**Abstract**

My research goal is to investigate the role of SBP transcription factors, particularly SlySBP8b and SlySBP12a, in plant cell death pathways. While plant PCD is not well understood, it has been shown that ectopic expression of the Inhibitor of Apoptosis protein from Spodoptera frugiperda (SfIAP) can suppress PCD characteristics in plants by interacting with members of the SBP transcription factor family. SlySBP8b and SlySBP12a, two SBP proteins that induce cell death when transiently expressed in Nicotiana benthamiana, have been identified as potential regulators of PCD. To expand our understanding beyond Nicotiana benthamiana, which has an unannotated SBP gene family, we plan to use phylogenetic inference to investigate the SBP gene family in other plant species. While some closely related species such as Nicotiana tabacum and Solanum lycopersicum have partial annotations of SBP genes, other model species such as Arabidopsis thaliana have complete annotations with partial experimental validation. Through this analysis, we hope to identify PCD-associated genes and characterize their function in a variety of plant species. This will provide insights into the evolution and diversification of the SBP gene family and their roles in stress responses and plant development.

**Introduction**

Programmed cell death (PCD) is a highly regulated process that is required in all multicellular organisms for roles in development and responses to both abiotic stresses and biotic insults. In plants, while clear evidence of PCD has been characterized, the underlying mechanisms regulating its activity are largely unknown. Investigation into these mechanisms has been focused on comparison to animal apoptosis, a well understood form of PCD, which is characterized by the release of cytochrome C from the mitochondria and subsequent activation of cell-executing caspases. While plant PCD shares similar morphologies to animal apoptosis, orthologs of key apoptotic regulators are absent in plant genomes, complicating research (1).

Despite this lack of conservation, it has been shown that apoptotic regulators from animals can suppress classical PCD characteristics in plants when expressed ectopically. The Inhibitor of Apoptosis protein from Spodoptera frugiperda (SfIAP) confers salt and heat stress tolerance in tobacco, and resistance to necrotrophic pathogens in tomato, both responses demonstrating a cell protective effect (2). Protein-protein interaction studies of the function of SfIAP revealed interactions with members of the plant-specific SQUAMOSA promoter-binding protein (SBP or SPL, depending on species) transcription factor family (3). Of the SBP proteins identified via protein interaction with SfIAP, the two members SlySBP8b and SlySBP12a induced cell death when transiently expressed in Nicotiana benthamiana, indicating their role in PCD (3). Subcellular localization analysis of the two SBPs revealed that while SBP8b is entirely nuclear localized, SlySBP12a expression is split between the nucleus and the ER membrane (3). Sequence analysis of the two proteins revealed a transmembrane domain (TMD) in SlySBP12a that controls localization, with truncated SlySBP12a lacking the TMD being entirely nuclear localized. Comparing the cell death effects of truncated SlySBP12a to wild type, the resulting cell death is attenuated, indicating that localization from the ER membrane is a level of regulation for this gene. The ER membrane has been widely studied as a critical sensor and regulatory hub for stress responses, with many membrane-tethered transcription factors functioning to regulate biotic and abiotic responses and determine cell fate. With the localization of SlySBP12a and the importance of membrane-tethered transcription factors highlighted in stress, it has been hypothesized that SlySBP12a is proteolytically cleaved from the ER membrane in response to stress factors then translocates to the nucleus to activate genes involved in cell death.

The overall goal of my research progress is to characterize role that SBP transcription factors, particularly SlySBP8b and SlySBP12a, have on cell death pathways within plants. SfIAP works through an unknown mechanism to inhibit cell death, and interacts with these transcription factors, so it stands to reason that regulation of SlySBP8b and SlySBP12a is important for pro-survival signaling in planta, and that these transcription factors must have some manner of interaction with PCD regulatory mechanism. Identification of the associating proteins and the genes activated by these transcription factors are of key interest for identifying PCD-associated genes and characterizing their function.

SBP family members (or SPL as they are referred to in other plant species) have been shown to play a diverse role in plant development including juvenile to adult phase transition, trichome development, apical dominance, and pollen sac development (4). Silencing of the SBP-box gene Colorless non-ripening (Cnr) in tomato results in fruit with delayed ripening, a phenotype that is seen in tomatoes overexpressing SfIAP (2, 5). Although much is known about the role of SBP family members in plant development, only a few studies have linked these transcription factors to stress responses. For example, the N immune receptor of *Nicotiana benthamiana* has been shown to associate with NbSPL6 and is required for the activation of HR cell death (6). In response to copper deficiency, Arabidopsis SPL7 activates the transcription of microRNAs and other genes involved in copper uptake (7). Another Arabidopsis SBP family member, AtSPL14, has been implicated in cell death caused by the mycotoxin FB1. The Arabidopsis atspl14 mutant was shown to be insensitive to FB1 (8), another phenotype that we have observed in SfIAP-overexpressing tomato plants (2).

While we have performed a majority of our analyses in N. benthamiana due to its ease of transformation as a model plant, we are interested in expanding this proposed mechanism of cell death regulation we have uncovered to other more economically relevant species. SBP transcription factors are found in all land plants, but the family has undergone extensive duplication and functional divergence, requiring experimental validation to expand hypotheses beyond a single species (9,10). While all SBPs are named for their characteristic DNA binding domain, the sizes and functions of these transcription factors varies between family members and species. For example, Arabidopsis thaliana contains 16 SBP genes, 19 members in tomato, 19 in rice, and 27 in apple.

My analysis will focus on using phylogenetic inference as a method to investigate the SBP gene family of *Nicotiana benthamiana*, which remains unannotated. While closely related species such as *Nicotiana tabacum* and *Solanum lycopersicum* have partial annotation of the SBP family genes, more distant model species such as *Arabidopsis thaliana* have complete annotations with partial experimental validation. To identify closely related SBP family members in *N. benthamiana* and predict function, I collected coding sequences (CDS) of all predicted SBP sequences from *N. benthamiana*, *A. thaliana*, and *S. lycopersicum* and generated a tree to guide investigation of homologous function.

**Methods**

Data Source

Sequences were obtained from the Plant Transcription Factor Database v5.0, an online database from the developers of PlantRegMap. The PlantTFDB has lists of TFs with detailed annotation based on literature with links to external databases for further study. PlantTFDB also contains predicted TFs from released draft genomes, but lack annotation. Querying the SBP family on PlantTFDB, I downloaded all CDS from *N. benthamiana, S. lycopersicum,* and *A. thaliana* and merged them into a single FASTA file. As an outgroup, I chose the copper responsive transcription factor CrCRR1 from the green algae *Chlamydomonas reinhardtii,* as was shown to have homology to the *A. thaliana* SBP gene AtSPL7. CDS of CrCRR1 was obtained from the Uniprot database.

Multiple Sequence Alignment

I performed my alignment with the MUSCLE multiple sequence alignment program MUSCLE (11), which uses UPGMA-based iterative refinement of trees to generate an alignment of input sequences. I chose MUSCLE for its speed and accuracy with larger datasets, as well as some prior experience with the software. A limitation with MUSCLE is the dependence on initial tree generation guiding the alignment, and that it assumes a constant rate of evolution with its UPGMA iterative process. MUSCLE gives options to set total alignment steps and optimize models, however I performed on default parameters. Output files were viewed in the UGENE viewer to assess alignment quality.

Maximum Likelihood

Maximum likelihood trees were generated with the software IQ-TREE2 (12), a versatile and fast tool for tree inference. IQ-TREE2 uses an NNI hillclimbing algorithm with random tree perturbation to avoid local optima. IQ-TREE2 was ideal for its speed and automatic model selection, but has still has limitations with local optima and limited runs before reporting output tree. IQ-TREE2 gives users options for model selection and optimization, however I chose to run with default parameters to perform the automatic model testing. The substitution model chosen by IQ-TREE2 was TPM2u+F+R4. A consensus tree was generated in Newick Format, which was visualized with the Interactive Tree of Life (iTOL) web application.

Bayesian Inference

Bayesian inference was performed with the MrBayes3 software (13), a powerful tool for phylogenetic analysis. MrBayes3 is a flexible package that allows for a wide range of priors and models for large datasets. MrBayes relies heavily on the model priors to run analysis, which was viewed for confirmation with the Tracer software. MrBayes is computationally expensive, and requires some fore-knowledge of data. MrBayes is extremely customizable and requires users to specify a detailed model selection. MrBayes output files contained a consensus tree in Newick format, which was visualized with the Interactive Tree of Life (iTOL) web application.

A picture containing text, screenshot, font

Description automatically generated

Figure 1: MrBayes model parameters used to run phylogenetic inference with CrCRR1 as an outgroup

Github Repository with Reproducible Script (log-notebook.md)

<https://github.com/AustVDT/myProject>

**Results**

Maximum Likelihood tree generation with IQ-TREE2

Running all of my sequences through IQ-TREE with automatic model selection, the following tree was generated and visualized with iTOL. *A. thaliana* sequences are named by their annotation, while *N. benthiana* and *S. lycopersicum* sequences are listed by their gene IDs. Within this gene tree, I observe the several A. thaliana members that are known to be similar are in close proximity of each other. For example, the SBP genes SPL10 and 11 work redundantly in A. thaliana, and show up as neighbors on this tree. Similarly, SPL1 and 12 are redundant members from A. thaliana that have close domain similarity with SPL14, an association which is supported by this tree model.

Figure 2: IQ-TREE2 generated Maximum likelihood tree. Visualized in iTOL viewer.

A picture containing circle, sketch, text, drawing

Description automatically generated

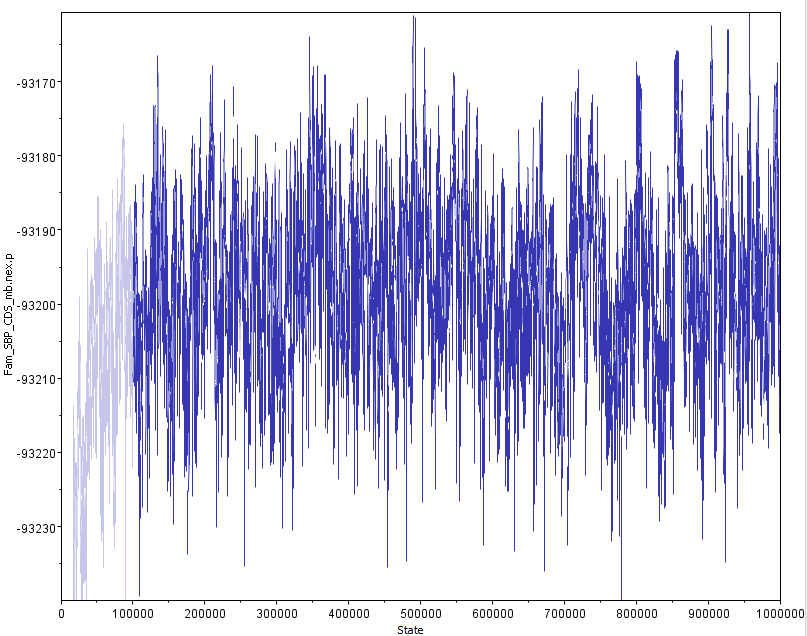
Bayesian Inference with MrBayes 3

Sequences were configured into Nexus file format and run with MrBayes3. MrBayes was first run without data to visualize just the prior distribution of the model. The prior distribution was opened in Tracer to view the parameter fit. MrBayes output files were read in iTOL for analysis.

**A**

**B**

A picture containing diagram, screenshot, plot, line

Description automatically generated

A picture containing sketch, circle, drawing, diagram

Description automatically generated

**C**

Figure 3: MrBayes analysis output. A) Trace Plot and B) alpha parameter distribution of prior. C) Circular view of MrBayes consensus tree visualized in iTOL.

Similar to the IQ-TREE2 consensus tree, MrBayes kept several known clades from A. thaliana together, particularly preserving the close similarities between AtSPL3/4/5, AtSPL10/11, and AtSPL1/12. Due to these experimentally validated similarities showing on the MrBayes tree, this gives my inference a high degree of support.

**Discussion**

Based on my analysis, the outputs from IQ-TREE2 and MrBayes both place key subgroups within this family properly, demonstrating near equivalent accuracy with my dataset. With MrBayes reliance on prior knowledge of the model, the logical workflow for future analysis would begin with IQ-TREE2 maximum likelihood for initial tree generation and informing model selection for a follow-up Bayesian inference with Mr.Bayes. Future work for refining this analysis would start with expansion of my sample data, broadening which plant species are included. While my analysis thus far has focused on *Solanaceae* with *Arabidopsis* as an outgroup, inclusion of even more distant relatives from annotated monocots such as rice and wheat would be interesting to show.

One consideration in my analysis is the assumption of fixed evolutionary rates in my analysis with MUSCLE. Plant genomes have evidence of extensive cycles of gene duplications and reductions that often generate gene copies with redundant or diverging functions, contributing to unique function of SBP genes between species. With this evolutionary history in mind, challenging the assumption of uniform evolution is important for finding the most appropriate model fit for my dataset. Exploration of different substitution models within he already used software, as well as expanding to compare with variable evolutionary models will accurately predict gene changes over time.

**References:**

1. Dickman et al. (2017) Reassessing Apoptosis in Plants. *Nature Plants* 3: 773-779

2. Li et al. (2010) Transgenic expression of an insect inhibitor of apoptosis gene, SfIAP, confers abiotic and biotic stress tolerance and delays tomato fruit ripening. *Phys. and Mol. Plant Path*. 74: 363-375

3. Kessens et al. (2018) An inhibitor of apoptosis (SfIAP) interacts with SQUAMOSA promoter-binding protein (SBP) transcription factors that exhibit pro-cell death characteristics. *Plant Direct.* 2 (8) 2018

4. Preston JC & Hileman LC (2013) Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. *Frontiers in plant science* 4:80.

5. Manning K, et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature genetics* 38(8):948-952.

6. Padmanabhan MS, et al. (2013) Novel positive regulatory role for the SPL6 transcription factor in the N TIR-NB-LRR receptor-mediated plant innate immunity. *PLoS pathogens* 9(3):e1003235.

7. Yamasaki H, Hayashi M, Fukazawa M, Kobayashi Y, & Shikanai T (2009) SQUAMOSA Promoter Binding Protein-Like7 Is a Central Regulator for Copper Homeostasis in Arabidopsis*. The Plant cell* 21(1):347-361.

8. Stone JM, Liang X, Nekl ER, & Stiers JJ (2005) Arabidopsis AtSPL14, a plant-specific SBP-domain transcription factor, participates in plant development and sensitivity to fumonisin B1. *The Plant journal : for cell and molecular biology* 41(5):744-754.

9. Salinas M, Xing S, Hohmann S, Berndtgen R, & Huijser P (2012) Genomic organization, phylogenetic comparison and differential expression of the SBP-box family of transcription factors in tomato. *Planta* 235(6):1171-1184.

10. Xie K, Wu C, & Xiong L (2006) Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant physiology* 142(1):280-293.

11. Edgar R (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32(5):1792–1797.

12. Minh B et al (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534

13. Ronquist F and Huelsenbeck J (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19(12):1572-1574