shows a phylogenetic tree color coded by geographic location. Analysis of this tree suggests a hypothesis regarding the method of dispersal that allowed the wingless seeds to be spread globally as well as the origin of some species of pines. Pines may have spread during the breakup of Pangea, the prehistoric single continental landmass, which could account for the clusters in Europe and Eastern North America (**Figure 6**). In addition, pines may have dispersed from Asia across the Bering Strait into North America presumably with the help of other organisms which accounts for the clusters in Asia and western North America. The location tree (**Figure 8**) supports the movement of pines from Asia to North America since the ancestral species (G) of clade 3b, which include some North American pines, are all Asian (color coded in red). **Figure 8B** shows the worldwide distribution of pines

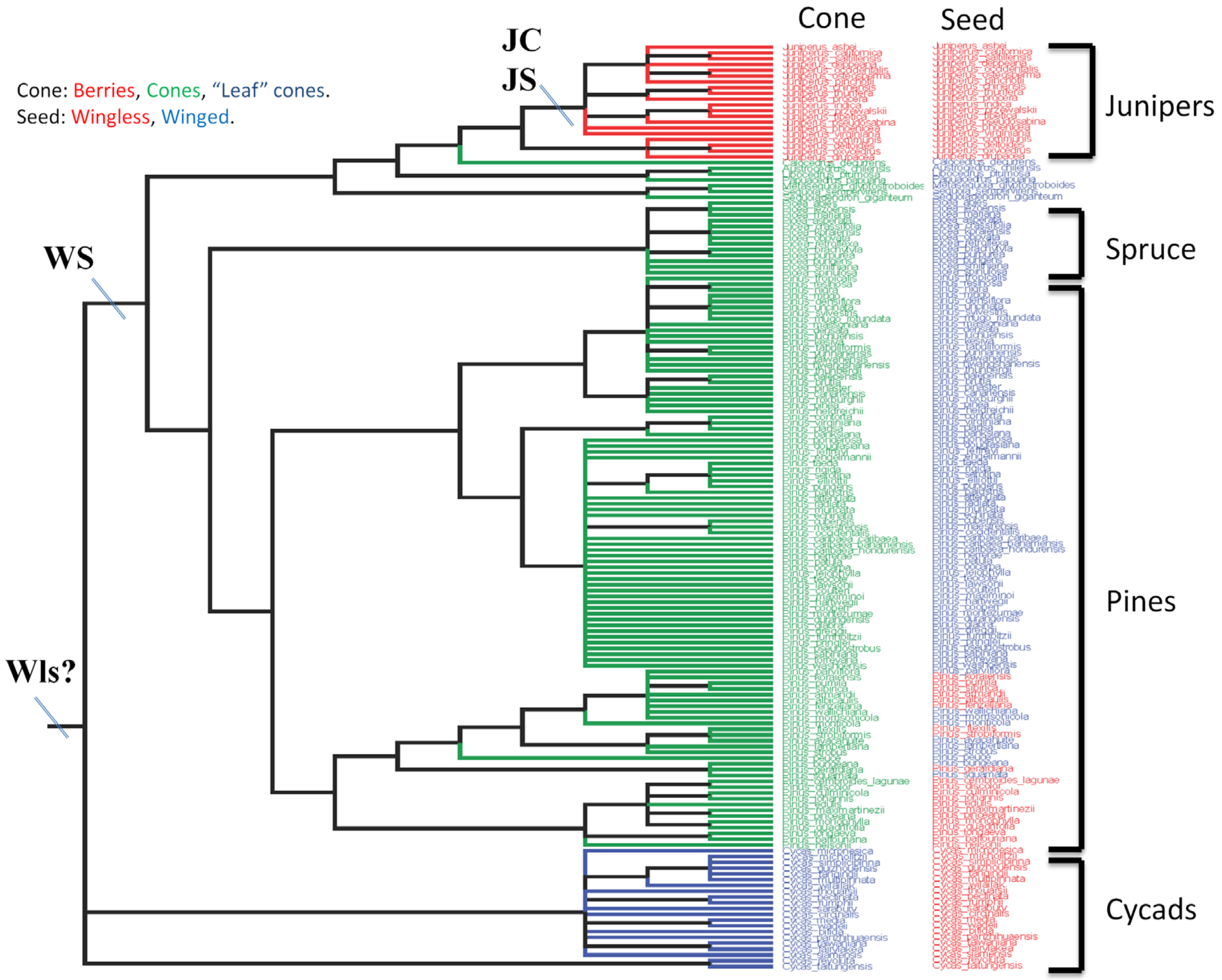

Figure 4: Tree of Cones and Winged/Wingless Seeds. A phylogenetic tree of gymnosperms color coded based on seed type and cone type. WS indicates that winged seeds are at least as old as the indicated clade (ancestral clade of all conifers of interest in this study). Since cycads are wingless, “WIs?” indicates possibility that the ancestral species to all species on this tree had wingless seeds. JC and JS indicate that the unique traits of junipers evolved independently since their ancestoral species, other conifers such as sequoias, differ from junipers in their seed and cone phenotypes.

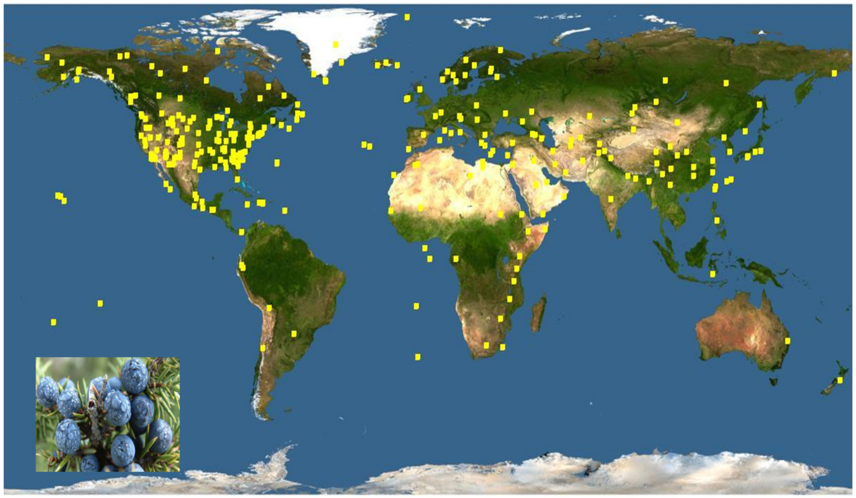

Figure 5: Distribution Map of Junipers. The map shows a wide distribution range of junipers that have wingless seeds. Note the cluster of junipers along the Swahili Coast and islands across the oceans that may indicate relationship with trade. A picture of juniper berries provides context for its edibility.

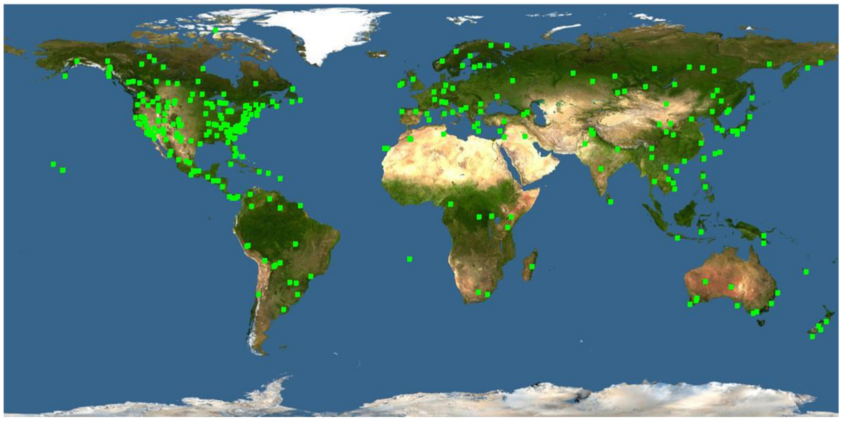

Figure 6: Distribution Map of Pines. The map shows the distribution of pines. The clusters of pines concentrated in Eastern and Western North America, Eastern Asia, and Europe suggests possibility that pine distribution has been largely affected by the breakup of the prehistoric single continental landmass, Pangea.

with edible seeds. Distribution patterns show clusters in Eastern Asia, some even in Russia, as well as in western North America. Such patterns, along with **Figure 8**, add support to the hypothesis that some pines crossed over from Asia to North America to create the species that exist today.

Discussion

Evolution of Leaf Type, Seeds, and Cones

Figure 4 maps the phenotypes of winged and wingless seeds as well as the shape of the cones. As mentioned in the results section, ginkgos have wingless seeds. This observation increases the likelihood that the common ancestor of conifers, ginkgos, and cycads were wingless. Regardless of the original seed type, junipers mutated individually to form their wingless seeds because **Figure 4** shows that junipers branched from cedars which branched out from sequoias, both of which have winged seeds. It is likely that junipers evolved independently of conifers in general since both their cones and seeds are different from the rest of the conifers. Therefore, junipers likely mutated just before the common ancestor of all junipers (JC and JS). **Figure 1** adds support for this speculation because it reveals repeated differences in several mapped phenotypes between junipers and the rest of the conifers.

Figure 5 revealed interesting patterns of juniper distribution. The generated hypothesis was that juniper distribution has been affected by humans and birds. A more specific scenario stems from **Figure 5**. **Figure 5** shows juniper populations clustered along the Swahili Coast and islands in the middle of the oceans. The Swahili Coast has long been a major trading center for many items including gin. It is also very likely that many of the islands were rest stops for trading vessels in the past as well. We also know that juniper cones are fleshy and berry-like (**Figure 5**). The berry encloses wingless seeds. Due to its fruit-like structure, juniper berries have been used for centuries by humans and animals as food and as medicine. Speculation produces an explanation for the global dispersal of junipers despite their wingless seeds. Junipers may have been spread by humans who carried the berries for food. In addition, the potential relationship between juniper distribution and trade networks indicates that juniper berries were possibly spread by ocean trade, which could explain the clusters of junipers on the Swahili Coast and the numerous islands.

Figure 6 provides a basis for speculation about the distribution of pines. Two major clusters of pines are in Eastern North America and Europe. These two areas were joined in Pangea. Pangea existed about 255-300 million years ago (mya). This fact supports the idea that pines originated from Pangea, which puts them at least 255-300 mya. Research indicates that such a prediction from my data would be consistent with the accepted origin date of pines. Pines are thought to be from the Carboniferous Period, which is about 300 mya (6).

One observation from the analysis of **Figure 7** revealed a discrepancy in clade 3 between winged and wingless seeds. However, all of clade 1 and clade 2 have winged seeds. A possible scenario that could explain this discrepancy is that a common ancestor of clade 3 mutated and became wingless. This seems possible because half of clade 3, clade 3b, has wingless seeds. The individual discrepancies in clade 3a could be independent species mutations. **Figure 8** revealed that the distribution maps of pines with winged seeds were similar to that of pines with wingless seeds. This is significant because it implicates the influence of external factors, such as other organisms, given that wingless seeds cannot be dispersed over a distance as great as the winged seeds which have been adapted for wind dispersal. In an effort to explain this pattern, it is important to note that certain wingless pine seeds are edible and have been eaten for centuries by humans. This information proposes that as

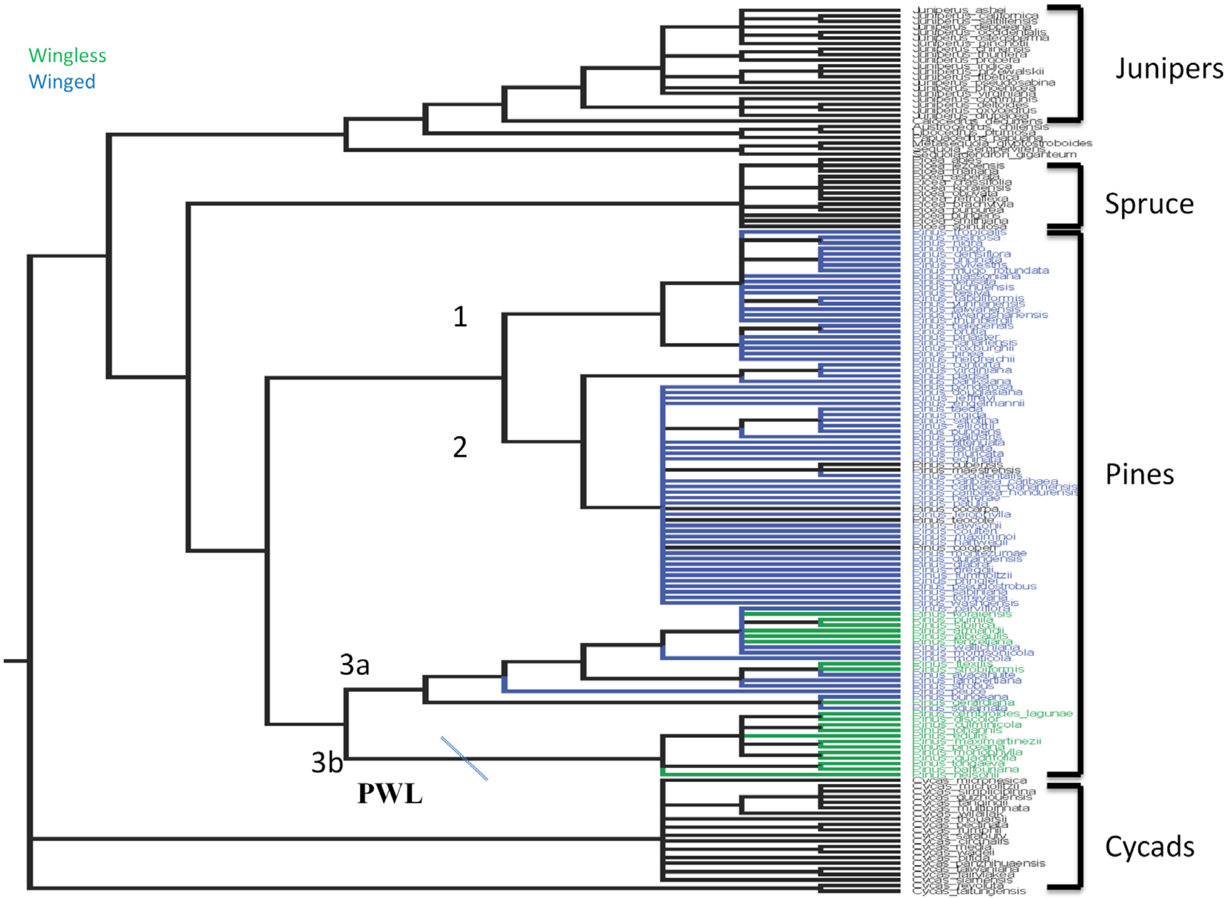

Figure 7: Tree of Seed Type of Pines. A tree focusing on the seed types of pines only; winged vs. wingless. Clade 1 and 2 species have winged seeds, but there are discrepancies in clade 3 with the majority of the species having wingless seeds. Refer to **Figure 8** for a combined analysis.

humans crossed from Asia to Western North America through the Bering Strait, people brought their stores of these edible seeds to North America. Nuts are nutritious and do not spoil easily, therefore it is reasonable to hypothesize that people carried edible pine nuts for food as they migrated. The location tree shown in **Figure 8A** supports this speculation. **Figure 8B** reveals that all the pines with edible seeds are clustered mainly in East Asia and Western North America, the two areas that were once connected by the Bering Strait. The location tree of **Figure 8** lends support since it reveals that the clade G of Asian pines is the ancestral species of clade 3a which include some North American pines. This presents the possibility that at least some of the North American pines are descendents of Asian pines, implying a previous connection between the two areas. Given the existence of humans on earth is relatively recent compared to the 300 million year history of pines, the descendents of the Asian pines are relatively recent divergences.

Future Experiments and Possible Errors

One major addition that would add support to this tree and the analysis are divergence dates. Divergence dates would require dated nodes or some external information that would require research of the fossil record of these species. Additional outgroups would also make the conclusions of the evolutionary relationships between these organisms more accurate. The current tree has been constructed using nucleotide sequences of the *matK* gene taken from the NCBI database. It is possible that there may be mistakes in the NCBI sequences, which would result in the production of different trees. There was a mistake discovered during the process: Pinus cembra. After coloring the tree based on phenotypes, this species was constantly different then all the other species in its clade and research indicated that it was a Strobus pine when the tree showed that it was related to the Pinus. Such mistakes could occur through errors in sequencing. Other possible errors may result from the programs used for analysis. However, since the data in this manuscript is similar to previously published research, such as the Tree of Life, it is likely to be an accurate analysis (5).

Conclusion

The phylogenetic tree discussed in this manuscript was constructed with the *matK* gene and depicts evolutionary relationships between gymnosperms.

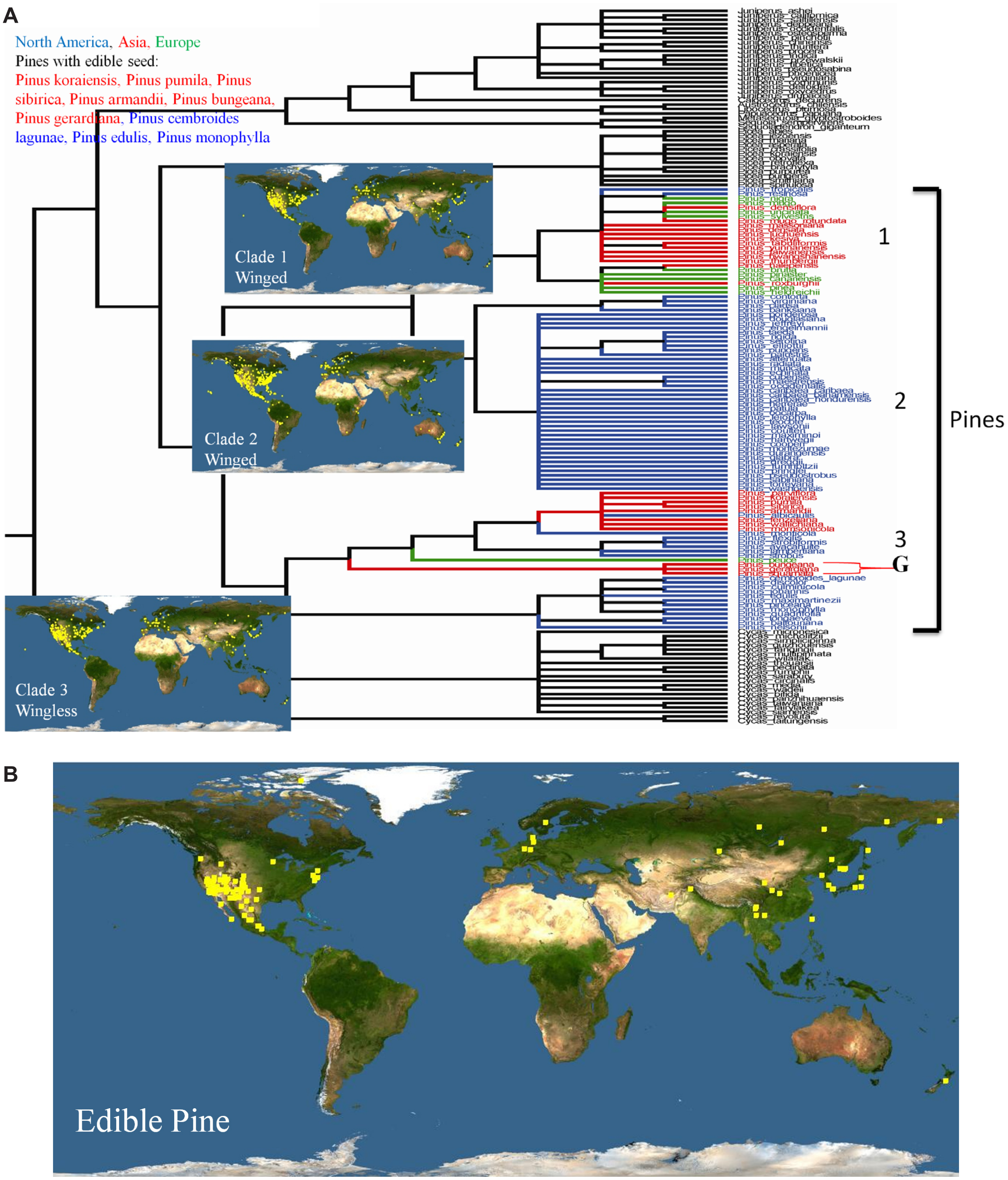

Figure 8: Location Tree with Maps. A. This tree is color coded by location (North America, Europe, Asia) that each pine specie is found in. The maps depict distribution patterns of the pines divided up by clade (1-3) as previously marked in Figure 1. Referring to **Figure 7**, the maps show that winged (clade 1 and 2) and wingless (clade 3) seed pines have similar distribution. Clade G in the location tree reveals that Asian pines are ancestors to at least some North American pines. **B**. A distribution map of pines containing edible seeds. The species are *Pinus koraiensis*, *Pinus pumila*, *Pinus sibirica*, *Pinus armandii*, *Pinus bungeana*, and Pinus gerardiana (color-coded red) from Asia and *Pinus cembroides lagunae*, *Pinus edulis*, and *Pinus monophylla* (blue) from North America. There are large clusters of pines containing edible seeds in Western North America and Eastern Asia indicating a potential connection between these clusters.

Mapped phenotypes of this tree are consistent published studies of the species. Pines, spruces, sequoias, and cedars have similar mapped phenotypes whereas junipers and cycads appear to be distinct. The leaf phenotype appears to have evolved through environmental selection and likely originated from Pangean times. Also, winged seeds evolved in the conifer clade from ancestral wingless seeds. Based on the phylogenetic analysis in this study, it appears that junipers evolved independently of conifers. The distribution map of pines generated in this study suggests that pines originated from Pangea. Harsh environments and competition in Pangea led to the evolution of a highly versatile type of leaves, the needles, of conifers. The distribution of wingless juniper seeds was almost certainly influenced by world history. Junipers spread to America likely through the movement of humans across the Bering Strait. Trade and the use of Juniper berries in gin/food/medicine influenced its distribution (for example, the Swahili Coast). Similarly, wingless seed species of pines spread around the world and from Asia to North America through trade and migration across the Bering Strait as people brought their edible nuts from pines in Asia to the Americas. The mixture of Asian/North American Pines in clade 3b suggests that the possible origin of North American Pines in clade 3b is Asia since the ancestral species are Asian. Overall, the results from the analysis of this phylogenetic tree further develop our understanding of conifers which are some of the oldest and most successful group of plants.

Methods

NCBI (www.ncbi.nlm.nih.gov) was used to obtain the nucleotide sequences for the *matK* gene. All of the sequences were compiled into one file in the FASTA format. The FASTA file was aligned using SEAVIEW. Maft was used as the alignment tool. The resulting alignment was saved as a NEXUS file. PAUP was the phylogenetic program used to create the tree from the saved NEXUS file. Cycads were defined as the outgroup. Parsimony and Heuristic Search were the conditions used to create the tree. To generate Bootstrap values, the “Bootstrap/ Jacknife Analysis” was used under the Analysis tab in PAUP. The distribution maps of all the species were produced by inputting the species’ scientific names into the following site: http://www.discoverlife.org/mp/20m?act=make_map. Finally, FigTree was used to color the different clades of the tree for analysis. The traits mapped were pollen, sex, cone type, leaf type, seed type, and sperm type for all species. Additional traits were mapped only for the pines. These traits were the following: location, winged/wingless seeds, number of leaves per fascicle, edibility, and cone size.

Editor’s Note

Higher resolution images are available online at: http://emerginginvestigators.org/articles/2013/10/phylogenetic-study-of-conifers

References

1. *NCBI*. n.p. n.d. <http://www.ncbi.nlm.nih.gov/guide/>.

2. Nagalingum, N. S., C. R. Marshall, T. B. Quental, H.S. Rai, D. P. Little, S. Mathews. “Recent Synchronous Radiation of a Living Fossil.” *Science*. 334 (2011): 796-799.

3. *Dendrome, A Forest Tree Genome Database*. SDA, NSF, UC Davis Plant Sciences. Web. n.p. n.d. <http://dendrome.ucdavis.edu/index.php>.

4. *The Gymnosperm Database*. Christopher J. Earle, n.d. Web. 27 May 2012. <http://www.conifers.org/index.php>.

5. Maddison, D. R. and K.-S. Schulz (eds.) *The Tree of Life Web Project*. 2007. <http://tolweb.org>

6. Raven, P.H., R. F. Evert, S. E. Eichhorn. *Biology of Plants*. Chapter 20: Gymnosperms. New York: Freeman, 1990.

7. *Global Mapper*. The Polistes Corporation, n.d. Web. 21 January 2012. <http://www.discoverlife.org/mp/20m?&act=make_map>.

8. *GeneCard*. n.p. n.d. Web. 4 Feb. 2012. <http://www.genecards.org/ >.

9. *Lovett Pinetum*. Robert Lovett, n.d. Web. 25 Jan. 2012. <http://www.lovett-pinetum.org/>.

10. *EMBL-EBI*. n.p. n.d. Web. 3 Feb. 2012. <http://www.ebi.ac.uk/interpro/IEntry?ac=IPR002866>.

11. Yu, Jing, Jian-Hua Xue, and Shi-Liang Zhou. “New universal matK primers for DNA barcoding angiosperms.” *Journal of Systematics and Evolution* 49(2011):176-181.

12. Gouy, M., S. Guindon and O. Gascuel. “SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building.” *Molecular Biology and Evolution* 27.2 (2010):221-224.

13. Rambaut, A. (2009). Figtree (Version 1.3.1) [software]. Available from http://tree.bio.ed.ac.uk/software/figtree/.

14. Swofford, D. (2012). Paup (Version 4.0) [software]. Available from http://paup.csit.fsu.edu/index.html.