**Title page**

A phylogeny-based analysis of bioactive metabolites in *Scutellaria* reveals multiple species with significant medicinal potential

**Footnotes**

**Abstract (250 word limit)**

The *Scutellaria* genus contains multiple species of plants that have been used extensively in traditional medicines due to their various anti-inflammatory, sedative, and neuroprotective effects. Bioactive metabolites, especially 4’-hydroxyflavones and 4’-deoxyflavones, have been identified as a significant source of these effects. S*. baicalensis*,one of the most well-known species of *Scutellaria*, accumulates 4’-deoxyflavones in its roots and 4’-hydroxyflavones in its shoots. Although the biochemistry of *S. baicalensis* responsible for this organ-specific pattern has been studied, the *Scutellaria* genus contains more than 470 species of plants, many of which have not been chemically analyzed. The primary goal of this work was to characterize the chemodiversity of *Scutellaria*, and use this information to identify to species with significant medicinal value and biotechnology potential. We compared aerial metabolite profiles for six 4’-hydroxyflavones, eight 4’-deoxyflavones, and one other metabolite to a phylogenetic tree constructed from chloroplast genome sequences for 51 species of *Scutellaria*. Based on this initial profiling, we selected 13 species for further organ-specific and genome size analysis. Overall, our phylogeny and metabolite profile data were only weakly correlated, but we did find 4’-deoxyflavones to be much more well-conserved across the genus than 4’-hydroxyflavones. Additionally, we identified several species with significant 4’-deoxyflavone accumulation in their aerial tissues, suggesting a divergence from the organ-specific pathway described in *S. baicalensis*. From the species we selected for further profiling, *S. racemosa*, *S. wrightii*, and *S. dependens* stood out candidates for medicinal study due to their high flavonoid accumulation and relatively small genome size.

**Key words**

Chemodiversity, flavonoid, skullcap, medicine, biotechnology

**INTRODUCTION**

Medicinal plants have played an important role in the traditional medicines of many indigenous populations for thousands of years (Shang et al., 2010). Due to this widespread usage, modern research techniques are being applied to identify the specific compounds responsible for these medicinal properties, and better characterize their method of action (Shang et al., 2010). A negative consequence of this increased attention to and demand for medicinal plants, is the endangerment of native plant populations resulting from overharvesting. Production efficiency and scale is also limited by this approach (I. B. Cole et al., 2007). Therefore, development of mass production systems for these medicinal compounds is extremely desirable. As chemical synthesis methods are limited by their expense and relative inefficiency, biotechnology-based methods are currently the most promising means of mass production (Yang et al., 2016). In this work, we analyze the metabolite diversity of a genus of medicinal plants, *Scutellaria*, and identify several species which are promising candidates for biotechnology improvement.

Part of the mint family Lamiaceae, the *Scutellaria* genus contains multiple species of plants renowned for their medicinal properties. *S. barbata* is commonly used in traditional Eastern medicines to treat swelling and inflammation, and multiple reports have recently been published describing its effectiveness in cancer treatments (Q. Wang et al., 2018). *S. lateriflora* is used in traditional Native American treatments as a nerve tonic and sedative, with recent studies demonstrating its anticonvulsant and anxiolytic properties (Awad et al., 2003; Zhang et al., 2009). Perhaps the most well-known species is *S. baicalensis* due to its extensive use in traditional Eastern remedies. More commonly called Huang Qin in Chinese medicine, the roots of *S. baicalensis* are prescribed to treat a variety of ailments, including edema, dysentery, pneumonia, jaundice, and more (T. Zhao et al., 2019). In clinical studies, *S. baicalensis* root extracts have been found to exhibit neuroprotective, antibacterial, antitumor, antioxidant, and other beneficial health effects (Tao et al., 2018; Venkatarame Gowda Saralamma et al., 2017; Zhu et al., 2016).

These beneficial effects can be largely attributed to the bioactive metabolites which these *Scutellaria* species accumulate in high concentrations (Karimov & Botirov, 2017; Q. Zhao, Chen, et al., 2016). 4’-hydroxyflavones, which include apigenin and its derivatives, have been isolated from multiple *Scutellaria* species, but are also widely distributed throughout multiple families in the plant kingdom. 4’-deoxyflavones however, which includes chrysin and its derivatives, have been demonstrated to be mostly specific to the *Scutellaria* genus (Q. Zhao, Zhang, et al., 2016). Only several species of plants outside of the *Scutellaria* genus have been found to be capable of synthesizing this class of flavones (Kato et al., 1992; V. M. Rao et al., 2009; Y. K. Rao et al., 2002). The specific biological activities of many flavones from both classes has been studied (Lin et al., 2012; Salehi et al., 2019; Shang et al., 2010). In addition, the accumulation patterns of these flavones have been characterized for several more well-known *Scutellaria* species, such as the previously mentioned *S. barbata*, *S. lateriflora*, and *S. baicalensis* (I. Cole et al., 2008; L. Wang et al., 2020).

In *S. baicalensis*, an organ-specific accumulation pattern can be observed, where 4’-hydroxyflavones accumulate in the aerial parts at higher concentrations than in the roots, and 4’-deoxyflavones accumulate in the roots at higher concentrations than in the aerial parts. The biosynthetic pathway responsible for this organ-specific accumulation of flavones has been largely elucidated (Q. Zhao, Zhang, et al., 2016). Specifically, after the formation of cinnamic acid, the pathway diverges into a 4’-hydroxyflavone pathway, and a 4’-deoxyflavones pathway. In the aerial parts of the plant, a cinnamate 4-hydroxylase uses cinnamic acid to form 4-coumarate, which is then used to synthesize 4’-hydroxyflavones. However, in the roots, a cinnamate-CoA ligase forms cinnamoyl-CoA, which is then used to synthesize 4’-deoxyflavones (Q. Zhao, Zhang, et al., 2016). In addition to being well-characterized biochemically, a reference genome for *S. baicalensis* was published in 2019 (Q. Zhao et al., 2019).

Although many studies have focused on the biochemistry of *S. baicalensis*, the *Scutellaria* genus contains more than 470 species, distributed nearly worldwide (Yoonkyung & Kim, 2020). Because metabolite profiles for most of these species are not well-studied, it is unknown how well conserved the flavonoid biosynthetic pathway is in the genus. It’s also unknown how well the flavonoid organ-specific accumulation pattern observed in *S. baicalensis* is conserved. Thus, the limited number of *Scutellaria* species that have been chemically analyzed leaves the possibility of an uncharacterized species having a medicinally significant flavonoid accumulation profile. In this case, if the species is determined to be diploid, it should be a strong candidate for biotechnology efforts for the mass production of medicinal flavonoids.

Therefore, to better characterize flavonoid diversity in *Scutellaria*, we performed a phylogenetic-based analysis of 14 flavonoids, and one other metabolite across 76 species of *Scutellaria*. A subset of 13 species was then selected for additional chemical and genome size analysis. With the results from this method, we identified several species which could serve as targets for biotechnology improvement.

**MATERIALS AND METHODS**

***Chloroplast genome sequencing and phylogenetic tree construction –***

Ask Yoonkyung

***Plant materials and growth conditions for fresh tissue samples –***

Viable seeds were obtained for *S. altissima*, *S. arenicola*, *S. baicalensis*, *S. leonardii*, and *S. racemosa*. Seeds for all these species, except *S. racemosa*, were ordered from retailers. *S. racemosa* mature plants were taken from a field in Hattiesburg, Mississippi, USA and further grown in indoor conditions at the University of Florida before harvesting seeds. To improve germination rate, seeds were first incubated in a 100 µM solution of gibberellic acid for 1 hour with gentle shaking. Seeds were then planted on moist, Sungro© Propagation Mix soil, and watered every 5-8 days. Plants were grown at the University of Florida (Florida, USA) in indoor, climate-controlled conditions at 21-23°C under fluorescent lighting with a light intensity of 140 mE m-2s-1 in a 16 hour light/ 8 hour dark photoperiod. Various tissue samples were taken at six weeks after germination.

The species of *S. insignis*, *S. indica var. coccinea*, *S. barbata*, *S. strigillosa*, *S. dependens*, and *S. pekinensis var. alpina* were collected from the field in Korea and transplanted in the greenhouses of the Sungshin Women’s University, Korea. grown at Sungshin Women’s University (Seoul, Korea). Fresh tissue samples from mature plants were then frozen, and shipped to the University of Florida for metabolite analysis.

Tissue samples for *S. wrightii* and *S. suffrutescens* were taken from fully mature plants having open flowers obtained from Far South Wholesale Nursery in Austin, Texas, USA, where they were cultivated in outdoor greenhouse conditions. Approximately one week before tissue samples were taken, the plants were moved to indoor, climate-controlled conditions as stated previously. Voucher specimens of fresh plants that were used in this study were deposited in the herbariums (Table S1). Herbarium samples were obtained from the FLAS herbarium and the NY herbarium. Voucher information of herbarium samples used in this study is provided in Table S1.

***Metabolite extraction and quantification –***

*Herbarium tissue samples –*

14 flavonoids and one other metabolite were extracted and quantified from the aerial tissues of herbarium vouchers of 67 unique *Scutellaria* species. The flavonoids quantified included 6 4’-hydroxyflavones, which were apigenin, apigenin-7-glucuronide (apigeninG), scutellarein, scutellarin, hispidulin, and hispidulin-7-glucuronide (hispidulinG). The remaining 8 flavonoids were 4’-deoxyflavones, which were chrysin, chrysin-7-glucuronide (chrysinG), baicalein, baicalin, oroxylin A, oroxyloside, wogonin, and wogonoside. Acteoside was the final metabolite quantified. Tissue samples were first weighed with an analytical balance to determine their dry weight. An extraction buffer of 80% High Performance Liquid Chromatography (HPLC) grade methanol added to each so that the following ratio was achieved: 10 mg dry tissue/1 mL solvent. Samples were sonicated for 1 hour at room temperature. The extraction buffer from each sample was withdrawn and further diluted by adding 80% HPLC grade methanol so that the following ratio was achieved: 1 mg tissue/1 mL solvent. To remove any remaining tissue particles, the diluted extraction buffer was centrifuged at 15,000 rpm for 5 minutes and ran through a syringe filter with a pore size of 0.45 µM. Metabolite concentrations in this final solution were quantified with a Thermo Scientific UltiMate 3000 HPLC system. Metabolites were separated with an 3 x 100 mm Acclaim RSLC 120 C18 column, and eluted by a mixture of 0.1% formic acid (A) and 100% acetonitrile (B) with the following gradient: -8 to 0 min, 5% B; 2 min, 25% B; 2 to 6 min, 25% B; 9 min, 50% B; 9 to 11 min, 50% B; 15 min, 95% B; and 15 to 23 min, 95% B. A flowrate of 0.5 mL/min was used and the column oven temperature set to 40°C. Calibration mixes of 0.1, 0.5, 1, 5, 10, 25, 50, and 100 ppm were used to convert peak areas to concentrations in ppm. Preparation of calibration mixes + source of metabolites. With the molecular weight of each metabolite, concentrations in µmol/g dry weight were then calculated. To roughly account for the difference in water weight between herbarium and fresh tissues so that comparisons could be made, all measured metabolite concentrations for herbarium tissue samples were divided by 10. Thus, all final concentrations for dried tissue samples are expressed in units of µmol/0.1 g dry weight, which is approximately equivalent to µmol/g fresh weight.

*Fresh tissue samples –*

The same set of 15 metabolites were extracted and quantified from the root, stem, and leaf tissues of 13 *Scutellaria* species. Tissue samples for each organ were collected in triplicate from mature plants. A mixture of young and mature tissues were selected for each sample so that the average chemical state of the entire plant was represented. Root tissue samples were gently washed to remove soil. The fresh weight of all tissue samples was determined with an analytical balance immediately after harvesting. For the species whose tissue samples were frozen and shipped, fresh weight was measured before freezing. The metabolite extraction and quantification method closely follows that used for the herbarium samples. However, the initial extraction buffer was added to achieve a ratio of 30 mg tissue/1 mL solvent instead of 10 mg tissue/1 mL solvent, and then after sonication, diluted to achieve a ratio of 5 mg tissue/1 mL solvent instead of 1 mg tissue/1 mL solvent. Measured metabolite concentrations were then divided by 5, and the molecular weight of each metabolite was used to calculate final concentrations in units of µmol/g fresh weight.

***Phylogeny-metabolite comparison by multiple correspondence analysis –***

Comparisons between phylogenetic clade and metabolite profile were made using a multiple correspondence analysis-based approach. This analysis used HPLC data for the 15 metabolites extracted from aerial tissue samples of 76 *Scutellaria* species. 67 of these tissue samples were herbarium vouchers, and the remaining 9 were fresh tissue. To reduce bias resulting from mixing data from herbarium and fresh tissue samples, all metabolite concentrations were converted into a binary format where only the presence or absence of a metabolite was considered. The multiple correspondence analysis was conducted in R with the “MCA” function from the FactoMineR package with all settings left at their default values (ver. 2.3). 80% confidence ellipses were drawn from the covariance matrix calculated with the “covMcd” function from the robustbase package (ver. 0.93-6).

***Estimation of genome size with flow cytometry –***

Seeds of *Solanum lycopersicum L*. ‘Stupické polní rané’ (2C = 1.96 pg), and *Glycine max Merr.* ‘Polanka’ (2C = 2.50 pg) were requested from Dr. Jaroslav Doležel (Institute of Experimental Botany, Olomouc, Czech Republic) who suggested them as size-standards for flow cytometry (Doležel et al., 2007). Leaves of these plants were used as size standards for genome-size estimation. In each estimation, a standard sample was selected based on the previously reported genome size information in *Scutellaria* (Lee & Kim, 2017).

The genome sizes of 8 species of *Scutellaria* were estimated using flow cytometry as described in Doležel et al. (2007). Fresh leaves from a standard plant and a sample for estimation (each ca. 5 mm²) were co-chopped using a razor blade in a petri dish with DAPI Prep DNA Staining Solution (SONY, Biotechnology Inc., USA) and incubated two minutes for staining. Cell sorter SH800 (SONY, Biotechnology Inc., USA) was used to measure the fluorescence of the stained cells and ensure each sample measured more than 5,000 particles. Each analysis was repeated three or more times using different leaves or different individuals, and their average and standard deviation were calculated. The 2C-value was estimated based on the relative counts between G1 (growth 1 stage on the cell division) peak from a standard plant and that from a sample for the estimation. The genome size (bp) was estimated based on relative rate between 1C-value and number of bases: 1C (pg) DNA = 0.978 × 109 bp (Doležel et al., 2007).

**RESULTS**

***Phylogeny of selected species as revealed by chloroplast genome sequences –(Fig 1)***

***Aerial tissue metabolite diversity and comparison with phylogeny –***

We applied high performance liquid chromatography (HPLC) to analyze the concentrations of 14 flavonoids and 1 other metabolite from aerial tissue samples of 76 species of *Scutellaria* (Fig 2A). 51 of the 76 species were included in the phylogenetic tree of Fig 1. 67 of the tissue samples were from herbarium vouchers with one replicate, and the remaining 9 from fresh tissue sampled in triplicate. To roughly account for the difference in water weight between herbarium and fresh tissues, the fresh weight of herbarium tissues was estimated by multiplying their dry weight by a factor of 10. Thus, concentrations for dried tissue samples are expressed in units of µmol/0.1 g dry weight, which is approximately equivalent to µmol/g fresh weight. Aerial tissues were chosen for this analysis because they were more readily available from herbariums than root tissues, and because preliminary results indicated that, between species, aerial metabolite profiles were much more diverse than root metabolite profiles. More diversity in metabolite profile would facilitate the metabolite-phylogeny comparison method we planned to apply.

The most commonly occurring metabolite was chrysinG, which we detected in 50 of the 76 species. As its non-glycosylated form, chrysin, is a precursor for all 4’-deoxyflavones we quantified, this supports the role of chrysin as an important metabolic intersection in *Scutellaria*. Interestingly, we detected chrysin in only 31 species. This rarity in detection as compared to its glycosylated form likely reflects the lower stability the non-glycosylated form, and could possibly indicate that in most species, chrysin, once synthesized, is very quickly converted into other downstream 4’-deoxyflavones. A similar observation can be made regarding apigenin, a precursor for all 4’-hydroxyflavones we quantified. We detected apigenin in 23 species, but its glycosylated form, apigeninG, in 43 species. The rarest flavonoid we quantified was hispidulinG, which was detected in only one species, *S. coerulea*.

In general, we detected the glycosylated form of each metabolite (chrysinG, wogonoside, apigeninG, baicalin, scutellarin, and oroxyloside) more often and in greater amounts, than its non-glycosylated counterpart (chrysin, wogonin, apigenin, baicalein, scutellarein, and oroxylinA). This finding reinforces that the glycosylated form is a more stable, storage form. The exception was hispidulinG and hispidulin, as hispidulin was detected in 27 species, but hispidulinG in only one. Out of the five most commonly occurring metabolites, four were 4’-deoxyflavones (chrysinG, wogonoside, baicalin, and wognonin), and only one was a 4’-hydroxyflavone (apigeninG). This result suggests that the 4’-deoxyflavone biosynthetic pathway is more well conserved across the *Scutellaria* genus than the 4’-hydroxyflavone pathway. However, it is also possible that the 4’-deoxyflavones we quantified are more stable than the 4’-hydroxyflavones, and as a result, were less likely to be degraded over time in the herbarium vouchers.

As this data was generated from aerial tissue samples, our finding of 4’-deoxyflavones in all but 11 species suggests that the root-specificity of the 4’-deoxyflavone pathway proposed for *S. baicalensis* may not be well-conserved across the genus. However, we were able to detect low concentrations of several 4’-deoxyflavones, including chrysin, baicalein, wogonin, and their glycosylated forms, in our *S. baicalensis* tissue sample. This suggests that transport of 4’-deoxyflavones from the roots to the aerial tissues, is possible.

Although the species in Fig 2A are ordered based on the phylogenetic tree shown in Fig 1, it is difficult to detect any clear relationship between phylogeny and metabolite profile. Therefore, to make any relationship more apparent, we used a multiple correspondence analysis (MCA) approach. MCA is a technique similar to principal component analysis (PCA), which seeks to summarize a multivariate dataset into only several variables. These variables, called principal components, are calculated to retain the maximum amount of variance possible that is present in the original dataset. MCA was chosen over PCA because it is more appropriate for data in a binary format, which the metabolite dataset was converted into prior to the analysis. This conversion to a binary format was completed by assigning concentration datapoint with a value of TRUE if the metabolite was detected in that species, and FALSE if it was not. A binary format was chosen for this use in this analysis to minimize the effects resulting from the varying ages of the herbarium samples, as well as our use of fresh tissue samples for several species.

No significant separation of species by clade was observed in the MCA plot, indicating that metabolite profile cannot be used to definitively determine the phylogenetic relationship of those species we included (Fig 2B). This is supported by the relatively low percentage of variance in the original dataset that was able to be explained by the first two principal components (21.92% and 13.80%). The small number of species included in clade 5 limits detection of any patterns in metabolite accumulation. Confidence ellipses for both clades 2 and 4 were very large, reflecting the diversity in aerial metabolite profiles for the species in this clade. Although no conclusions besides these can be drawn from the confidence ellipses of clades 2, 4, and 5, several general patterns in metabolite profile are illustrated by the locations of the confidence ellipses for clades 1 and 3.

Some grouping of species in clade 3 to the right side of the plot can be observed, possibly because of trace accumulation of 4’-hydroxyflavones in many of these species, Globally, apigeninG, hispidulin, scutellarin, and scutellarein were detected in 56.58%, 35.52%, 25.00%, and 21.05% of all species analyzed, respectively, but in 69.23%, 69.23%, 38.46%, and 38.46% of species in clade 3. This increase in detection rate of 4’-hydroxyflavones can be confirmed as a cause for the slight grouping of clade 3 species in Fig 2B by considering the associated variable loading plot (Fig 2C). This plot illustrates the role that each metabolite plays in determining the position of points in Fig 2B. The variable loading plot shows that a positive detection of 4’-hydroxyflavones can drive the movement of species in Fig 2B to the upper right quadrant of the plot area, which is where the 80% confidence ellipse for clade 3 is centered. The variable loading plot was also used to identify the negative detection of baicalin as another causative factor in the grouping of species in clade 3. Negative detection of baicalin is shown by the variable loading plot to be well-represented in the first principal component, and accordingly, plays a significant role in the movement of species to right side of the plot area. Confirming this is our detection of baicalin in only 23.08% of species in clade 3, but globally, in 53.94% of all species analyzed.

A similar method of analysis can be applied to clade 1, which has a confidence ellipse centered towards the bottom of the plot area in Fig 2B. The variable loading plot shows that this corresponds to a negative detection of both 4’-hydroxyflavones and 4’-deoxyflavones. Specifically, the negative detection of apigenin, apigeninG, chrysin, and chrysinG is most well-represented on the negative axis of the second principal component. Accordingly, apigenin, apigeninG, chrysin, and chrysinG are detected in only 14.29%, 14.29%, 14.29%, and 28.57% of species in clade 1, but globally, in 30.26%, 56.57%, 40.78%, and 65.78% of all species analyzed.

Although differentiation of clades by metabolite profile is not possible, the variable loadings calculated in our MCA form several significant patterns that are illustrated in the variable loading plot. First, flavonoids of the same class and detection (i.e. TRUE or FALSE) cluster together in the same quadrant. This indicates that in most species, accumulation of a flavonoid of a given class is positively correlated with accumulation of other flavonoids from that same class. Second, flavonoids of different class are positioned in quadrants that are directly adjacent to one another. This indicates that in most species, accumulation of one class of flavonoid is not correlated with accumulation of flavonoids from the other class. In other words, in the aerial tissues of *Scutellaria* species that we analyzed, 4’-deoxyflavone accumulation was independent of 4’-hydroxyflavone accumulation, and vice versa.

***Organ-specific metabolite diversity –***

From the initial set of 76 species selected for aerial tissue metabolite profiling, we identified 13 species with interesting accumulation patterns to investigate further. Comment on morphology of selected species? (Fig 3). At least one species was selected from 4 of the 5 clades identified in our phylogenetic analysis. Due to limited access to plant material, no species were selected from clade 1. The selected species were grown fresh, and tissue samples taken in triplicate from the roots, stems, and leaves of mature plants. We applied HPLC to quantify concentrations of the same 15 metabolites as in our aerial tissue analysis.

Based on our root-specific metabolite profiling results, the 4’-deoxyflavone pathway appears to be very well conserved across all of the species we selected (Fig 4). We detected at least two 4’-deoxyflaonves in all of the species we analyzed, and at least four in all but one species, *S. insignis*. *S. insignis* accumulated only wogonin and wogonoside at relatively low concentrations as compared to those in other species.

Interestingly, although chrysin is proposed to serve as a precursor for all the 4’-deoxyflavones we quantified, we only detected chrysin in the root tissues of two species, and its glycosylated form, chrysinG, in seven. This rarity in detection is possibly a result of chrysin rapidly being converted into downstream products before it accumulates to a level detectable by our HPLC method.

The absence of 4’-hydroxyflavones in the root tissues of all but two species (*S. arenicola* and *S. leonardii*) indicates the specificity of the 4’-hydroxyflavone biosynthetic pathway to the aerial tissues of the plant, in line with that described for *S. baicalensis* (Q. Zhao, Zhang, et al., 2016). However, *S. arenicola,* one of the two species we found to contain 4’-hydroxyflavones in their root tissues, accumulated a greater amount of scutellarin in its roots than in its leaves or stems. *S. wrightii* is notable because of its large accumulation of baicalein and baicalin, many times that of *S. baicalensis*. We also found *S. havanensis* to accumulate greater concentrations of baicalein and baicalin than *S. baicalensis*.

Although root-specific metabolite profiles were relatively consistent across the 14 species we selected, aerial tissue-specific profiles were much more varied. There are two species in which we did not detect any 4’-hydroxyflavones in our leaf samples of (*S. strigillosa* and *S. pekinensis* var. *alpina*). In other species, the pathway seems to be conserved to greatly varying degrees. In our leaf tissue samples, we were able to detect apigenin, which is a precursor for all other 4’-hydroxyflavones analyzed, or its glycoside, apigeninG, in ten out of the 12 species in which we detected 4’-hydroxyflavones. Two species (*S. havanensis* and *S. dependenens*) did not accumulate any 4’-hydroxyflavones downstream from apigenin. Out of the ten species in which we detected apigenin or its glycoside, we were able to detect scutellarein or its glycoside, scutellarin, in 7. Scutellarein is immediately downstream of apigenin in the 4’-deoxyflavone pathway, and is also a precursor for the final set of 4’-deoxyflavones we analyzed, hispidulin and its glycoside, hispidulinG. We identified only two species (*S. arenicola* and *S. leonardii*) that accumulated apigenin, scutellarein, and hispidulin, or their glycosylated forms. There were three species (*S. insignis*, *S. racemosa*, and *S. writghtii*) which accumulated hispidulin or hispidulinG without scutellarein or scutellarin. Of these three, we found that *S. insignis* and *S. racemosa* did not accumulate even apigenin or apigeninG.

HispidulinG was exceptionally rare, as we detected it in only two species, *S. racemosa* and *S. arenicola*. Only a small amount accumulated in the stems of *S. racemosa*. However, a significant amount accumulated in the leaves of *S. arenicola.*

Our detection of at least one 4’-deoxyflavone in the leaves of 11 out of the 14 species we selected indicates that 4’-deoxyflavones are not root-specific to the same degree that 4’-hydroxyflavones are aerial tissue-specific. One explanation is provided by considering the flavonoid profiles of *S. wrightii*, S*. baicalensis*, *S. altissima*, and *S. tournefortii*. We detected higher concentrations of many of the same 4’-deoxyflavones in the roots of these species as compared to the leaves. A mix of 4’-hydroxyflavones from the leaves, and 4’-deoxyflavones from the roots, can also be observed in the stems. Therefore, in these species, it’s likely that some amount of 4’-deoxyflavones being synthesized in the roots are being transported to the aerial parts. It’s also possible that biosynthetic enzymes acting in the 4’-deoxyflavone pathway of these species are more active in the roots, but are also active at a much lower level in the leaves. Chrysin and chrysinG are the only 4’-deoxyflavones which don’t seem to follow this pattern of accumulation due to their scarcity in the root tissues.

There are several species which do not follow this root-heavy accumulation pattern for 4’-deoxyflavones. We found *S. racemosa* to accumulate oroxylinA and oroxyloside at significantly greater concentrations in its leaves as compared to in its roots. *S. strigillosa* is also a notable exception due to its high accumulation of baicalein in its leaves relative to its roots. Finally, we found *S. dependens* to accumulate higher concentrations of wogonin and wogonoside in its leaves are compared to its roots. In these species, it is likely that at least some biosynthetic enzymes in the 4’-deoxyflavone pathway are primarily active in the aerial parts of the plant, instead of in the roots.

In general, acteoside does not seem to share the same organ-specificity as the flavonoids we analyzed. We detected acteoside in the roots of eight species, but also in the stems or leaves of six out of these eight species. In these species that accumulated acteoside in both their roots and aerial parts, concentrations were very similar. Interestingly, we found *S. havanensis* to accumulate a significant amount of acteoside in its leaves, but none in its roots.

Several general patterns of accumulation become apparent by considering the phylogeny of the 14 species we selected, as was shown in Fig 1. First, three species within clade 2 (*S. insignis, S. indica var. coccinea*, and *S. barbata*) display a very similar flavonoid profile. The root specificity of 4’-deoxyflavones, and aerial tissue specificity of 4’-hydroxyflavones is very well conserved in these species. Additionally, the concentrations of accumulated flavonoids are similar, and relatively low as compared to other species that we analyzed. In line with this observed similarity in accumulation profile, *S. insignis, S. indica var. coccinea*, and *S. barbata* are positioned nearby each other in our phylogenetic tree.

In contrast to this similarity, the other three species we selected from clade 2 (*S. racemosa*, *S. strigillosa*, and *S. dependens*) display very diverse accumulation patterns, especially in their aerial tissues. The root specificity of 4’-deoxyflavones is not well conserved in these species. Not only did we detect multiple 4’-deoxyflavones in leaf and stem tissues, but we also found that each species accumulates significantly higher concentrations of several 4’-deoxyflavones in its aerial tissues as compared to its roots. Interestingly, a different 4’-deoxyflavone was accumulated in significant amounts in the aerial tissues of each of the three species.

Another pattern in metabolite profile can be observed between those species in clades 4 and 5. *S. baicalensis*, the only species which we selected from clade 4, displays an organ-specific accumulation pattern that is closely matched by *S. tournefortii* and *S. altissima* in clade 5. These three species accumulate a similar set of 4’-deoxyflavones in their roots. However, the high accumulation of baicalin by *S. baicalensis* sets it apart from *S. tournefortii* and *S. altissima*. Aerial tissue accumulation for all three species was very similar. We detected limited concentrations of only several 4’-hydroxyflavones in our aerial tissue samples, and higher concentrations of 4’-deoxyflavones that also accumulated in the roots of each species.

***Genome size estimations –***

Genome sizes for 10 species of *Scutellaria* were estimated with flow cytometry by comparison of 2C-values to a standard plant. We were not able to find any previously published genome size data for seven of these species included in our analysis. To facilitate comparison, we also collected genome size and chromosome number data for several species from literature (Table 1). Unfortunately, we did not include *S. wrightii* in our analysis, and no previous publications have reported its genome size.

Overall, genome sizes for the species we selected were similar at around 0.40 Giga base pairs (Gbp). One notable exception was *S. arenicola*, which we found to have a genome size approximately double that of all other species analyzed. It’s likely that this is a result of a genome duplication event, and *S. arenicola* is a tetraploid organism. We estimated a slightly larger genome size as compared to previously published data for several species, including *S. barbata*, *S. racemosa*, *and S. baicalensis*. This is possibly a result of ... We also collected chromosome number counts from literature to assess how well genome structure is conserved across the 14 species we selected. *S. racemosa* and *S. baicalensis* have both been reported to have a chromosome number of 18. However, *S. barbata* has been reported as having a chromosome number of 26, despite having a similarly sized genome to *S. baicalensis* and *S. racemosa*.

**DISCUSSION**

Include a summary of conclusions and a take-home message for the generally informed reader in the DISCUSSION.

Differences with *S. baicalensis* biosynthetic pathway and pattern of accumulation

4’-deoxyflavones are more well-conserved – serve important core roles in plant growth/signaling

4’-hydroxyflavones are less well-conserved – serve accessory roles, specific to environmental conditions (e.g. attracting specific pollinators, pest or pathogen defense)

Therefore, the 4’-hydroxyflavone pathway seems to slightly more well-conserved in clade 3 species as compared to other non-clade 3species.

Phylogeny-metabolite-genome size relationship

Chrysin not detectable in roots despite being a precursor for all 4’-deoxyflavones – possible metabolon? - Interestingly in *S. altissima* and *S. tournefortii*, we detected much greater concentrations of chrysin and chrysinG in the leaves of the plants than we did in the roots. We observed a similar pattern in several other species as well, including *S. baicalensis*, *S. havanensis*, *S. leonardii*, and *S. pekenensis var. alpina*. Possible explanations – downstream 4’-deoxyflavone biosynthetic enzymes are not active in the aerial parts of these species. Thus, instead of being quickly converted into downstream products as occurs in the roots, chrysin is able to accumulate to much higher concentrations.

*S. arenicola* has hispidulin forming enzymes in aerial parts have high affinity for scutellarein, therefore, not much is able to accumulate

OroxylinA and oroxyloside accumulation in aerial parts of *S. racemosa* (Fig 5). Other species recommendations for biotechnology development – *S. wrightii* (baicalin in roots), *S. depenedens* (wogonin + wogonoside in aerial parts), and *S. arenicola* (hispudulin + hispidulinG in aerial parts).

**Acknowledgments**

**Author contributions**

**Data availability statement**

Github? - need to clean up

Chloroplast genome sequences?

**Literature Cited**

Awad, R., Arnason, J. T., Trudeau, V., Bergeron, C., Budzinski, J. W., Foster, B. C., & Merali, Z. (2003). Phytochemical and biological analysis of Skullcap (Scutellaria lateriflora L.): A medicinal plant with anxiolytic properties. *Phytomedicine*, *10*(8), 640–649. https://doi.org/10.1078/0944-7113-00374

Cole, I. B., Saxena, P. K., & Murch, S. J. (2007). Medicinal biotechnology in the genus scutellaria. *In Vitro Cellular & Developmental Biology - Plant*, *43*(4), 318–327. https://doi.org/10.1007/s11627-007-9055-4

Cole, I., Cao, J., Alan, A., Saxena, P., & Murch, S. (2008). Comparisons of Scutellaria baicalensis, Scutellaria lateriflora and Scutellaria racemosa: Genome Size, Antioxidant Potential and Phytochemistry. *Planta Medica*, *74*(4), 474–481. https://doi.org/10.1055/s-2008-1034358

Doležel, J., Greilhuber, J., & Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols*, *2*(9), 2233–2244. https://doi.org/10.1038/nprot.2007.310

Karimov, A. M., & Botirov, E. Kh. (2017). Structural Diversity and State of Knowledge of Flavonoids of the Scutellaria L. Genus. *Russian Journal of Bioorganic Chemistry*, *43*(7), 691–711. https://doi.org/10.1134/S1068162017070068

Kato, M. J., Yoshida, M., & Gottlieb, O. R. (1992). Flavones and lignans in flowers, fruits and seedlings of Virola venosa. *Phytochemistry*, *31*(1), 283–287. https://doi.org/10.1016/0031-9422(91)83055-P

Lee, Y., & Kim, S. (2017). Genome size of 15 Lamiaceae taxa in Korea. *Korean Journal of Plant Taxonomy*, *47*(2), 161–169. https://doi.org/10.11110/kjpt.2017.47.2.161

Lin, T.-Y., Lu, C.-W., Wang, C.-C., Lu, J.-F., & Wang, S.-J. (2012). Hispidulin inhibits the release of glutamate in rat cerebrocortical nerve terminals. *Toxicology and Applied Pharmacology*, *263*(2), 233–243. https://doi.org/10.1016/j.taap.2012.06.015

Rao, V. M., Damu, G. L. V., Sudhakar, D., & Rao, C. V. (2009). Two New Bio-active Flavones from Grangea maderaspatana (Artemisia maderaspatana). *Asian J. Chem.*, *21*(2), 7.

Rao, Y. K., Reddy, M. V. B., Rao, C. V., Gunasekar, D., Blond, A., Caux, C., & Bodo, B. (2002). Two new 5-deoxyflavones from Albizia odoratissima. *Chemical & Pharmaceutical Bulletin*, *50*(9), 1271–1272. https://doi.org/10.1248/cpb.50.1271

Salehi, B., Venditti, A., Sharifi-Rad, M., Kręgiel, D., Sharifi-Rad, J., Durazzo, A., Lucarini, M., Santini, A., Souto, E. B., Novellino, E., Antolak, H., Azzini, E., Setzer, W. N., & Martins, N. (2019). The Therapeutic Potential of Apigenin. *International Journal of Molecular Sciences*, *20*(6). https://doi.org/10.3390/ijms20061305

Shang, X., He, X., He, X., Li, M., Zhang, R., Fan, P., Zhang, Q., & Jia, Z. (2010). The genus Scutellaria an ethnopharmacological and phytochemical review. *Journal of Ethnopharmacology*, *128*(2), 279–313. https://doi.org/10.1016/j.jep.2010.01.006

Tao, Y., Zhan, S., Wang, Y., Zhou, G., Liang, H., Chen, X., & Shen, H. (2018). Baicalin, the major component of traditional Chinese medicine Scutellaria baicalensis induces colon cancer cell apoptosis through inhibition of oncomiRNAs. *Scientific Reports*, *8*(1), 14477. https://doi.org/10.1038/s41598-018-32734-2

Venkatarame Gowda Saralamma, V., Lee, H. J., Hong, G. E., Park, H. S., Yumnam, S., Raha, S., Lee, W. S., Kim, E. H., Sung, N. J., Lee, S. J., Heo, J. D., & Kim, G. S. (2017). Korean Scutellaria baicalensis Georgi flavonoid extract induces mitochondrially mediated apoptosis in human gastric cancer AGS cells. *Oncology Letters*, *14*(1), 607–614. https://doi.org/10.3892/ol.2017.6184

Wang, L., Chen, W., Li, M., Zhang, F., Chen, K., & Chen, W. (2020). A review of the ethnopharmacology, phytochemistry, pharmacology, and quality control of Scutellaria barbata D. Don. *Journal of Ethnopharmacology*, *254*, 112260. https://doi.org/10.1016/j.jep.2019.112260

Wang, Q., Acharya, N., Liu, Z., Zhou, X., Cromie, M., Zhu, J., & Gao, W. (2018). Enhanced anticancer effects of Scutellaria barbata D. Don in combination with traditional Chinese medicine components on non-small cell lung cancer cells. *Journal of Ethnopharmacology*, *217*, 140–151. https://doi.org/10.1016/j.jep.2018.02.020

Yang, L., Yang, C., Li, C., Zhao, Q., Liu, L., Fang, X., & Chen, X.-Y. (2016). Recent advances in biosynthesis of bioactive compounds in traditional Chinese medicinal plants. *Science Bulletin*, *61*(1), 3–17. https://doi.org/10.1007/s11434-015-0929-2

Zhang, Z., Lian, X., Li, S., & Stringer, J. L. (2009). Characterization of chemical ingredients and anticonvulsant activity of American skullcap (Scutellaria lateriflora). *Phytomedicine*, *16*(5), 485–493. https://doi.org/10.1016/j.phymed.2008.07.011

Zhao, Q., Chen, X.-Y., & Martin, C. (2016). Scutellaria baicalensis, the golden herb from the garden of Chinese medicinal plants. *Science Bulletin*, *61*(18), 1391–1398. https://doi.org/10.1007/s11434-016-1136-5

Zhao, Q., Yang, J., Cui, M.-Y., Liu, J., Fang, Y., Yan, M., Qiu, W., Shang, H., Xu, Z., Yidiresi, R., Weng, J.-K., Pluskal, T., Vigouroux, M., Steuernagel, B., Wei, Y., Yang, L., Hu, Y., Chen, X.-Y., & Martin, C. (2019). The Reference Genome Sequence of Scutellaria baicalensis Provides Insights into the Evolution of Wogonin Biosynthesis. *Molecular Plant*, *12*(7), 935–950. https://doi.org/10.1016/j.molp.2019.04.002

Zhao, Q., Zhang, Y., Wang, G., Hill, L., Weng, J.-K., Chen, X.-Y., Xue, H., & Martin, C. (2016). A specialized flavone biosynthetic pathway has evolved in the medicinal plant, Scutellaria baicalensis. *Science Advances*, *2*(4), e1501780. https://doi.org/10.1126/sciadv.1501780

Zhao, T., Tang, H., Xie, L., Zheng, Y., Ma, Z., Sun, Q., & Li, X. (2019). Scutellaria baicalensis Georgi. (Lamiaceae): A review of its traditional uses, botany, phytochemistry, pharmacology and toxicology. *Journal of Pharmacy and Pharmacology*, *71*(9), 1353–1369. https://doi.org/10.1111/jphp.13129

Zhu, D., Wang, S., Lawless, J., He, J., & Zheng, Z. (2016). Dose Dependent Dual Effect of Baicalin and Herb Huang Qin Extract on Angiogenesis. *PLOS ONE*, *11*(11), e0167125. https://doi.org/10.1371/journal.pone.0167125

**Tables**

Table 1: Genome sizes as estimated by flow cytometry, and a summary of previously published genome sizes and chromosome numbers.

**Appendices**

Table S1: Table w/ voucher info for all herbarium samples

Table S2: .csv of all flavonoid profiling results

**Figure Legends**

**Figure 1.** Maximum likelihood phylogenetic tree inferred from 3 chloroplast genome regions for 51 species of *Scutellaria* and 1 outgroup*.* To facilitate downstream analysis, the tree is subdivided into 5 color-coded clades based on grouping of species within the tree.

**Figure 2.** Aerial metabolite concentrations measured with High Performance Liquid Chromatography (HPLC) compared to phylogeny for 76 species of *Scutellaria*. (A) Heatmap of collected data. Data is shown in units of µmol/g fresh weight. Fresh weight of herbarium samples was estimated by multiplying their dry weight by a factor of 10. Species for which fresh tissue was used are indicated by an asterisk on the right side of the heatmap. Samples for all other species were prepared from herbarium vouchers. Colored circles next to species names indicate phylogenetic clade, as shown in Figure 1. An empty circle indicates that the species was not included in the tree from Figure 1. (B) Multiple Correspondance Analysis (MCA) individual results with overlaid 80% confidence ellipses generated from binarized metabolite data. Each colored circle represent a species, and the color of the circle represents phylogenentic clade, as shown in Figure 1. The percentage of total variance explained by each principal component is shown next to each axis title. (C) Variable loadings from MCA. Each variable represents the prescence (indicated with “TRUE) or absence (indicated with “FALSE) of a metabolite, and are color-coded according to metabolite class.

**Figure 3.** Representative images of 14 species of *Scutellaria* selected for additional metabolite and genome size profiling. Scale bar in bottom left of images represents a length of 5 cm. Images without a scale bar did not have a ruler included in the image. Colored circles next to species names indicate the clade which the species belongs to, as indicated in Figure 1.

**Figure 4.** Organ-specific metabolite data collected from 14 *Scutellaria* species via High Performance Liquid Chromatography (HPLC). Data is shown in units of µmol/g fresh weight. Species on x-axis are ordered based on phylogenetic relationship determined from chloroplast genome data, and colored circles next to species names indicate phylogenetic clade, as shown in Figure 1. An empty circle indicates that the species was not included in the tree from Figure 1.

**Figure 5.** Organ-specific oroxylinA (bottom) and oroxyloside (top) concentrations in 14 *Scutellaria* species, as determined via High Performance Liquid Chromatography (HPLC). Data is shown in units of µmol/g fresh weight. Concentrations were averaged from tissue samples taken from 3 biological replicates, and error bars represent standard error. Species are ordered based on phylogenetic relationship determined from chloroplast genome data, and colored circles next to species names indicate phylogenetic clade, as shown in Figure 1. An empty circle indicates that the species was not included in the tree from Figure 1.

**Table 1.** Genome size and chromosome number data for 14 *Scutellaria* species. Genome sizes for 10 species were measured in this study using flow cytometry. Other genome size, and all chromosome number data was collected from literature. *S. wrightii* was not included in the flow cytometry procedure, and neither genome size nor chromosome data has been published. Cell colors indicate phylogenetic clade, as shown in Figure 1.