**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 phytochemicals 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 phytochemical 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 metabolite analysis of 14 flavonoids, and one phenol across 76 species of *Scutellaria*. A subset of 14 species was then selected for additional chemical and genome size analysis. With the results from this method, we identified several species with promising metabolite profiles 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 tissue samples –***

Viable seeds were obtained for several species, including *S. altissima*, *S. arenicola*, *S. baicalensis*, *S. havanensis*, *S. leonardii*, and *S. racemosa*. Seeds for all these species, except *S. racemosa*, were ordered from an online retailer. *S. racemosa* seeds were provided by … 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 73-74°C under fluorescent lighting. Intensity and photoperiod. Tissue samples were taken at # weeks after germination.

The species of *S. insignis*, *S. indica var. coccinea*, *S. barbata*, *S. strigillosa*, *S. dependens*, and *S. pekinensis var. alpina* were grown at Sungshin Women’s University (Seoul, Korea). fresh tissue samples were taken from … Need KR growing info – collected from field? Tissue samples were then frozen, and shipped to the University of Florida for metabolite analysis.

Tissue sample for one species, *S. wrightii*, were taken from plants 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 and grown at 73-74°C under fluorescent lighting. Intensity and photoperiod. The specific age of the plants when tissue samples were taken is not known.

***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 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. The phenol quantified was acetoside. 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. 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 ppm were used to convert peak areas to concentrations in ppm. Concentrations in µM were then calculated using the molecular weight of each flavonoid. Preparation of calibration mixes + source of phytochemicals. So that rough comparisons could be made between herbarium and fresh samples, all measured phytochemical concentrations for herbarium tissue samples were divided by 2. Thus, all final concentrations for dried tissue samples are expressed in units of µmol / 0.5 g dry weight.

*Fresh tissue samples –*

The same set of 15 phytochemicals were extracted and quantified from the root, stem, and leaf tissues of 14 *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 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. Thus, all final concentrations for fresh tissue samples are expressed in units of µmol / 5 g fresh weight.

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

Comparisons between phylogenetic clade and phytochemical profile were made using a multiple correspondence analysis-based approach. This analysis used HPLC data for the 15 phytochemicals extracted from aerial tissue samples of 76 *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 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 for each phylogenetic clade 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 and the University of Florida. Voucher specimens were deposited in each herbarium of the university (Table S1). 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 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).