Modulation of alternative splicing by methylation in salt-stressed rice (*Oryza sativa*)

1) Introduction and Background

Genes in higher eukaryotes contain introns, sequences of DNA that are transcribed but then removed during pre-mRNA splicing. The process of splicing is highly variable in that different regions of the pre-mRNA transcript can be removed depending on environmental or intracellular factors. Thanks to this variability, called alternative splicing (AS), one gene can encode multiple protein products with different functions. Thus, the flexibility of splicing enables new functions and new modes of gene regulation to evolve.

From RNA-Seq and EST studies, we now know that around 50% or more of multi-exon plant genes can produce multiple transcript variants through AS (Reddy et al., 2013). Genes involved in stress response are especially prone to AS (Liu et al., 2013; Palusa et al., 2007; Staiger and Brown, 2013). These changes in AS can have a profound effect on the transcribed protein, leading to a protein with similar function, unique function, or even opposing function. However, the mechanism through which stress pathways regulate splicing is not well understood.

During splicing, a dynamic, macro-molecular machine called the spliceosome catalyzes intron removal via a multi-step process (Will and Luhrmann, 2011). Current thinking is that RNA-binding proteins control splice site selection by interacting with regulatory elements in the pre-mRNA. These proteins retard or promote spliceosome assembly at specific sites (Reddy et al., 2013). However, with the finding that splicing can occur co-transcriptionally, there is mounting evidence that epigenetic marks, especially DNA methylation, can also regulate splice site selection (Braunschweig et al., 2013; Shukla et al., 2011; Vargas et al., 2011).

Evidence for the role of methylation in AS comes from studies in humans and *Arabidopsis* showing that methylation is enriched in exons compared to the surrounding introns (Chodavarapu et al., 2010; Laurent et al., 2010; Takuno and Gaut, 2013). Furthermore, a study of alternatively spliced exons in human cell lines found that frequently skipped exons contained fewer methylated residues (Maunakea et al., 2013). However, similar work examining the role of methylation in AS has not been carried out in plants.

Considering the importance of AS in stress response, there is a large gap in knowledge as to how alternative splice sites are selected. Therefore, *this project will investigate the idea that DNA methylation influences splice site selection in plants*. Because abiotic stress, especially salt stress, has been shown to alter DNA methylation in rice and other plants, I will focus on investigating AS and DNA methylation in rice undergoing salt stress treatment (Angers et al., 2010; Karan et al., 2012; Lukens and Zhan, 2007).

2) Research Objectives, Methods, and Significance

Aim 1 - Analysis of alternative splicing in salt stressed rice via RNA-Seq.

An estimated 25% of the world's daily calories come from rice (*Oryza sativa)*, making it one of the most important grain crops (GRiSP, 2013). However, environmental stresses including extreme temperatures, drought and soil salinity pose significant problems for rice production. Rice is considered moderately sensitive to salinity, with yields decreased by up to 50% when grown in moderately salty conditions (Maas and Hoffman, 1977; Zeng et al., 2002). In North America alone there are over 15 million hectares of salt-affected soils (Shahid et al., 2013). Excess salinity is a particular problem for rice production in California and Louisiana, the number two and three rice producers in the US respectively, due to irrigation practices and salt-water intrusion (Chang and Brawer Silva, 2014; Scardaci et al., 1996; Personal Communication, Dr. Steven D. Linscombe, Station Director LSU AgCenter Rice Research Station).

As no studies have examined the role of AS and methylation in commercially grown rice undergoing salt stress, I will use *japonica* varieties Agami (salt-tolerant) and M103 (salt-sensitive) grown in control or saline growth media. *I hypothesize that changes in splicing patterns contribute to salt stress responses in rice, and that these changes depend in part on changes in DNA methylation status*. To test these hypotheses, I propose the following experiments.

Aim 1.1 - Rice genotypes and salt treatment.

The rice *japonica* varieties Agami (salt-tolerant) and M103 (salt-sensitive), a commercial variety grown in California, will be grown hydroponically in a temperature-controlled growth chamber according to standard protocols in a complete randomized design (Lee et al., 2003; Yoshida et al., 1976). Three replicates will be harvested for each analysis, however, double this number of plants will be grown to ensure enough samples live to the proper stage (Fig. 1). Rice will be exposed to salt by the addition of NaCl to reach an electrical conductivity of 12 dS m^{-1}, as measured by an electrical conductivity meter, for 14 days starting at the four-leaf stage. On the 14th day, the effects of the salt stress will be evaluated using the Modified Standard Evaluation System for the effects of salt treatment, in addition to shoot height and root length (Lee et al., 2003). Roots and shoots will then be harvested separately and stored at -80C.

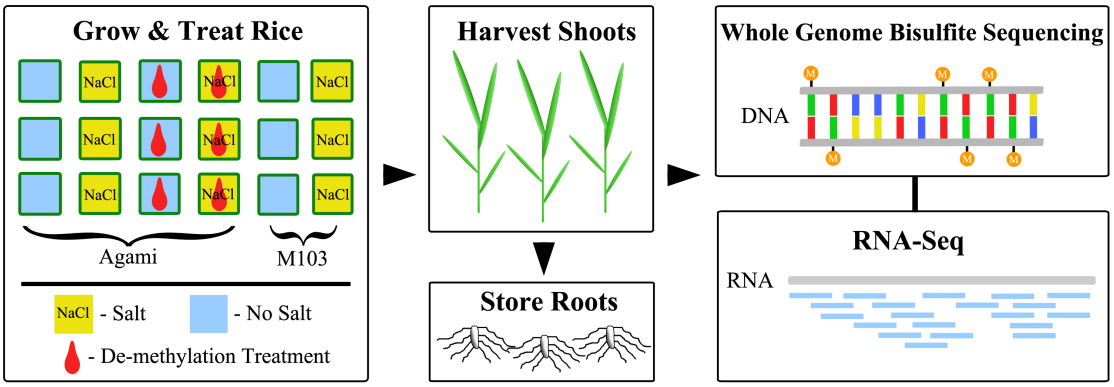

Figure 1: Diagram illustrating experimental design. Hydroponically grown rice Agami (salt-tolerant) and M103 (salt-sensitive) will undergo salt and/or de-methylation treatment. Shoots will be harvested and RNA-Seq and whole genome bisulfite sequencing carried out.

Aim 1.2 - RNA-Seq and analysis.

Salt-tolerant plants are able to withstand salt stress by controlling ion flow across their plasma membranes, maintaining a low Na^+ concentration in the cytoplasm, and a low rate of salt transport to the leaves (Baisakh et al., 2012). Mounting evidence indicates that AS plays a role in enabling salt tolerance. For example, in *Arabidopsis* the SOS4 gene (salt overly sensitive) undergoes AS, and the relative abundance of transcript variants changes upon salt exposure (Shi et al., 2002). RNA-Seq analysis of *Arabidopsis* undergoing salt stress treatment identified multiple AS events that were induced by salt stress (Ding et al., 2014). *I hypothesize that AS promotes salt tolerance by favoring transcripts that produce protein forms involved in salinity tolerance*.

To identify salt-tolerance AS isoforms and investigate the effect of salt exposure on AS, I will carry out RNA-seq analysis on the control and salt-stressed samples from the salt-tolerant and salt-sensitive varieties. As the overall goal of this proposal is to examine the relationship between AS and methylation, I will use shoot tissue, as methylation within the shoot is strongly affected by salt exposure compared to root tissue (Karan et al., 2012). Shoot tissue will be ground, and RNA and DNA (for bisulfite sequencing in Aim 2 and 3) from the same plants extracted using standard methods (root tissue will be saved for potential future comparison) (Loraine et al., 2013). The Loraine lab is highly experienced in RNA-Seq, and will assist in setup, analysis, and troubleshooting. I will prepare libraries for directional sequencing of both small and long RNAs on an Illumina HiSeq2000 with six samples per lane (Loraine et al., 2013).

Following sequencing, reads will be aligned onto the reference genome using TopHat (for long RNA libraries) and ShortStack (for short RNA libraries) (Trapnell et al., 2009). The Loraine Lab collaborates with Michael Axtell's group at Penn State, which develops ShortStack, and they will advise me on running the program and interpreting results.

Gene expression differences at the whole gene level will be assessed using EdgeR, DESeq, and the CuffLinks suite, three standardized methods for analysis of RNA-Seq data. For AS analysis, I will use a

combination of approaches, as methods for detecting splicing changes arising from experimental treatments are still an active area of research (Liu et al., 2014).

A potential complication is that most methods were developed to analyze mammalian data. Mammalian introns are typically much larger than in plants and exhibit less intron retention, which in plants is the dominant mode of AS (English et al., 2010; Gupta et al., 2004; Wang and Brendel, 2006). To accommodate this, the Loraine group developed an alternative approach based on the ArabiTag algorithm, and has experimentally validated the results (English et al., 2010). Using their method and others, I will identify genes that are differentially spliced in response to salt stress in both salt-tolerant and salt-sensitive rice varieties. Such genes would provide insight as to how salt tolerance is regulated in rice and may suggest strategies for breeding salt resistance in rice and other crops.

Aim 2 - Modulation of alternative splicing by methylation.

The knowledge that splicing occurs co-transcriptionally opened the door to understanding new mechanisms controlling AS. A landmark study performed using human cell lines discovered that methylation within alternatively spliced exons correlated with how frequently they were included in transcripts (Maunakea et al., 2013). This observation suggests a mechanism in which methylation enables spliceosomal assembly on splice sites flanking alternative exons. In plants, nothing is known about any such relationship between splicing and methylation. *I hypothesize that such a relationship exists, and that I can detect it by comparing whole-genome methylation data (from bisulfite sequencing) to whole-genome splicing data (from RNA-Seq)*.

Aim 2.1 - Whole genome bisulfite sequencing and analysis.

Exposure of rice varieties to salt stress induces changes in methylation within stress response genes, indicating methylation plays a role in response to salt stress (Karan et al., 2012). However, this study only took into account ten genomic locations where methylation differed between control/salt-treated. To better understand the role of methylation in stress response I will generate whole genome bisulfite sequence data from the same control and salt stressed shoot samples from Aim 1.1 (Fig. 1). Two PIs located at the North Carolina Research Campus (NCRC) routinely use this technique and have offered to assist me; they are Dr. Tzung-Fu Hsieh (North Carolina State University), who studies methylation in *Arabidopsis* and rice, and Dr. Folami Ideraabdullah (University of North Carolina at Chapel Hill), who studies epigenetic marks in mouse (Hsieh et al., 2009; Hsieh et al., 2011; Ibarra et al., 2012; Ideraabdullah et al., 2011). Both have offered to prepare a letter of collaboration upon request.

I will carry out library preparation for Illumina sequencing, followed by bisulfite conversion using published methods (Hsieh et al., 2009). Sequencing will be done on an Illumina HiSeq with three samples per lane to obtain ~20X coverage.

Methylation analysis of the bisulfite sequencing data will be carried out using Bismark (Krueger and Andrews, 2011). Bismark aligns bisulfite-converted sequences by computationally converting all cytosines to thymines in both sample and reference sequences, which are then aligned. Aligned samples are then compared against the original non-converted reference sequence, and frequency of methylation calculated. Custom scripts in python or R will be developed as needed to further process and analyze the data. My host lab has ample experience in bioinformatics programming and I intend, as part of the training component of the fellowship, to master basic scripting in both languages.

Analysis will compare changes in observed methylation between control/salt-treated samples, as well as between the two varieties of rice. By subtracting differences in methylation between controls, any differences between the two varieties following salt-treatment should indicate differences in how the two varieties of rice react to salt-treatment. This analysis will reveal methylation patterns unique to salt-tolerance in Agami and salt-sensitivity in M103, highlighting loci that are critical for salt tolerance in rice.

Aim 2.2 - Comparison of RNA-Seq and whole genome bisulfite sequencing data.

To test the hypothesis that methylation of alternatively spliced exons (ASE) increases their inclusion, I will compare the RNA-Seq (from Aim 1.2) and whole genome bisulfite sequencing (from Aim 2.1) data sets. Using the RNA-Seq expression data, I will identify ASE that are included at different frequencies using the Percent-Spliced-In (PSI) metric (Katz et al., 2010). PSI is a calculation of the percentage of RNA-Seq aligned reads mapping to an alternatively spliced region that supports the exon-included form.

Analyses of ESTs and RNA-Seq data sets from blueberry, *Arabidopsis*, and rice by the Loraine lab found that, in general, the majority of transcripts either include (>90%) or exclude (<90%) the ASE (Fig. 2A). By splitting these ASE into two groups, I can compare levels of methylation between those that are generally included (red) versus excluded (blue). *I hypothesize that ASE at the extreme ends of the distribution, e.g., exons that are nearly always included or excluded from their respective transcripts, differ significantly with respect to their methylation status* (Fig. 2B).

To achieve greater power to distinguish links between methylation status and AS, I will run this analysis both within and between treatment groups. I will examine the sub-population of exons where PSI (exon inclusion) differs between varieties or treatments and test whether methylation increases or decreases with exon inclusion. If my hypothesis is correct, then increasing exon inclusion will correlate with increased methylation.

A potential pitfall with this approach is that it has been optimized for exon skipping events, which are found frequently in animal models, but not in plants (English et al., 2010; Gupta et al., 2004; Wang and Brendel, 2006). In plants, the majority of AS events are intron retention, and for many genes with ASE, the introns flanking the ASE may be especially prone to intron retention (Gulledge et al., 2012). To account for this, I will also examine intron retention events and their levels of methylation. If lower levels of methylation within ASE cause them to be skipped, then I predict that higher methylation within introns would lead to their retention.

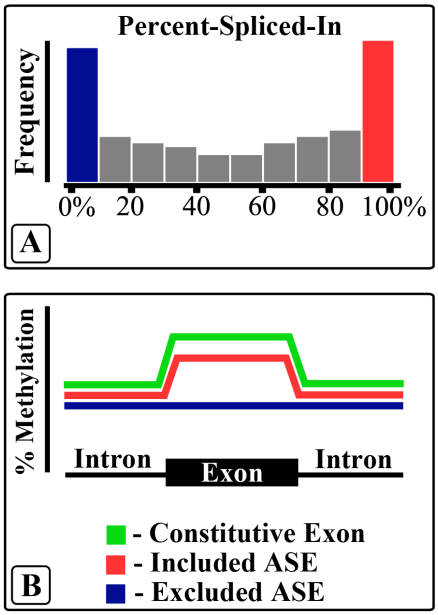

Figure 2: (A) - Illustration of Percent-Spliced-In frequency generally observed in plants. Majority of transcripts do not include (blue) or do include (red) alternatively spliced exons (ASE). (B) - Diagram of predicted outcome for percentage of methylation at exon, where that exon is constitutively expressed (green), usually included as an ASE (red) or usually excluded (blue).

Aim 3 - Experimentally modify methylation to test effects on alternative splicing.

To functionally demonstrate my hypothesis that methylation modulates AS in rice, I will experimentally alter methylation and look for the associated changes in AS. To disrupt methylation, I'll treat rice seedlings with 5-aza-2'-deoxycytidine (5-azaC), which inhibits DNA methyltransferase activity, decreasing methylation in rice (Eun et al., 2012; Kumpatla et al., 1997; Tsugane et al., 2006). *I hypothesize that reducing methylation of ASE will increase exon skipping in alternatively spliced transcripts*.

Aim 3.1 - Treatment of rice with methylation altering chemical 5-aza-2'-deoxycytosine.

Agami rice seeds will be soaked for 16 hours in 3mM of 5-azaC before being grown hydroponically. To make all results from this experiment comparable with data from previous aims, rice for all experiments will be grown simultaneously, allowing the non-5-azaC treated Agami seedlings from Aim 1 and Aim 2 to act as controls (Fig. 1). As 5-azaC can be dissolved in water, no mock treatment will be necessary, further simplifying the experimental setup. To compare between both control and salt-treated Agami samples, 5-azaC treated Agami will also undergo control (no salt) and salt treatment. Both shoots and roots will be harvested and stored at -80C.

Aim 3.2 - Effects of experimentally reduced methylation on alternative splicing.

The goal of this aim is to functionally demonstrate the role of methylation in AS. DNA and RNA will be isolated from shoots, followed by both RNA-seq and whole genome bisulfite sequencing following the same guidelines outlined above. ASE and retained introns identified in Aim 2.2 will be analyzed using this new data. In the event that Aim 2.2 does not identify sufficient candidate AS regions, I will also examine genes known to be involved in salt stress tolerance.

I will compare the data generated from Aim 2.2 (no 5-azaC) and Aim 3.2 (5-azaC). Where ASE or retained introns had been previously located, and where there is a significant difference in methylation between no 5-azaC and 5-azaC treated samples, I will analyze the frequency of AS. *I hypothesize that experimentally reduced methylation of ASE and retained introns will decrease their inclusion rate*.

3) Training Objectives

My training objectives are to gain expertise in plant biology, high throughput pipelines for RNA-Seq and whole genome bisulfite sequencing, and bioinformatic data analysis. The sponsoring lab will provide the majority of training for working with rice and RNA-Seq data, and collaborations with Dr. Tzung Fu Hsieh and Dr. Folami Ideraabdullah, who are based at the NCRC, will provide training in whole genome bisulfite sequencing and methylation analysis.

To acquire the necessary skills in bioinformatics and connect with other biologists transitioning into bioinformatics, I plan to attend training events such as the iPlant workshop on large-scale data analysis, the annual Cold Spring Harbor Programming for Biology and Computational and Comparative Genomic courses, or one of many Software Carpentry courses in programming and UNIX. Another option will be to enroll in the new Summer Institute in Bioinformatics offered at UNC Charlotte; this course is designed for researchers seeking a fast route to gaining expertise in bioinformatics.

Dr. Loraine and I have also discussed a six-month program in which I'll gain the core skills in bioinformatics data analysis: ability to work productively in a Unix environment, including working with a computer cluster or Amazon Web Services; ability to do data analysis using R and tools from the Bioconductor libraries; and ability to write simple data processing and "data munging" scripts in python. Toward this end, Dr. Loraine developed a set of training modules that covers these topics. For data analysis training, I plan to take the R Programming course taught by Roger Peng as part of the Johns Hopkins on-line Data Science courses (Coursera), the Genome Analysis course taught by Rafael Irizarry (EdX), and the Data Analysis course taught by Jeff Leek (Coursera). To build python programming skills, I will work through the Rosalind on-line training modules. As I learn, I plan to develop teaching materials I can then use for courses I teach in my first faculty position.

Other training objectives include developing professional manuscript and grant-writing skills. This will occur through close tutelage with Dr. Loraine, as well as the Catalyst Postdoctoral Group at the NCRC, which provides regular training experiences in professional writing and editing.

4) Career Development

This project differs from my dissertation research by using bioinformatic analysis to understand mechanisms of splicing in plants. My dissertation research was in developmental biology and genetics in animals, where I studied posterior vertebrae formation in a chicken breed (Araucana) that lacks a tail (Freese et al., 2014; Noorai et al., 2012). My Ph.D. research introduced me to bioinformatics and genetics, but the primary focus was in embryology and developmental biology techniques. Following my PhD, I worked for five months at the Max Planck Institute for Chemical Ecology. I studied the resistance of maize to western corn rootworm larvae, where I was able to gain new skills in plant biology.

Training with the host lab will help establish myself within the field of plant bioinformatics and AS. This will enable me to launch an all-new career in plant science. In addition, the research I have proposed will generate several data sets that will expose new mechanisms for salt tolerance, splicing, and epigenetic modifications in plants. These data sets will enable me to propose many new hypotheses and experiments, making me an appealing candidate for a faculty position, my ultimate goal.

5) Sponsoring Scientist and Host Institution

My proposed sponsor, Dr. Loraine, works on abiotic stress and AS in plant models, including *Arabidopsis* and rice (English et al., 2010; Gulledge et al., 2012; Loraine et al., 2013). Dr. Loraine's lab has expertise in programming, bioinformatics pipelines, and is regularly available for training in AS methodology. Dr. Tzung Fu Hsieh, also at the NCRC, will provide training in whole genome bisulfite sequencing, his area of expertise (Hsieh et al., 2009; Hsieh et al., 2011; Ibarra et al., 2012).

I will carry out my research at the NCRC, which provides access to plant growth facilities, computer clusters, and a genomics core facility equipped with Illumina HiSeq sequencers. Instrumentation needed to prepare libraries for bisulfite sequencing is also available. Therefore, I will have immediate access to expertise and training opportunities needed to carry out my proposed work.

My current position is postdoctoral research fellow in Dr. Loraine's lab. I was recruited to lead outreach and training efforts for the Integrated Genome Browser (IGB) project. However, Dr. Loraine has stressed that my postdoctoral position also include extensive training opportunities. As such, I have been working with the Loraine lab to build new expertise in data analysis, and then use that expertise to connect with new collaborators for the IGB project. For instance, this December I was able to help Dr. Flenniken of Montana State University parse *Crithidia mellificae* genomic and transcriptomic data into BED format, enabling it to be viewed in IGB. I propose to remain in the Loraine Lab to continue the proposed work as the resources here are outstanding and will provide everything I need, including a supportive atmosphere, training in bioinformatics and plants science, and access to instrumentation.

6) Yearly Goals and Outcomes

Aim	Objectives	Outcomes	Y1		Y2		Y3	
1	**Treatment & AS**							
1.1	Stress Treatment	Shoot RNA and DNA samples	■					
1.2	RNA-Seq	Whole genome AS analysis		■				
2	**Methylation (M) & AS**							
2.1	Bisulfite-sequencing	Whole genome methylation analysis			■			
2.2	AS-M Comparison	Analysis of M in alternatively spliced transcripts				■	■	
3	**Alter Methylation**							
3.1	Treat hydroponic rice	Growth, RNA, and DNA samples	■					
3.2	Analyze AS & M	Changes in AS due to altered methylation						■

7) Broader Impacts

Carrying out the proposed research in rice, a model cereal grain with vast economic value, will maximize potential benefits to agriculture. By understanding the role of methylation in AS as it applies to salt tolerance, better programs for breeding and identifying salt tolerant varieties could be found.

My training will also involve mentoring junior scientists. As a graduate student, I mentored undergraduate students in the Eureka (Experiences in Undergraduate Research, Exploration and Knowledge Advancement) program. Eureka helped students transitioning from high school to college. Inspired by this, I am developing a similar program here at the NCRC. This program - the Research Campus Plant-Bio Summer Internship - is a collaboration with A.L. Brown High School (letter of collaboration available upon request). This paid internship will give high school juniors first-hand experience with plant biology research in the summer before their senior year. To start, the Loraine Lab has committed to hosting two students each summer. Interns will learn basic lab procedures, and will carry out their own supervised research project related to salt stress in rice. The goal of this internship will be to inspire the students to pursue higher education in science.

I am also volunteering to serve as a multi-session guest speaker for the FILM @ 6 Program, an after school program that offers 6th graders tutoring and support, with the overall goal being to make STEM relevant to young students. As such, I have volunteered my time as a multi-session guest speaker for the program, covering topics such as plant genetics (letter of collaboration available upon request).