

ABSTRACT

The Lamiaceae is one of the largest plant families incorporating over 7,000 species. It is home to a variety of ornamental and medicinal plants, but perhaps the most well-known are the culinary herbs, such as basil, oregano, rosemary, and thyme. The often-fragrant leaves of these plants are in part thanks to their terpenoid content. Terpenes are a class of natural compounds that are highly prolific in plants, and specifically the Lamiaceae family. These specialized metabolites can serve a variety of functions in the plant, including communication and defense. Terpenoids have been co-opted for human applications, as many can be used as medicines, pesticides, flavors, and fragrances. Occasionally, specialized metabolism, or specifically terpenoid metabolism, can be physically clustered in the genome. Biosynthetic gene clusters (BGCs) are a set of genes in the genome that are biochemically associated, phylogenetically distinct, and are often transcriptionally linked. One piece of the puzzle making the discovery of BGCs more feasible is the increase in robustness and quality of genome sequencing today. The following dissertation integrates the ideas of terpenoid metabolism and BGCs in the Lamiaceae family. Specifically, I explore the genomes of several Lamiaceae species, including *Callicarpa americana* and *Teucrium chamaedrys*, in search of BGCs and to explore terpenoid metabolism. I discovered a Lamiaceae family-wide gene cluster in *C. americana* that appears to predate the family at around 65 million years old. Additionally, I sequence and analyze the genome of the medicinally relevant plant *T. chamaedrys*, and find not only the same large BGC, but also discover the most putative diTPSs identified in a single species to date. I also functionally characterize several diTPSs and associated Cytochrome P450s. My work presented here aims to lay another brick in the larger building of understanding plant genomes and elucidating plant specialized metabolism.