**Title page**

Footnotes

Abstract

Key words

**INTRODUCTION**

In the first paragraph of the introduction, include the theoretical or conceptual basis for your work in a context accessible to the diverse botanical readership that AJB attracts.

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. Biotechnology as mass production system for medicines: <https://link.springer.com/article/10.1007/s11627-007-9055-4>

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 (Shang et al., 2010; 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 phytochemicals which these *Scutellaria* species accumulate in high concentrations (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 the majority of both these classes of flavones have 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* (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 of the plant, and 4’-deoxyflavones in the roots. The biosynthetic pathway responsible for this organ-specific accumulation of flavones has been largely elucidated (Q. Zhao, Zhang, et al., 2016). In addition, a reference genome for *S. baicalensis* was published in 2019 (Q. Zhao et al., 2019).

Although the phytochemical profile of *S. baicalensis* has been well-characterized, the *Scutellaria* genus contains more than 470 species, distributed nearly worldwide (Yoonkyung & Kim, 2020). Because phytochemical profiles for most of these species are not well-studied, the stability of the flavonoid biosynthetic pathway across the genus is largely unknown. To better characterize flavonoid diversity in *Scutellaria*, we performed a phylogenetic-based survey of 14 flavonoids in 75 species of *Scutellaria*. One phenol was also quantified. A subset of 13 species was then selected for additional chemical and genome size analysis. Our results indicate a relationship between phylogeny and phytochemical profile, and we identified several species which could serve as targets for biotechnology improvement.

**MATERIALS AND METHODS**

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

Ask Yoonkyung

***Planting and growing conditions for fresh samples –***

Viable seeds were obtained for several species, including *S. altissima*, *S. arenicola*, *S. baicalensis*, *S. barbata*, *S. havanensis*, *S. leonardii*, and *S. racemosa*. To improve germination rate, seeds were incubated in a 100 uM solution of gibberellic acid for 1 hour with gentle shaking. Soil/water/light?

Need KR growing info – collected from field?

***Phytochemical extraction and quantification –***

*Herbarium tissue samples –*

14 flavonoids and one phenol were extracted and quantified from the aerial tissues of herbarium vouchers of 67 unique *Scutellaria* species. The flavonoids quantified were apigenin, apigenin-7-glucuronide (apigeninG), scutellarein, scutellarin, hispidulin, hispidulin-7-glucuronide (hispidulinG), chrysin, chrysin-7-glucuronide (chrysinG), baicalein, baicalin, oroxylin A, oroxyloside, wogonin, and wogonoside. The phenol quantified was acetoside. Tissue samples were weighed with an analytical balance, and an extraction buffer of 80% High Performance Liquid Chromatography (HPLC) grade methanol added to each so that the following ratio was achieved: 10 mg 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 between 0.2 and 0.45 µM. Phytochemical concentrations in this final solution were quantified with a Thermo Scientific UltiMate 3000 HPLC system. Phytochemicals 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 µM were used to convert peak areas to concentrations in µM. Preparation of calibration mixes + source of phytochemicals. So that rough comparisons could be made between herbarium and fresh samples, all measured phytochemical concentrations were divided by 2 to represent a ratio of 0.5 mg tissue / 1 mL solvent.

*Fresh tissue samples –*

The same 15 phytochemicals 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. Root tissue samples were gently washed to remove soil. The phytochemical 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.

***Phylogeny-phytochemical comparison by multiple factor analysis –***

Comparisons between phylogenetic clade and phytochemical profile were made using a multiple factor analysis-based approach. This analysis used HPLC data for the 15 phytochemicals extracted from aerial tissue samples of 75 *Scutellaria* species. 67 of these tissue samples were herbarium vouchers, and the remaining 8 were fresh tissue. To reduce bias resulting from mixing data from herbarium and fresh tissue samples, all phytochemical concentrations were converted into a binary format where only the presence or absence of a phytochemical was considered. The multiple factor analysis was conducted in R with the “MFA” function from the FactoMineR package (ver. 2.3). Following the biosynthetic pathway for *S. baicalensis* described by Zhao et al. (2016), phytochemicals were split into two group – aerial-specific flavonoids, root-specific flavonoids, and phenols. Aerial-specific flavonoids included apigenin, apigeninG, scutellarein, scutellarin, hispidulin, and hispidulinG. Root-specific flavonoids included chrysin, chrysinG, baicalein, baicalin, oroxylin A, oroxyloside, wogonin, and wogonoside. Phenols included acetoside only. 95% 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 –***

Ten species of *Scutellaria* were collected from the field and transplanted in the greenhouses of the Sungshin Women’s University (Seoul, Korea) and the University of Florida (Florida, USA). Voucher specimens were deposited in each herbarium of the university (Table #). We requested seeds of *Solanum lycopersicum L*. ‘Stupické polní rané’ (2C = 1.96 pg), and *Glycine max Merr.* ‘Polanka’ (2C = 2.50 pg) 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 and Kim, 2017).

The genome size of each plant was 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 –***

***Aerial phytochemical diversity and comparison with phylogeny –***

**10.1007/s11306-010-0269-9**

We applied high performance liquid chromatography (HPLC) to analyze the concentrations of 14 flavonoids and 1 phenol from aerial tissue samples of 75 species of *Scutellaria* (Figure 2A). 50 of the 75 species were included in the phylogenetic tree of Figure 1. 67 of the tissue samples were from herbarium vouchers with one replicate, and the remaining eight from fresh tissue. So that rough comparisons could be made between data from herbarium and fresh tissue samples, concentrations for herbarium samples were calculated at 0.5 mg tissue / 1 mL solvent, and concentrations for fresh samples at 5 mg tissue / 1 mL solvent.

The most widepsread phytochemical we measured was chrysinG, which was detected in 49 of the 75 species. The rarest was hispidulinG, which was detected in only one species, *S. coerulea*. In general, we detected the glycosylated form of each phytochemical (chrysinG, wogonoside, apigeninG, baicalin, scutellarin, and oroxyloside) more often than its non-glycosylated counterpart (chrysin, wogonin, apigenin, baicalein, scutellarein, and oroxylinA). The exception was hispidulinG and hispidulin, as hispidulin was detected in 27 species.

Abundance: aerial-specific vs root-specific

General conclusions about profiling results – abundance/nonabundance of flavonoids

General conclusions about flavonoid results (e.g. most common and least common flavonoids, organ-specificity). Grouping of species by flavonoid profile (MFA).

***Organ-specific flavonoid diversity –***

***Genome size estimations –***

**DISCUSSION**

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

**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., 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

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

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

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 #s for all herbarium samples

Table S2: .csv of all flavonoid profiling results

**Figure Legends**

Figure 1: Maximum likelihood phylogenetic tree inferred from 3 chloroplast genomic regions for 51 species of *Scutellaria* and 1 outgroup. To facilitate downstream analysis, the tree was subdivided into 5 color coded clades.

Figure 2: Aerial flavonoid concentrations measured with High Performance Liquid Chromatography (HPLC) for 75 species of *Scutellaria*. (A) Heatmap of collected data. 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. Data for fresh tissue samples is shown at 5000 ppm (5 mg sample / 1 mL solvent). Data for herbarium tissue samples is shown at 500 ppm (0.5 mg sample / 1 mL solvent). (B) Multiple Factor Analysis (MFA) results generated from binary version of flavonoid data. (C) Variable representations from MFA.

Figure 3: Representative images of 13 species of *Scutellaria* selected for additional flavonoid 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.

Figure 4: Organ-specific flavonoid data collected from 13 *Scutellaria* species via High Performance Liquid Chromatography (HPLC). Species are ordered based on phylogenetic relationship determined from chloroplast genome data, shown in Figure 1. 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 phylogenetic tree. Flavonoids are ordered based on proposed flavonoid pathway for *S. baicalensis*.