Combining evolutionary inference and metabolomics to identify plants with medicinal potential
Mawalagedera, Sundara M. U. P.; Callahan, Damien L.; Gaskett, Anne C.; Rønsted, Nina; Symonds, Matthew R. E.

Published in:
Frontiers in Ecology and Evolution

DOI:
10.3389/fevo.2019.00267

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Mawalagedera, S. M. U. P., Callahan, D. L., Gaskett, A. C., Rønsted, N., & Symonds, M. R. E. (2019). Combining evolutionary inference and metabolomics to identify plants with medicinal potential. Frontiers in Ecology and Evolution, 7, [267]. https://doi.org/10.3389/fevo.2019.00267
Combining Evolutionary Inference and Metabolomics to Identify Plants With Medicinal Potential

Sundara M. U. P. Mawalagedera 1, Damien L. Callahan 2, Anne C. Gaskett 3,4, Nina Rønsted 5 and Matthew R. E. Symonds 1*

1 Faculty of Science, Engineering and the Built Environment, Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Burwood, VIC, Australia, 2 Faculty of Science, Engineering and the Built Environment, School of Life and Environmental Sciences, Deakin University, Burwood, VIC, Australia, 3 School of Biological Sciences, University of Auckland, Auckland, New Zealand, 4 Te Kura Mātauranga Koiora, Te Whare Wānanga o Tāmaki Makaurau, Auckland, New Zealand, 5 Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

Plants have been a source of medicines in human cultures for millennia. The past decade has seen a decline in plant-derived medicines due to the time-consuming nature of screening for biological activity and a narrow focus on individual candidate plant taxa. A phylogenetically informed approach can be both more comprehensive in taxonomic scope and more systematic, because it allows identification of evolutionary lineages with higher incidence of medicinal activity. For these reasons, phylogenetics is being increasingly applied to the identification of novel botanic sources of medicinal compounds. These biologically active compounds are normally derived from plant secondary or specialized metabolites generally produced as induced responses and often playing a crucial role in plant defense against herbivores and pathogens. Since these compounds are typically bioactive they serendipitously offer potential therapeutic properties for humans, resulting in their use by traditional societies and ultimately drug lead development by natural product chemists and pharmacologists. The expression of these metabolites is likely the result of coevolutionary processes between plants and the other species with which they interact and effective metabolites are thus selected upon through evolution. Recent research on plant phylogeny coupled with metabolomics, which is the comprehensive analysis of metabolite profiles, has identified that related taxa produce similar secondary metabolites, although correlations are dependent also on environmental factors. Modern mass spectrometry and bioinformatic chemical networking tools can now assist high throughput screening to discover structurally related and potentially new bioactive compounds. The combination of these metabolomic approaches with phylogenetic comparative analysis of the expression of metabolites across plant taxa could therefore greatly increase our capacity to identify taxa for medicinal potential. This review examines the current status of identification of new plant sources of medicine and the current limitations of identifying plants as drug candidates. It investigates how ethnobotanic knowledge, phylogenetics and novel approaches in metabolomics can be partnered to help in characterizing taxa with medicinal potential.

Keywords: evolution, medicinal plants, metabolomics, phylogenetics, secondary metabolites
PLANTS AS A SOURCE OF MEDICINE

Plants have been a perennial source of human therapeutics. The use of plants for medicine has been documented as early as 2600 BC (Gurib-Fakim, 2006) and plants still play a major role in treating human diseases. People in developing countries rely heavily on plant-derived traditional medicines (De Luca et al., 2012), accounting for close to 80% of world population (Gurib-Fakim, 2006). In the United States, a considerable proportion of approved drugs are derived from plants (9.1% according to the USFDA; United States Food and Drug Administration (Newman and Cragg, 2016). Furthermore, there are many drugs that are modified analogs of plant derived secondary metabolites that are in wide use (e.g., aspirin derived from the natural product salicin, from the willow tree Salix alba L., codeine and morphine from the opium poppy Papaver somniferum L.) (Dias et al., 2012). There are over 28,000 medicinal plants recorded in the scientific literature (Allkin, 2017) that are based on knowledge from traditional medicine.

Despite the prevalence of plants in traditional medicine, the majority of plant taxa have not received formal experimental appraisal for their medicinal properties. During the early twenty-first century there has actually been a decline in the use of plants as a source of medicine in the pharmaceutical industry (Atanasov et al., 2015). Of the plant species recorded as having therapeutic properties, only 16% have been tested for biological activity (Willis, 2017). Furthermore, the pharmaceutical industry more commonly uses synthetic chemistry for drug design, even though success in finding new leads are limited (Li and Vederas, 2009) and there is a vast diversity of plants with biological activity that could be drug candidates. This decline is a result of the challenges faced in the conventional plant-based drug discovery process which make the overall process time consuming (Figure 1). The process starts with identifying the species of plant known to have medicinal function, sourcing the plant material, obtaining approval according to international legislations such as the convention on biological diversity and the Nagoya protocol on access to genetic resources, isolating the bioactive compound or compounds, identification of it, and ultimately synthesizing the candidate compound (Liu and Wang, 2008; Atanasov et al., 2015). For example, in the case of the drug artesunate derived from artemisinin originally discovered from Artemisia annua L., used to treat malaria, there was an 18 month lag between initial identification and production of the candidate drug. The time delay was due to the 6 month Artemisia cultivation period followed by the bioactive compound purification period (Wells et al., 2015). This was prior to the longer process of bioassaying, and before clinical testing. In addition to these practical issues, identifying the biologically active compounds from the plant source is challenging due to the enormous chemical diversity of plants and possibility of synergistic effects among compounds (Kingston, 2011). Furthermore, elucidating the mechanism of action of plant-based drugs also remains a slow process. These are only the delays after a plant source has been identified, and do not consider the time taken to identify that original botanic source in the first place. Overall this process of screening plant taxa for biological activity is costly, so the pharmaceutical industry tends to opt for other sources for medicines.

The study of therapeutic traditional use of plants is a multidisciplinary area, because both pharmacology of the plant and human behavior has affected whether and how the plant is used. As an example in traditional medicine there are many instances where complex herbal mixtures are used (Leonti and Casu, 2013). In such instances the metabolites in the herbal mixture may need to act in synergy to yield the pharmacological effects (Wagner, 2011). Further, some plant-based traditional medicines may also have different preparation methods. Based on the preparation methods the mode of action of the medicinal metabolites may differ (Heinrich et al., 2009). In addition to that some traditional communities may also select plants based on specific traits (De Medeiros et al., 2015). Therefore, consideration should be given to all of these related processes in ethnopharmacological studies. Despite these complications, ethnopharmacology still does provide a credible starting point in narrowing the target taxa for natural product chemistry research.

There are also significant challenges for managing how ethnopharmacological knowledge is partnered with scientific discovery methods and the pharmaceutical industry. Cultural appropriation and exploitation of traditional knowledge is a serious and ongoing issue given that intellectual property is often regulated by researchers and pharmaceutical companies (Mcgonigle, 2016). Thus a mechanism to acknowledge the indigenous knowledge of ethnobotany also needs to be adopted in drug discovery (Gupta et al., 2005). The International Platform on Biodiversity and Ecosystem Services (IPBES) established in 2012, acknowledges by developing and governing a set of principles that recognize the contribution and knowledge of indigenous and local cultures in the use of biodiversity (Pert et al., 2015). Thus, the use of a proper legal framework and collaborations of indigenous communities, scientist and policymakers may also greatly help in sharing and utilizing the wealth of ethnopharmacological knowledge (Shane, 2004).

Largely in the pharmaceutical industry, drug discovery novelties are few and the need for bioactive compounds are persisting. Given the lengthy process of sampling, screening and biological testing, many plants are not investigated. Despite evidence that plants are a rich source of bioactive compounds. Hence the need for alternative approaches, to identify plants in a systematic manner for their medicinal potential is required. Novel approaches for identifying medicinal plant taxa more effectively need to be explored (Liu and Wang, 2008).

Recent advancements in the screening processes for natural products are stimulating renewed effort in plant-based drug discovery. Most notably, the 2015 Nobel Prize for medicine was awarded for identifying an antimalarial compound from Artemisia annua L. using the knowledge of traditional Chinese medicine (Tu, 2011). Currently the two main drugs used to treat malaria are plant derived (artemisinin and quinine from Cinchona L. barks) highlighting that plant-derived pharmaceuticals are highly relevant for drug discovery. Furthermore, during the last decade, several new plant-derived drugs (apomorphine hydrochloride; water lily (Nymphaea L. sp), galanthamine hydrobromide; snowdrop (Galanthus nivalis L.),...
nitisinone; bottlebrush plant (*Callistemon citrinus* Curtis.), tiotropium bromide; devil’s snare (*Datura stramonium* L.) and Cytisine; golden rain acacia (*Cytisus laborinum* L.) have been approved by the USFDA (Das, 2017; Auffret et al., 2018). According to World Bank projections, a growth rate of 5–15% annually in the plant-based medicinal market is anticipated (Liu and Wang, 2008). Thus, identifying biologically active potential medicinal plant sources is currently a commercially viable and beneficial research area, albeit one with significant cultural issues.

**ROLE OF PLANT SECONDARY/SPECIALIZED METABOLITES IN TREATING DISEASES**

A medicinal plant is defined as a botanical source able to cure, prevent or relieve a disease, or a plant that is utilized as a drug or a precursor for a drug (Rates, 2001). Most of the active components of newly developed drugs are plant secondary metabolites (Balunas and Kinghorn, 2005). Plant secondary metabolites are
Many plant secondary metabolites serve as defensive compounds produced to overcome the challenges plants face in the environment, including pathogens and herbivores but also abiotic stressors (Wink, 1988). In addition, some of these secondary metabolites also serve as allelopathic defensive compounds, synthesized in response to competition from surrounding plants (Reigosa and Pazos-Malvido, 2007). These chemical entities are specialized/secondary metabolites, which are typically expressed systemically, but occasionally the expression of these secondary metabolites is localized.

Plant secondary metabolites are structurally specialized by having highly active functional groups (i.e., aldehyde, sulphydryl groups, epoxides, hydroxyl, and carbonyl groups) to act on cellular targets, such as enzymes, cell receptors and transporters. Therefore, when a medicinal plant product is ingested by humans, the secondary metabolites mediate a chain of reactions that may be beneficial in treating diseases (Acamovic and Brooker, 2005; Wink, 2015). As a result of their effect on the biological systems of hosts, secondary metabolites exhibit major properties as for example antiviral, antimicrobial, antifungal, antimalarial, analgesic, antiarrhythmic, antihypertensive, psychoactive, and tumor inhibition agents. However, out of c.100,000 known bioactive secondary metabolites, remarkably few have been scientifically screened for their activity (Wink, 2015).

The biggest challenge is identifying the target bioactive compound from the plant material, since every crude plant extract contains thousands of compounds, most of which are unknown (Atanasov et al., 2015). Thus, identification of biologically active compounds is done through comparing the chemical structures of compounds identified from the plant, to those already existing in compound libraries typically used in analytical processes (i.e., mass spectrometry (MS) and nuclear magnetic spectroscopy (NMR) (Dias et al., 2016). Examples of these include KNapsack (Afendi et al., 2012), the Universal Natural Product Database (UNPD) which is an open source library with bioactivity information, the Global Natural Product Database (GNPS) and Massbank (Johnson and Lange, 2015). However, these libraries do not include all the natural products that plants produce (Schwager et al., 2008). Furthermore, inevitably if searching for new candidate drugs, many plant compounds are likely to be new to science (Takeuchi et al., 2018), and hence not present in compound libraries. Therefore, studies providing broad identification of the secondary metabolites with biological activity from plants, can provide an enormous amount of information applicable to the discovery of new therapeutic plants (Harvey et al., 2015).

**IDENTIFYING THERAPEUTIC PLANTS ACROSS THE PLANT KINGDOM**

Known medicinal activity is not randomly distributed across the plant kingdom. There are specific distribution patterns where certain lineages are significantly more bioactive (Coley et al., 2003). This pattern is also reflected in the traditional use of plants by humans. Different cultures around the world tend to use related plant taxa for similar medicinal purposes (Saslis-Lagoudakis et al., 2011). This ethnomedical convergence suggests that therapeutic potential may have phylogenetic constraints (Zhu et al., 2011; Kowiyou et al., 2015; Garnatje et al., 2017). Therefore, by investigating ethnomedical uses of plants, phylogenetically related target species for further investigation could be identified.

Phylogenetic exploration of approved drugs of natural origin has identified 11 major plant clades that have delivered the significant part of these approved drugs (Zhu et al., 2011). Likewise, there is evidence of particular plant lineages expressing specific secondary metabolites with medicinal function (Ronsted et al., 2012). For example, in-depth studies of the synthesis of the secondary metabolites of terpene structure reveal that there are seven groups of gene families responsible for terpene synthesis. These genes are specific to plant lineages (land plants, vascular plants, gymnosperms) and the types of terpenes they synthesize varies, respectively between clades (Feng et al., 2011). Thus, a broader understanding of how the genes responsible for secondary metabolism have evolved across plant taxa will provide information on the type of secondary metabolites synthesized by a plant lineage.

Secondary metabolites can be either ubiquitous or expressed in specific plant lineages (Chezem and Clay, 2016). Since secondary metabolites confer survival benefits to plants, they are under natural selection (Wink, 2003). The genes responsible for plant secondary metabolism are versatile and have high plasticity to adapt to environmental pressures. In contrast, the genes responsible for primary metabolism are stringent and expression is generally relatively inflexible (Hartmann, 2007). Genes responsible for secondary metabolism therefore have the ability to be upregulated and downregulated across lineages in response to selection imposed by the other species with which they interact (e.g., herbivorous insects) or environmental stress factors. This presents extra challenges to drug discovery as environmental factors can also play a role in the amount of a particular bioactive compound present in a natural product extract.

Differential gene expression is responsible for the lineage specificity of secondary metabolites. For example, in Arabidopsis thaliana (L.) Heynh. and other closely related Brassicaceae plants, glucosinolate expression is controlled by the methylthioalkylmalate (MAM) synthase gene cluster (Benderoth et al., 2006). Studies of differential expression of MAM genes have shown that positive natural selection has shaped the synthesis of glucosinolates as a defense mechanism to insect herbivory (Benderoth et al., 2006; Figueiredo et al., 2008). The phylogenetic history of Arabidopsis and its close relatives reveals specific gene duplication points, where new MAM genes were positively selected for the combinations of glucosinolates that those genes encoded (Benderoth et al., 2006). Synthesis of many plant secondary metabolites is regulated by different gene families, and they could have similar underlying gene duplication and selection histories explaining their diversity.
Differential gene expression in secondary metabolism is also driven by coevolution—the correlated evolution between two groups of ecologically interacting organisms, resulting in reciprocal evolution of both groups in response to each other (Ehrlich and Raven, 1964). Coevolution-mediated expression of secondary metabolites has been a major hypothesis in understanding the differential expression of secondary metabolites across plant lineages (Speed et al., 2015). One of the best studied models of coevolution concerns the chemical diversity of the family Apiaceae (which includes such well-known plants as dill; *Anethum graveolens* L., coriander; *Coriandrum sativum* L., cumin; *Cuminum cyminum* L. and fennel; *Foeniculum vulgare* Mill.). The coevolution of the Apiaceae and *Papilio* butterflies is reflected in their shared patterns of phylogeny where chemical shifts in the phylogeny of the butterflies has governed host plant diversifications (Berenbaum, 2001; Berenbaum and Zangerl, 2008). Several other well-documented examples of reciprocal adaptations between host plants and their specialist herbivores include *Barbarea vulgaris* (Aiton) (rocketcress; Brassicaceae) and *Phyllotreta nemorum* L. (turnip flea beetle), Brassicaceae plants and Pierine butterflies, and *Pastinaca sativa* L. (Parsnips; Apiaceae) and *Depressaria pastinacella* (Duponchel) (Parsnip webworm) (Zangerl et al., 2003; Toju and Sota, 2006; Agrawal et al., 2012b; Edger et al., 2015).

Plant secondary metabolite expression can also be driven by generalized, rather than specialized, herbivores. Recent research on Amazonian trees (from the tribe Proteae; Burseraceae) showed that secondary metabolites able to act as herbivore repellents were found frequently in high abundances in Proteae tribe plants, even in the absence of specialized herbivores. This suggests that regardless of the type of herbivore, the mere existence of any natural enemy can affect how secondary metabolites are expressed in plants (Salazar et al., 2018). Conversely, *Oenothera biennis* L. (common evening-primrose) populations grown in controlled insect pest-free environments, had reduced defenses against insects pests over time, i.e., they reduced their production of toxic secondary metabolite in response to decreased herbivory (Agrawal et al., 2012a).

Some of the chemical drivers behind the expression of these secondary metabolites are herbivore-associated elicitors (metabolites released into the plant by the herbivores). Many of these elicitors are produced by insects, and range from enzymes and modified lipids to sulfur containing fatty acids (Bonaventure et al., 2011). These elicitors have the ability to stimulate biosynthetic pathways within a plant (Waterman et al., 2019). Further, the action of these elicitors tends to be associated with certain plant-insect associations. For example the lepidopteran *Manduca sexta* L. (tobacco hornworm) releases fatty acid-amino acid conjugates into the plant that stimulate the biosynthesis of jasmonic acid in *Nicotiana attenuata* Torr. ex S.Watson (wild tobacco) (Wu et al., 2007; Meldau et al., 2011). Jasmonic acid afterwards mediates the synthesis of secondary metabolites as defense compounds in *N. attenuata* (Kallenbach et al., 2010). Thus, expression of secondary metabolites is governed both by evolution (i.e., one-way adaptation in response to generalized herbivory) and coevolution (i.e., specific reciprocal adaptation between the interacting host and a herbivore species) (Caseys et al., 2015).

Overall, the genetic mediation of plant secondary metabolite expression does not act in isolation. There are abiotic factors such as temperature, soil pH, and frost exposure that may have effects on overall plant secondary metabolism. In *Zea mays* L. (corn) changes in abiotic factors as soil humidity, temperature and light had significant effects on the volatile secondary metabolite production (Gouinguen and Turlings, 2002). These environmental stresses are also correlated with the abundance of natural enemies or herbivores and plant pathogens which themselves affect the expression of secondary metabolites (Müller and Orias, 2018). Therefore, studying secondary metabolite expression, in the light of coevolution would provide a more holistic and biologically relevant method to understand differential gene expression, and thus the distribution and occurrence of plant secondary metabolites. This has theoretical implications for the tracking the expression of traits according to phylogeny, as well as for medicinal and drug discovery.

**USE OF PHYLOGENY AND METABOLOMICS TO IDENTIFY BIOLOGICAL ACTIVITY IN PLANTS**

Understanding how secondary metabolite gene expression varies across clades is a potential key to identifying biologically active plants. For example, the kinds of enzyme-encoding gene clusters used by algae and mosses are distinctively different from those found in grasses and eudicots (Chae et al., 2014). Furthermore, analysis of the metabolomic reactions within these plant groups has shown that closely-related taxa express the same set of reactions, thus suggesting phylogenetic constraints on secondary metabolite expression. If true, in depth analysis of plant secondary metabolite evolution would help us to understand which taxa/clades are biologically active, and hence are good candidates as medicinal species. In other words, the phylogenetic patterns of secondary metabolite expression could potentially act as a marker for biological activity, and therefore potential medicinal use.

There have been numerous studies in the past decade describing phylogenetic patterns of secondary metabolite expression. Analysis of volatile terpenes from angiosperm taxa has shown phylogenetically conserved expression, with related species producing the same terpenoid across the phylogeny. In particular, species belonging to the Magnoliid clade such as orders Laurales, Magnoliales had very high terpene diversity (Courtois et al., 2016). This suggests that the origin of the Magnoliids 122–125 million years ago is a key evolutionary point at which plant volatile terpene synthesis increased significantly. The Magnoliids are also one of the eleven clades predicted to have significantly higher pharmacological activity (Zhu et al., 2011). Indeed, over a quarter (18 families) of the 66 drug productive families identified, are from the Magnoliid clade (Zhu et al., 2011). Other lineages with high degrees of bioactivity are mosses from the order Hypnales, conifers (Pinales, was Coniferales), plus angiosperms from the monocots (Commeliniids, and orders
Asparagales and Liliales), the basal or early diverging eudicots (Ranunculales) and more derived eudicots (Fabids and Malvids from the Rosids, and Lamids and Campanulids from the Asterids) (Zhu et al., 2011; APG IV 2016). Phylogenetic patterns in bioactivity (through secondary metabolite expression) have also been examined at the genus level for the angiosperms Euphorbia (Euphorbiaceae; Rosid) and Narcissus (Amaryllidaceae; Asparagales) with both genera yielding significant phylogenetic signals in biologically active secondary metabolite expression (Rønsted et al., 2008; Ernst et al., 2016).

The strong patterns of phylogenetically conserved expression of secondary metabolites could underlie the observed ethnobotanical convergence, in which similar plant taxa are used in similar ways in different parts of the world. Research on cross cultural patterns of medicinal plant usage reports the use of taxa from the same clades, across continents to treat similar diseases, signifying that there is phylogenetically non-random selection of therapeutic plants by indigenous cultures (Saslis-Lagoudakis et al., 2011, 2012). As an example, there are accounts for using Piper umbellatum L. (cow foot leaf; Piperaceae, Magnoliid) to treat kidney and digestive diseases across North America, Africa and South East Asia (Roersch, 2010). Further cross-cultural patterns have been revealed in a larger study across a number of genera e.g., Aloe Aloe vera (L.) Burm. f. (Asphodelaceae; Asparagales), the breadfruit Artocarpus altulis (Parkinson) Fosberg (Moraceae; Rosid), Papaw Carica papaya L. (Caricaceae; Brassicales) and love vine Cassytha filiformis L. (Lauraceae, Magnoliid) used for medicinal preparations across the Caribbean Islands (Halberstein, 2005). Thus, information from traditional medicine, along with biological active metabolite expression, coupled with phylogeny could be the strategy for next generation drug productive taxa identification.

Recent research has used phylogenetic approaches to predict medicinal potential of plants. For example, phylogenetic mapping was performed for plant species with compounds used for treating cardiovascular disease, based on ethnobotanic and pharmacological mode of action of the drug. Seven angiosperm plant families were identified as having similar pharmacology: Zingiberaceae (Commelinid monocot), Brassicaceae, Fabaceae, Malvaceae, Rosaceae (Rosids), and Apiaceae and Lamiaceae.
Figure 3 | Proposed integrated approach for identifying potentially medicinal plant taxa using metabolomics profiles of species A, B, C, and D (phylogenetically closely related species belonging to the same operational taxonomic unit/family) and species P, Q, and R (likewise but from a different operational taxonomic unit/family). (A) Hypothetical secondary metabolites identified using UHPLC-MSMS with respect to taxa A, B, C, D, P, and Q. Only taxa A and Q have previously known traditional medicinal uses. Taxa A and Q express the already identified bioactive metabolites X and Y, respectively. Taxon D is phylogenetically closely related to taxa A, B, and C. It does not have any known medicinal properties or identified bioactive metabolites. Taxon R is phylogenetically closely related to taxa P and Q. It does not have known medicinal properties or identified bioactive metabolites. (B) Hypothetical GNPS-based molecular network with the molecular phylogeny of the corresponding taxa. Molecular phylogenetic cluster producing molecular network related to bioactive compound X and Y, respectively. Taxa A and Q are also potentially medicinal based on both phylogeny and metabolomics.*In the molecular network the nodes are further divided into pie charts to represent the taxa sharing a particular metabolite. The area corresponds to the parent ion intensities in each node produced by respective taxa. The color code of the nodes correspond to the color code on the metabolites table and the phylogeny for ease of understanding the figure.

(Asteriids). Those families have been suggested as potential target clades for identifying novel leads in treating cardiovascular diseases (Guzman and Molina, 2018). Despite these promising advances, research to date has not specifically quantified the capacity of phylogeny to predict medicinal potential in specific lineages. This can be done by measuring the phylogenetic signal of a trait (Münkemüller et al., 2012), in this case, biological activity, across a large clade of plants with a high frequency of reported traditional medicinal use.

Phylogenetic signal reflects the tendency of closely-related species to be phenotypically similar in contrast to species drawn at random from the same phylogeny (Blomberg et al., 2003). It can be measured via a number of metrics for data that are continuous (quantitative) or discrete (e.g., presence/absence...
data) (Fritz and Purvis, 2010; Münkemüller et al., 2012; Kamilar and Cooper, 2013). In a medicinal use context, Fritz and Purvis's (2010) D statistic has been used to compare the phylogenetic distribution of presence/absence of both therapeutic use and expression of particular compounds across plant species, and whether this conforms to either a random distribution (with no phylogenetic signal) or distribution that would be expected if the trait had evolved by Brownian motion (gradual divergent evolution) (Saslis-Lagoudakis et al., 2015). This has been used to identify a phylogenetic signal of medicinal potential in the family Amaryllidaceae in relation to alkaloid diversity, where five of seven alkaloid groups exhibited weak, but significant, phylogenetic signal (Ronsted et al., 2012). The D statistic was also used to map the phylogenetic signal in leaf succulence and medicinal use across the genus Aloe finding that a succulent leaf habit is associated with the potential of a species to be used as a medicine (Grace et al., 2015). Furthermore, the D statistic has also been used to predict presence of artemisinin and antimalarial properties in other Artemisia (other than Artemisia annua L. (Pellicer et al., 2018).

**USING PHYLOGENETIC TRAIT MAPPING AND MOLECULAR NETWORKING TO PREDICT MEDICINAL POTENTIAL**

Novel techniques of metabolomics such as molecular networking can be used to identify metabolites that are structurally similar to already known bioactive metabolites (Allard et al., 2016). In molecular networking the fragmentation pattern of a metabolite (Tandem mass spectrometry-MS/MS fragmentation) is compared with other compounds and similar metabolites are identified based on the structure (Figure 2). These are then put into network clusters, where a single cluster would therefore refer to a single metabolite group that likely share many similar chemical (and hence potentially bioactive) properties. Therefore, this technique allows one to identify groups of potentially bioactive metabolites, and novel compounds with such properties. Further, by treating such networks as phenotypic characters, one can map their expression onto phylogenies and identify clades with related bioactive properties in a way that does not rely specifically on the identification of single compounds.

Tandem mass spectrometry-MS/MS based molecular networking has been extremely informative in chemical similarity studies identifying related chemical entities across taxa/clades. Once the chemical entities are identified using Global Natural Product Social Molecular Networking (GNPS), it is possible to identify which taxa share the same molecular network cluster. Furthermore, MS/MS spectra can be related to biosynthetic pathways and their biological activity using online databases for natural products, thus providing information on the medicinal potential of a taxon. One limitation with this approach is that it requires some pre-existing knowledge of chemical compounds' biological activity to identify analogs with similar potential. It may not be so effective in identifying novel biologically active compounds that have completely undescribed chemical composition.

There have been numerous studies identifying novel bioactive candidate compounds using molecular networking in species such as in the Euphorbiaceae species Euphorbia dendroides L. (tree spurge) and Codiaeum peltatum (Labill.) P.S.Green (croton), and identifying chemicals of cannabinomimetic activity from the cyanobacteria genus Moorea (Kleigrewe et al., 2015; Nothias et al., 2018; Olivon et al., 2018). However, the focus of these studies was on single taxa without considering their evolutionary history. Multi-species studies do exist, as in the recent study of acetylcholinesterase inhibitors produced by a range of Iranian monocot and eudicot flowering plant taxa, where new compounds as well as taxa responsible for expressing the new compounds were identified (Abbas-Mohammadi et al., 2018), but again the study taxa here were restricted only to known medicinal species, with little consideration of phylogenetic relatedness except in few recent research concerning Euphorbiaceae (Ernst et al., 2018). However, with the existence of phylogenetic and metabolomic information across multiple species, this approach could be further scaled up to comprising taxa with ethnobotanic medicinal use as well as their sister taxa for which no known records of ethnobotanical medicinal use exist, to identify target taxa for medicinal potential (see Figure 3).

In summary, the combination of ethnobotanic information, phylogeny and molecular networking provides a promising approach to plant natural product-based drug discovery. By investigating the metabolomics profiles of, a target highly bioactive plant clade comprising both known medicinal and understudied taxa, chemically related taxa could be identified via molecular networking. The phylogenetic signal of those chemical compounds (or chemical networks) could be quantitatively measured by constructing the phylogeny and using phylogenetic signal-measuring statistical methods. This will allow prediction of the medicinal potential of previously understudied taxa. Combining these approaches holds a key to the advancement of plant derived natural product drug discovery.

**AUTHOR CONTRIBUTIONS**

SM did the literature survey and wrote the paper. MS designed the research and edited the paper. DC edited and provided metabolomic insight to the paper. AG and NR edited the paper.

**REFERENCES**

Abbas-Mohammadi, M., Moridi Farimani, M., Salehi, P., Nejad Ebrahimi, S., Soniboli, A., Kelso, C., et al. (2018). Acetylcholinesterase-inhibitory activity of iranian plants: combined hplc/bioassay-guided fractionation, molecular networking and docking strategies for the dereplication of active compounds. *J. Pharm. Biomed. Anal.* 158, 471–479. doi: 10.1016/j.jpba.2018.06.026

Acmovic, T., and Brooker, J. D. (2005). Biochemistry of plant secondary metabolites and their effects in animals. *Proc. Nutr. Soc.* 64, 403–412. doi: 10.1079/PNS2005449
Wink, M. (2015). Modes of action of herbal medicines and plant secondary metabolites. *Medicines* 2, 251–286. doi: 10.3390/medicines2030251

Wu, J., Hettenhausen, C., Meldau, S., and Baldwin, I. T. (2007). Herbivory rapidly activates mapk signaling in attacked and unattacked leaf regions but not between leaves of *nicotiana attenuata*. *Plant Cell* 19, 1096–1122. doi: 10.1105/tpc.106.049353

Zangerl, A. R., Berenbaum, M. R., and Mallet, J. (2003). Phenotype matching in wild parsnip and parsnip webworms: causes and consequences. *Evol.* 57, 806–815. doi: 10.1111/j.0014-3820.2003.tb00292.x

Zhu, F., Qin, C., Tao, L., Liu, X., Shi, Z., Ma, X., et al. (2011). Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. *Proc. Nat. Acad. Sci USA.* 108, 12943–12948. doi: 10.1073/pnas.1107336108

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mawalagedera, Callahan, Gaskett, Rønsted and Symonds. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.