Bioprospecting in the age of phylogenetic trees:

influence of evolutionary data on evaluation of natural products

Lena Struwe[1, 3], Sasha Eisenman [2], and Mariusz Tadych [2]

[1] *Dept. of Ecology, Evolution, and Natural Resources, Rutgers University - Cook College, New Brunswick, NJ 08901, U.S.A.*

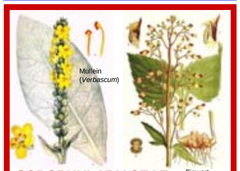
[2] *Dept. of Plant Biology and Pathology, Rutgers University - Cook College, New Brunswick, NJ 08901, U.S.A*

Abstract

Recent developments in evolutionary systematics and molecular sequencing have provided us with a stable family tree of all major plant groups on Earth. Bioprospecting of plants for new drugs and other useful properties can utilize this evolutionary-based information to focus screening on particularly high-yielding taxonomic groups. This require databases of samples to reflect the most recent evolutionary classifications so we can statistically identify the taxonomic groups that give the highest yield of positive screening results. For example, the taxonomy in the NAPIS database used for the ICBG global screening projects needs to be updated to permit scientific evaluation of the correlation between screening results and evolutionary data. An example is given of the discrepancies in the Lamiales (the Mint and Snapdragon plant order) between the most current classification and the NAPIS database. A road map describing the opportunities and processes of evolutionary-based bioprospecting is presented.

OROBANCHACEAE

LAMIACEAE



SCROPHULARIACEAE

VERONICACEAE

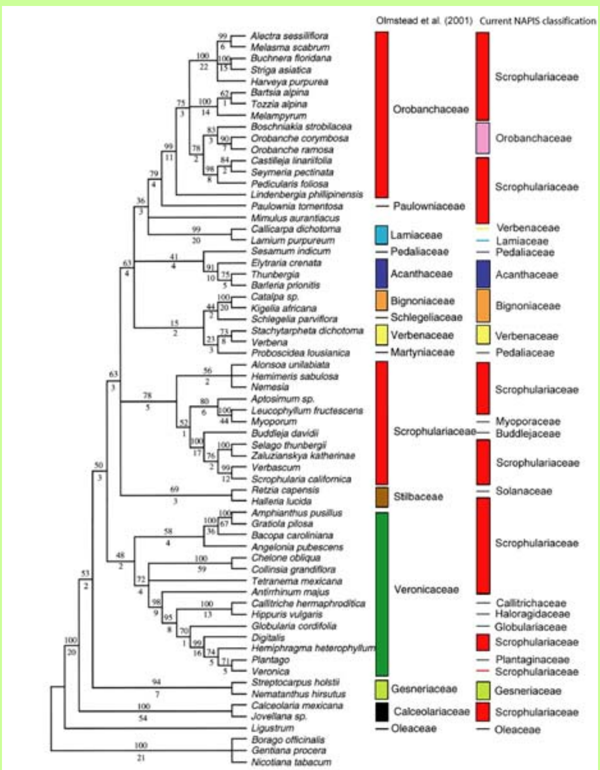


Figure 2. An evolutionary tree showing hypothesized relatiionships among mints, foxgloves, salvias, snapdragons, and related plants (order Lamiales). A new classification was based on this tree, with family groups as shown in the left-most column. The right-most column shows the previous classification, which is still used in the NAPIS database. For example, plantain (*Plantago*) was earlier placed in its own family, but its closest relative is *Veronica*, and they are now placed together in a new family called Veronicaceae. The older family Schrophulariaceae (in red) was shown to be an aggregate of many unrelated groups. Some representatives of medicinal plants of these families are shown to the very right. (Figure from Olmstead et al., 2001)

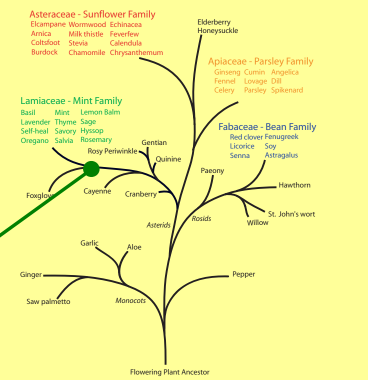


Figure 1. "Tree of Life" for some common medicinal plants based on most recently found evolutionary relationships using DNA sequencing. Note how many medicinal plants are aggregated into a few families that are not closely related, such as the Sunflower family, Mint family, and Bean Family. Drawing by Sasha Eisenman. (Evolutionary relationships after APG, 2003).

Introduction

There are over a quarter million known species of flowering plants on Earth, with only a small fraction investigated for their medicinal properties. Plants carry a great potential as sources for new drugs with their extensive production of secondary compounds and need to defend themselves against herbivore, viral, and fungal attacks. We present here a new approach to more efficient bioprospecting in these times of rising drug development costs and global biodiversity crisis (Farnsworth 1988; Gentry 1993).

Screening of plant diversity for new, bioactive compounds can take several approaches. Random samples of plants can be screened (the 'shot-gun' approach, as used by NCI; Cragg et al. 1994) or plants can be selected based on their use as herbals in native cultures (Cox & Balick 1994; Mendelson & Balick 1995; Singh et al. 2003). So far, plants have rarely been selected for screening based on their taxonomic neighborhood; i.e., close evolutionary relationship to plants with previously known activities. All approaches have yielded few new commercial drugs, but no study has compared the overall efficacy of the different approaches. Many efficacy studies have been done on a species-only basis, for example *Echinacea* in the sunflower family (Borchers et al. 2000). But so far, no study has looked at how sunflower plants (Asteraceae) score in comparison to other plant families, despite the abundance of medicinal plants in this family (e.g., chamomile, feverfew, milk thistle, wormwood, and yarrow; Fig. 1).

Plants derive a majority of their characteristics from their genetic heritage and ancestors, including their potentially bioactive compounds. Closely related species and genera are therefore more likely to share similar bioactive compounds, so invoking evolutionary data in the strategy for directed, successful pharmaceutical screening should lead to a higher success rate. This, however, has rarely been statistically tested, primarily due to the previous lack of a robust evolutionary-based family classification.

A new evolutionary framework

Extensive DNA sequencing among tens of thousands of plant species in the last few decades has lead to a revolution in plant systematics studies. New evolutionary hypotheses have been developed to construct a robust 'family-tree' of most plant families (e.g., Chase et al. 2003; Soltis et al. 2000; Qui et al. 2000; Fig. 1). Some traditional families were shown to include species from several different ancestors (polyphyletic) or did not include all descendants of a common ancestor (paraphyletic) and these have now been reclassified into new natural (monophyletic) families. This giant family tree has been translated into a new classification of all flowering plants, identifying only families and orders that are monophyletic, i.e., natural groups that strictly contain all (and only) descendants of a common ancestor (APG 1998, 2003). As seen in Figure 2, this has lead to extensive reclassification in some groups. The old large family Schrophulariaceae (the snapdragon family) is now smaller, and the snapdragon itself (*Anthirhinum*) has been moved to the new family Veronicaceae, which also includes foxglove (*Digitalis*), plantain (*Plantago*), and *Veronica* (Olmstead et al. 2001). All flowering plants are currently distributed in about 450 families and 45 orders (APG 2003). The recently published comprehensive hypotheses of vascular plant evolution will make it possible to place most genera of investigated plants in their evolutionary correct family. As shown in Figure 2, the taxonomy currently used in NAPIS is outdated and needs be modified to conform to modern standards as part of NIH drug development efforts.

Road map for evolutionary-based bioprospecting in ICBG

1.Continually update the taxonomic classification used in NAPIS to the most recent findings (with on-going funding for this provided directly from funding agencies)

2. Statistically evaluate the success-rate based on taxonomic groupings (can be based on particular disease targets or overall results)

3. Pinpoint evolutionary groups that yield higher than average positive screening hits

4. Start intense screening of additional species and populations within these taxonomic groups

Acknowledgments

Funding for the ICBG Central Asia project from NIH is gratefully acknowledged (PI Ilya Raskin: Rutgers University). The plant illustrations are drawings from *Nordens Flora* by C. A. M. Lindman (1901-05) and made freely available by Project Runeberg, Sweden (http://caliban.mpiz-koeln.mpg.de/~stueber/lindman/).

References

APG [= Angiosperm Phylogeny Group] II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Bot. J. Linnean Soc.* 141: 399-436.

APG [= Angiosperm Phylogeny Group]. 1998. An ordinal classification for the families of flowering plants. *Ann. Missouri Bot. Gard.* 85: 531-553.

Chase, M. W., et al. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. *Ann. Missouri Bot. Gard.* 80:528-580.

Farnsworth, N.R. 1988. Screening plant for new medicines. In: E.O.Wilson, ed., Biodiversity, pp. 83-97, National Academy Press, Washington, DC.

Gentry, A.H. 1993. Tropical forest biodiversity and the potential for new medicinal plants. *ACS Symposium Series* (Washington DC: American Chemical Society) 534: 13-20.

Cragg, G.M., M.R. Boyd, J.H. Cardellina, D.J. Newman, K.M. Snader, & T.G. McCloud. 1994. Ethnobotany and drug discovery: the experience of the US National Cancer Institute. *Ciba Foundation Symposium* 185: 178-190.

Cox, P.A. & M.J. Balick. 1994. The ethnobotanical approach to drug discovery. *Sci. Amer.* 270 (6): 82-87.

Mendelsohn, R. & M.J. Balick. 1995. The value of undiscovered pharmaceuticals in tropical forests. *Econ. Bot.* 49: 223-228

Singh, S.P., S. Tripathi, & R.S. Shukla. 2003. Ethnomedicinal heritage for bioprospecting and drug development in northeastern states of India. *J.Econ. Taxon. Bot.* 27(2): 384-395.

Borchers, A.T., C.L. Keen, J.S. Stern, & M.E. Gershwin. 2000. Inflammation and Native American Medicine: The role of botanicals. *Amer. J. Clin. Nutrition* 72(2): 339-347.

Soltis, D.E., et al. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. *Bot. J. Linnean Soc.* 133: 381-461.

Qiu, Y. L., J. H. Lee, F. Bernasconi-Quadroni, D. E. Soltis, P. S. Soltis, M. Zanis, E. A. Zimmer, Z. D. Chen, V. Savolainen, & M. W. Chase. 2000. Phylogeny of basal angiosperms: Analyses of five genes from three genomes. *Int. J. Plant Sci.* 161:S3-S27.

Presented as part of the ICBG Central Asia project by Rutgers University, University of Illinois-CHampaign, and partners in Uzbekistan and Kyrgyzstan

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS