Genome-wide Mapping of Nucleotide Excision Repair in *Arabidopsis* Root and Shoot & The Fate of LYM1/2 Proteins During Nodulation in *Medicago truncatula*

by

Duğçar Ebrar Erdoğan

A Dissertation Submitted to the

Graduate School of Sciences and Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Master of Science

in

Molecular Biology and Genetics

KOÇ
ÜNİVERSİTESİ

August 1, 2022

Genome-wide Mapping of Nucleotide Excision Repair in *Arabidopsis* Root and Shoot & The Fate of LYM1/2 Proteins During Nodulation in *Medicago truncatula*

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Duğçar Ebrar Erdoğan

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Asst. Prof. Onur Öztaş (Advisor)

_____

Asst. Prof. İsmail Kudret Sağlam

_____

Asst. Prof. Ogün Adebali

Date: 01.08.2022

To my beloved family

ABSTRACT

Genome-wide Mapping of Nucleotide Excision Repair in Arabidopsis Root and Shoot & The Fate of LYM1/2 Proteins During Nodulation in *Medicago truncatula*

Duğçar Ebrar Erdoğan

Master of Science in Molecular Biology and Genetics

August 1, 2022

Plants, unlike many other organisms, are not mobile, which makes them vulnerable to a variety of hazardous factors such as ultraviolet (UV) light. UV light is known to cause DNA damage, which in turn leads to genetic instability. Organisms have found ways to cope with DNA damage by different molecular mechanisms to preserve genetic information flow properly. Nucleotide excision repair (NER) is a conserved mechanism widely utilized by many organisms, including plants. In this study, The Excision Repair Sequencing (XR-seq) method has been applied to shoot, root, and seedling tissues in *Arabidopsis thaliana* to map genome-wide DNA damage types: CPD and (6-4) PP. *Arabidopsis thaliana* has been shown to have transcription-coupled repair (TCR) in addition to global genome repair (GG-NER). Interestingly, root tissue, despite not being exposed to UV light, has demonstrated a significant amount of nucleotide excision repair, whose levels are similar to tissues exposed to UV. Root tissue might exhibit such repair due to the chemicals in soil, which might give rise to the formation of various DNA damage products repaired by the NER mechanism. Together with this finding, the repair dynamics could be further investigated in the model organism Arabidopsis to understand better essential players of DNA damage, gene activity, and chromatin accessibility throughout the plant genome.

Medicago truncatula is a model legume that undergoes a mutualistic relationship with rhizobia under nitrogen-limited conditions to obtain ammonium (NH_3). In this interaction, rhizobia produce ammonium from atmospheric nitrogen (N_2) and provide it to the plant and in return, use the carbon source supplied by the plants. However, the question and mechanism of how the bacteria are allowed to maintain their cellular processes inside nodule organs within the root tissue of plants remain to be addressed in great detail. Lysin motif-containing receptor (LysM) proteins, LYM1 and LYM2, reside on plants' plasma membranes and are responsible for triggering innate immunity in response to patterns of bacterial peptidoglycan structures. Furthermore, upon arrival of rhizobia, the GPI anchors of LYM1 and LYM2 proteins are known to be cleaved that releases these proteins from symbiosome membrane. Thus, these proteins are inactivated to initiate an immune response so the intracellular rhizobia could survive, and the symbiotic relationship can be attained. This study includes cloning of *LYM1* and *LYM2* genes into plant expression vectors and genotyping *lym1* and *lym2* mutants. The questions of what the other interacting partners of LYM proteins are and what happens to LYM proteins in the absence of DNF2 protein need to be resolved in further studies.

ÖZETÇE

Arabidopsis Kökünde ve Sürgünde Nükleotid Eksizyon Onarımının Genom Çapında Haritalanması ve *Medicago truncatula*'da Nodülasyon Sırasında LYM1/2 Proteinlerinin Kaderi

Duğçar Ebrar Erdoğan

Moleküler Biyoloji ve Genetik, Yüksek Lisans

1 Ağustos 2022

Bitkiler, diğer birçok organizmanın aksine hareketli değildir ve bu da onları ultraviyole (UV) ışığı gibi çeşitli tehlikeli faktörlere karşı savunmasız hale getirir. UV ışığının DNA hasarına neden olduğu ve bunun da genetik kararsızlığa yol açtığı bilinmektedir. Organizmalar, genetik bilgi akışını düzgün bir şekilde korumak için farklı moleküler mekanizmalar yoluyla DNA hasarıyla başa çıkmanın yollarını bulmuşlardır. Nükleotid eksizyon onarımı (NER), bitkiler de dahil olmak üzere birçok organizma tarafından yaygın olarak kullanılan korunmuş bir mekanizmadır. Bu çalışmada, genom çapında DNA hasar tiplerini haritalamak için Arabidopsis thaliana'daki sürgün, kök ve fide dokularına Eksizyon Onarım Dizileme (XR-seq) yöntemi uygulanmıştır: CPD ve (6-4) PP. Arabidopsis thaliana'nın global genom onarımına (GG-NER) ek olarak transkripsiyon-bağlı onarıma (TCR) sahip olduğu gösterilmiştir. İlginç bir şekilde, kök dokusu, UV ışığına maruz kalmamasına rağmen, seviyeleri UV'ye maruz kalan dokularınkine benzer ve önemli miktarda nükleotid eksizyon onarımı göstermiştir. Kök dokusu, topraktaki kimyasallar nedeniyle böyle bir onarım sergiliyor olabilir ve bu da NER mekanizması tarafından onarılan çeşitli DNA hasar ürünlerinin oluşumuna neden olabilir. Bu bulgu ile birlikte, onarım dinamikleri, bitki genomu boyunca DNA hasarı, gen aktivitesi ve kromatin erişilebilirliğinin daha iyi temel oyuncularını anlamak için model organizma Arabidopsis'te daha fazla araştırılabilir.

Medicago truncatula, amonyum (NH3) elde etmek için azotla sınırlı koşullar altında rizobi ile karşılıklı ilişkiye giren bir model baklagildir. Bu etkileşimde rizobi atmosferik nitrojenden (N2) amonyum üreterek bitkiye verir ve karşılığında bitkilerin sağladığı karbon kaynağını kullanır. Bununla birlikte, bakterilerin, bitkilerin kök dokusu içindeki nodül organlarında hücresel süreçlerini sürdürmelerine nasıl izin verildiği sorusu ve mekanizması, ayrıntılı olarak ele alınmayı beklemektedir. Lizin motifi içeren reseptör (LysM) proteinleri, LYM1 ve LYM2, bitkilerin plazma zarlarında bulunur ve bakteriyel peptidoglikan yapı modellerine yanıt olarak doğuştan gelen bağışıklığı tetiklemekten sorumludur. Ayrıca, rizobinin gelmesi üzerine, LYM1 ve LYM2 proteinlerinin GPI çapalarının, bu proteinleri ortakyaşam zarından serbest bırakan parçalandığı bilinmektedir. Böylece, bu proteinler, hücre içi rizobinin hayatta kalabilmesi ve simbiyotik ilişkiye ulaşılabilmesi için bir bağışıklık tepkisi başlatmak için etkisiz hale getirilir. Bu çalışma, LYM1 ve LYM2 genlerinin bitki ekspresyon vektörlerine klonlanmasını ve *lym1* ve *lym2* mutantlarının genotiplendirilmesini içerir. LYM proteinlerinin diğer etkileşen partnerlerinin neler olduğu ve DNF2 proteininin yokluğunda LYM proteinlerine ne olduğu sorularının ileriki çalışmalarda çözülmesi gerekmektedir.

ACKNOWLEDGEMENTS

First of all, I would like to give my greatest acknowledgements to Dr. Onur Öztaş, who helped me understand how science should be applied in real life. His patience and belief in my progress have further contributed to the person I am today as a scientific researcher. It has been a wonderful experience to work with him.

It is also a pleasure to have such a great thesis committee members İsmail Kudret Sağlam and Ogün Adebali, who are leading scientists in their expertise. I must thank them as they have accepted to participate my thesis defense.

Moreover, Dr. Ogün Adebali and his Ph.D. student Sezgi Kaya have shown me excellent guidance on how to analyze biological data. It was my first and foremost experience with the use of computers cleverly. I feel grateful to have such collaboration with them.

A lab without diverse and kind-hearted lab members, who support your growth, not just as a scientist but also as a human being, is not the best lab. Therefore, I owe my one of the biggest appreciation to İlayda Göktepe, Aysun Gülünk, Deren Büyüktaş, Hidayet Pütün, Kutluhan Gülbahar, Sami Yalçıntaş, who have always been there for me. I feel so lucky to have been in such a stimulating, science-driven environment where everyone passionately carries out their own research and asks very fundamental questions related to nature. I will miss each of them and the fun times we had, and I wish each of them the best on their own journeys.

Other special thanks go to my dear friends Ege Ezen, İlayda Özcan, and Öznur Ataman, who not only have shared the most valuable memories of my life and but also always encouraged my personal growth. They have kept believing and supporting me regardless of how I am doing with my research. I cannot actually find a word to describe how grateful I am to them. I love you all.

Outside the lab and friends, I must thank my all family, who has always given me unconditional love and support. It feels wonderful to know that they will be always there for me.

Now, I am ready to start a new chapter of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ATP	Adenosine triphosphate
CAMK	Ca^{2+}/calmodulin-dependent protein kinase
cDNA	Complementary DNA
CEBIP	The chitin elicitor-binding protein
CERK	Ceramide kinase
CETN	Principal component analysis
CSA	Cockayne syndrome protein A
CSB	Cockayne syndrome protein B
CDS	Coding sequence
CTAB	Hexadecyltrimethylammonium bromide
CPD	Root Mean Squared Error
DDB	DNA damage-binding protein
DNA	Deoxyribonucleic acid
DNF	Defective in nitrogen fixation
ERCC1	Excision repair cross-complementation
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin adenine dinucleotide
FRET	Fluorescence resonance energy transfer
GGR	Global genome repair
HF	High fidelity
HPC	High performance computing
IP	Immunoprecipitation
IPTG	Isopropyl β- D-1-thiogalactopyranoside
LB	Lysogeny broth
LCO	Lipochitooligosaccharide
LRR	Leucine-rich receptor
LYM	Lysin motif containing receptor
LYMK	Lysin motif Receptor-like kinase
MAP	Mitogen-activated protein
NCR	Nodule-specific-rich
NFR	Nod factor receptor
NER	Nucleotide excision repair

NGS	Next generation sequencing
RNS	Reactive nitrogen species
NTS	Non-transcribed strand
PAMP	Pathogen-associated molecular pattern
PCA	Principal component analysis
PCR	Polymerase chain reaction
PIC	Pre-incision complex
PRR	Pattern recognition receptor
PP	Photoproduct
REC1	Reduced chloroplast coverage
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
ROS	Reactive oxygen species
RPA	Replication protein A
SNF	Symbiotic nitrogen fixation
SNM	Standart nodulation medium
SOC	Super optimal broth with Catabolites repression
SYP	Synatixin protein
TCR	Transcription-coupled repair
TES	Transcription ending site
TFIIH	Transcription factor IIH
TOPO	Topoisomerase based cloning
TS	Transcribed strand
TSS	Transcription start site
UV	Ultraviolet
Xgal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
XP(A-G)	Xeroderma pigmentosum (A-G)
XR-seq	Excision repair sequencing

Chapter 1:
GENOME-WIDE MAPPING OF NUCLEOTIDE EXCISION REPAIR IN ARABIDOPSIS ROOT AND SHOOT

1.1 Introduction

1.1.1 Overview of DNA repair

Organisms from prokaryotes to eukaryotes in their daily lives are exposed a variety of physical and chemical perturbations that lead to changes in their molecular systems. One of the most essential molecular perturbation takes place at DNA level. It is significant to understand the alterations in DNA as they might give rise to changes in the meaning of information such as mutations, which could lead to cancer progression. Life have found a way to overcome those perturbations to maintain the integrity of the genome through evolutionary mechanisms, so called DNA repair. Because of its chemical nature, DNA is susceptible to a variety of physical and chemical factors shown in Figure 1.1 below.

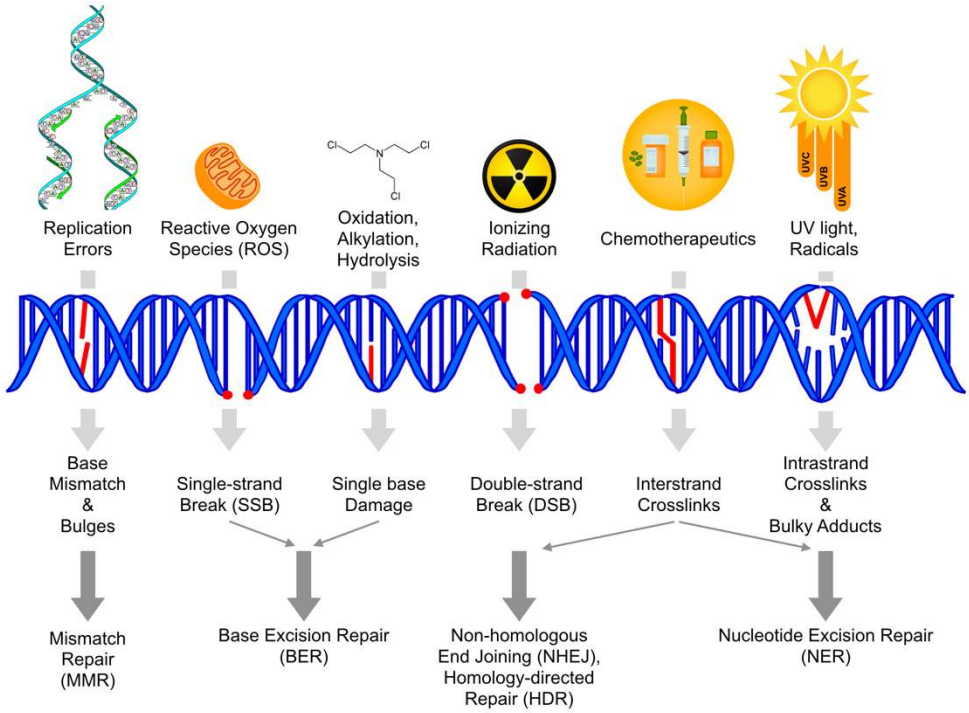

Figure 1.1: DNA damage factors, their alterations and repair types (from http://komorlab.ucsd.edu/research.html#).

Among them, ROS, oxidation agents, and chemotherapeutics are chemicals, which directly interfere with DNA while ionizing radiation and UV light are physical interferences, which disrupt and or catalyze some special chemical reactions as in the case of bulky adduct formation such as CPD and (6-4) PP ones via cycloaddition reaction. The damage might be either called covalent in the case of CPD and (6-4) PP and cisplatin or non-covalent such as base mismatches, bubbles, and loop formation depending on the nature of chemical interaction (Sancar et al., 2004). These CPD and (6-4) PP type of damages are called exogenous lesions as they are formed as a result of exterior factors such as UV light. On the other hand, reactive oxygen and nitrogen species (ROS, RNS) are called endogenous lesions since they are formed inside the cell as a result of metabolic reactions (Wozniak & Simmons, 2022). Cellular systems have evolved to respond to DNA damage through removing the damage by DNA repair mechanism, checking cell cycle tightly to make sure that DNA is successfully replicated before cell division, altering transcriptomic profile to trigger damage-specific signal pathways, and undergoing apoptosis where DNA is heavily damaged such that it is not possible to repair DNA (Sancar et al., 2004).

As in Figure 1.1, according to type of DNA damage, the whole repair systems might be summarized as mismatch repair, base excision repair, non-homologous end joining, homology directed repair, and nucleotide excision repair.

1.1.2 Types of DNA Damage Recognition

DNA damage could be recognized in several ways: direct recognition, multistep recognition, recognition via proximal distance, and intermediate molecules arisen from changes in DNA structure (Sancar et al., 2004).

Photolyase and DNA glycosylase enzymes can directly interact with the damage site and trigger a cascade of signal pathways associated with DNA damage.

Furthermore, Matchmakers help two molecules bring together via conformational changes and then leave the system at the end. One of the best example for such a molecule in humans is XPC in nucleotide excision repair (NER). Moreover, the synergetic interaction of two independent molecule associated towards a specific molecule is called

combinatorial recognition and the best example of this is the association of photolyase and UvrA towards DNA damage (Sancar et al., 2004)

In addition, RNA polymerase might encounter with a problem during transcription a specific gene and it leads to activation of some signaling pathways associated DNA damage. Since it does not directly look for specific damages in DNA structure but just transcribes transcribed strand (TS) of a gene but not non-transcribed strand (NTS) as other actively surveillance damage factors, it senses the damage via proximity and this type of recognition is called recognition by proximity (Sancar et al., 2004).

Lastly, some intermediate DNA molecules such as single stranded nicks and double strand breaks might be sensed to trigger a damage response across the cell (Sancar et al., 2004).

1.1.3 DNA repair mechanisms

Direct repair

Direct repair of a damaged DNA takes place with the help of methylguanine DNA methyltransferase and photolyase enzymes. Methylguanine DNA methyltransferase is recruited when a guanine nucleotide is chemically altered in a way that double bond oxygen of the molecule is methylated so the molecule becomes O^6methyl guanine nucleotide, which is reversed by the action of methylguanine DNA methyltransferase. Photolyase instead utilizes blue light to catalyze the reaction. Figure 1.2 shows the action mechanism of photolyase enzyme.

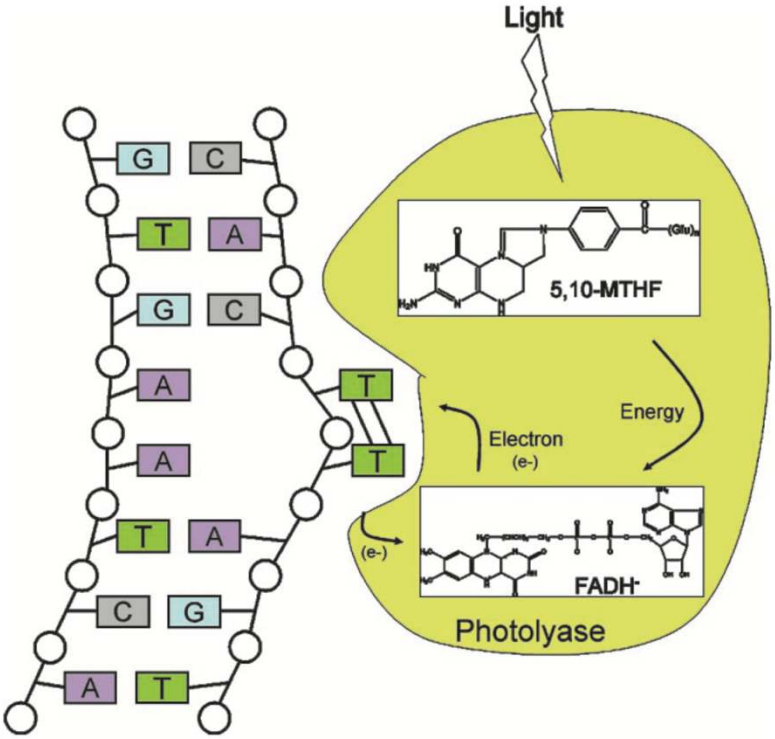

Figure 1.2: The mechanism of action performed by photolyase enzyme in repair of thymine-thymine dimers created by UV light.

The photolyase enzyme has been speculated to evolve in early times deep in oceans. Since the blue light of the invisible light spectrum has the most physical ability to penetrate the water compared to other visible light regions, organisms, which lived at that times learned to utilize it to cope with DNA damage exerted by UV light (Sancar et al., 2010). Light is the initiator of the reaction as 5,10-MTFH, or folate, has the ability to absorb energy of violet-blue region of the spectrum and responsible for transferring some of this excitation energy to $FADH^-$ molecule nearby the folate via FRET mechanism, implying that the proximity of these two molecules is essential. As a result, the $FADH^-$ is reduced by transferring an electron directly to the site where thymine-thymine dimer is present. The electron introduced to the dimer destabilizes the adduct structure and the molecule is rearranged to return its undamaged structure. Upon reversal of the damage, the electron transfers back to the FADH, and FADH becomes $FADH^-$, resulting in dissociation of the enzyme (Sancar et al., 2004).

Base Excision Repair

Base excision repair mechanism is employed in cells when chemical structure of a base is altered via oxidation, alkylation, and deamination or in the case of base mismatches. The enzyme initiating the mechanism is called DNA glycosylase which removes the damaged target base without altering sugar backbone (Sancar et al., 2004). As a key note, guanine base is the one most susceptible to oxidizing hazards in the cells (Wozniak & Simmons, 2022).

Nucleotide Excision Repair

Nucleotide excision repair (NER) is the mechanism where a variety of bulky adducts are repaired through cutting the damaged region of the DNA by introducing dual incisions (Petit & Sancar, 1999). Then a new strand is synthesized by DNA polymerase, using the complementary strand as a template, followed by ligation (Friedberg, 2001). The substrates of this mechanism are quite wide, including UV-induced (especially UVB) photoproducts such as CPD and (6-4) PP, various aromatic amines such as benzo[a]pyrene, and adducts formed as a result of therapeutics purposes such as cis-platin (Z. Hu et al., 2016; Mao et al., 2016). A representative chemical alteration for a dithymine nucleotide is shown in Figure 1.3.

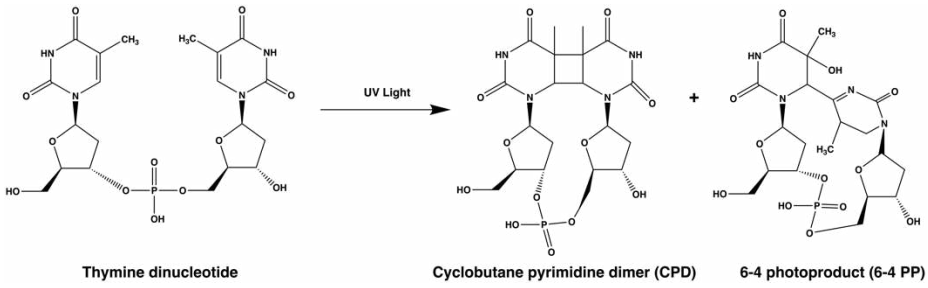

Figure 1.3: Chemical alteration of a thymine dinucleotide upon UV-exposure.

From the structure point of view, (6-4) PP is more distorting compared to CPD, resulting in faster repair of (6-4) PP compared to CPD photoproducts (Cleaver et al., 2009; J. Hu et al., 2015; Reardon & Sancar, 2005).

The fact that different types of damage molecules are not similar in structure does not explain how the mechanism is able to recognize such a variety of substrates but

reveals that those adducts are thermodynamically unstable in common, which make them important targets for nucleotide excision repair mechanism (de Laat et al., 1999; Sancar et al., 2010; Schärer, 2013). Moreover, the activity of NER has been reported to be influenced by chromatin accessibility and transcription factor binding (Mao et al., 2016; Sabarinathan et al., 2016). The mechanism follows a set of sequential events, including the action of a molecular matchmaker, unwinding of the double-strand by ATP hydrolysis, incision making at 3'end, then 5'end, elimination of the excised oligomer, and finally re-synthesis of eliminated region (Petit & Sancar, 1999).

The proposed mechanism of NER in mammals is demonstrated in Figure 1.4 (Reardon & Sancar, 2005).

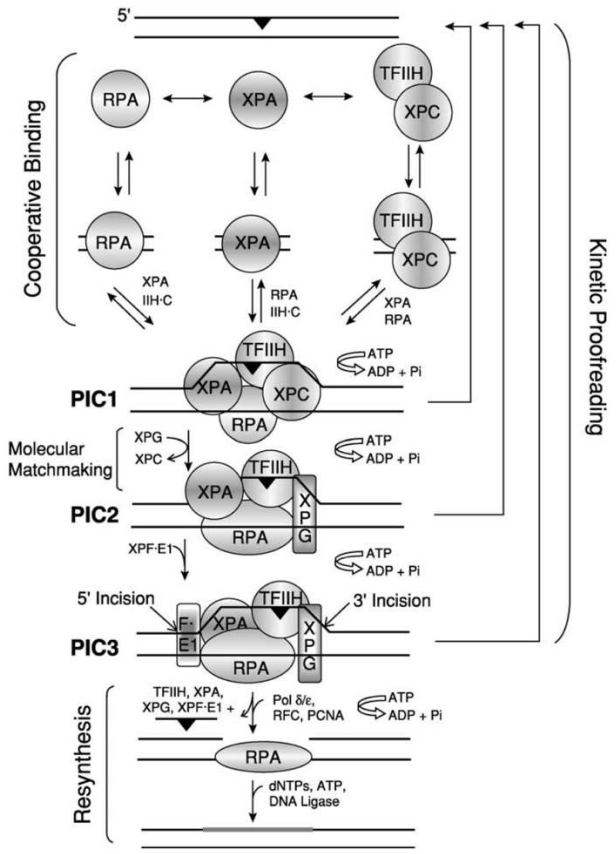

Figure 1.4: Proposed model for human nucleotide excision repair mechanism.

In this mechanism, transcription initiation and repair factor-II human (TFIIH) or Xerodema Pigmentosum A (XPA) or replication protein A (RPA) recognizes the damage randomly but in a cooperative manner to create an unstable machinery at the damage site. With the use of ATP molecule, TFIIH protein unwinds the helix, thereby creating pre-

incision complex 1 (PIC1), which is followed by the action of molecular matchmaker, XPC, which is substituted by XPG protein; resulting in the formation of pre-incision complex 2 (PIC2). With an additional ATP hydrolysis by TFIIH, XPF-ERCC1 complex associates with PIC2 complex, which results in the formation of pre-incision complex 3 (PIC3). At this stage, XPF-ERCC1 and XPG proteins incise the DNA lesion strand from the short distances at 5'end and 3'end, respectively, resulting in the formation of single strand excised product, which is believed to trigger damage signaling pathways in the cell (Hanawalt & Spivak, 2008). The gap is then filled by DNA polymerase, and newly synthesized DNA is ligated by DNA ligase as the last step of nucleotide excision repair(Gaul & Svejstrup, 2021). Although the mechanism seems to be energetically cost as undamaged nucleotides are also removed in addition to damaged ones, the cost is thought to be compensated with the diversity of DNA damages repaired by NER mechanism (Sancar, 1993).

The disruptions to the key players of NER result in a variety of diseases such as Xeroderma pigmentosum if XP proteins (from XPA to XPG) are not functional in NER, Cockayne syndrome and trichothiodystrophy, which drive cells to skin cancer, developmental and neurological anomalies (Cleaver et al., 2009; Schärer, 2013). Moreover, NER has been revealed to be tightly regulated by post-translational modifications and chromatin dynamics on the genome (Marteijn et al., 2014).

Nucleotide excision repair takes advantage of two different mechanisms, which are called transcription-coupled repair (TCR) and global genome repair (GGR) (Cleaver et al., 2009; Sabarinathan et al., 2016).

1.1.4 Global Genome Repair (GGR)

Global genome repair system constantly checks for any type of helix distortion as a consequence of abnormal base-pairing in both strands of DNA (Marteijn et al., 2014). The following Figure 1.5 summarizes the GGR and TCR pathways.

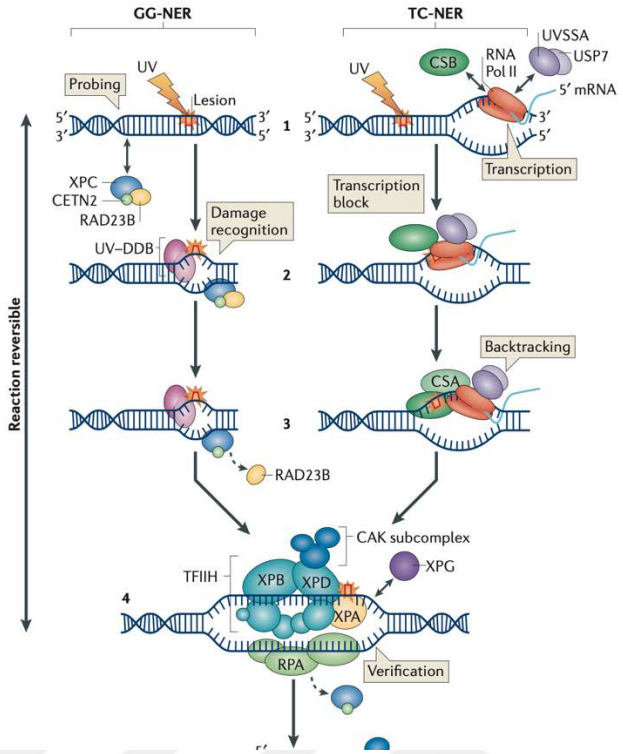

Figure 1.5: GGR and TCR pathways (Marteijn et al., 2014).

In GGR, the damage is first recognized by UV-DDB proteins, ultraviolet DNA damage binding protein 1 (UV-DDB1) and 2 (UV-DDB2) especially for CPD-type of damages as XPC is not successful alone to destabilize the helical structure without the aid of UV-DDBs. XPC is in complex with CETN2 and RAD23B proteins, which all surveille the genome for any type of helical distortion. As soon as the complex binds to the site of damage, RAD23B leaves the complex by dissociation (Marteijn et al., 2014).

1.1.5 Transcription-Coupled Repair (TCR)

As previously mentioned in Figure 1.5, RNA polymerase II (RNAP II) is responsible for synthesis of pre-mRNA molecule using the transcribed strand (TS) of a gene. However, in case of a DNA damage, which might occur in a particular gene being actively transcribed, RNAPII stalls at the site of lesion, that results in the recruitment of damage proteins (Pani & Nudler, 2017). Cockayne syndrome protein B (CSB), which constantly interacts with elongating RNAPII, is triggered to recruit Cockayne syndrome protein A (CSA) protein, which together with CSB recruits XPA-binding protein 2 (XAB2), TFIIS,

and high mobility group nucleosome binding domain containing protein (HMGN1) (Fousteri & Mullenders, 2008).

1.1.6 Excision Repair Sequencing (XR-seq)

Overview of XR-seq

Excision repair sequencing is an advanced technique to identify repair events performed by NER mechanism on genome scale. The technique basically involves the capture and sequencing of excision products, produced as a response to UV-light exposure. CPD and (6-4) photoproducts are immunoprecipitated by damage-specific antibodies. The technique has been applied to a variety of organisms including *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Arabidopsis thaliana*, mice, and mammalian cells (Adebali et al., 2017; Deger et al., 2019; J. Hu et al., 2015, 2019; W. Li et al., 2018; Oztas et al., 2018). The length of excision products differs based on the organism. For instance, mammals and plants produces an average of 26-27 nucleotide in length excision products, whereas the number is 23-24 nt for yeast and 12-13 nt for bacteria (J. Hu et al., 2019).

The level of TCR preference show variances across organisms. For instance, TCR preference of *Escherichia coli* has been reported to depend on MfD, which is a translocase (Adebali et al., 2017). Moreover, *Saccharomyces cerevisiae* has been found to prioritize TCR mechanism as a result of deep bioinformatic analyses (W. Li et al., 2018). Furthermore, although *Drosophila melanogaster* was previously known to be less dependent on TCR because of the lack of CSA and CSB orthologous in drosophila, XR-seq has revealed that TCR takes place in the cell probably with the help of CSA and CSB protein alternatives, repair factors, which need to be addressed (Deger et al., 2019). In the case of *Arabidopsis thaliana,* TCR has been demonstrated to be prioritized by plant cells and the relationship between TCR and circadian rhythm was concluded to be based on the activity of 10-30% of all genes (Oztas et al., 2018). In addition, chromatin structure, which affects the accessibility of damage sites by repair factors is also important for Arabidopsis to properly repair the damage by NER mechanism (Oztas et al., 2018). Moreover, CSA1 has the most contribution to TCR event compared to CSA2, whose absence has minor effect on TCR activity (Kaya et al., 2022). Last but not the least of

Arabidopsis NER, molecular mechanism is quite similar to that of humans except that Arabidopsis does not have an XPA ortholog, which is known to be essential for human NER. However, it is believed that another protein taking part the role of XPA in Arabidopsis might be present (Canturk et al., 2016).

XR-seq Workflow

According to XR-seq protocol published by Hu. J. et al. in 2019, the following workflow in Figure 1.6 is performed.

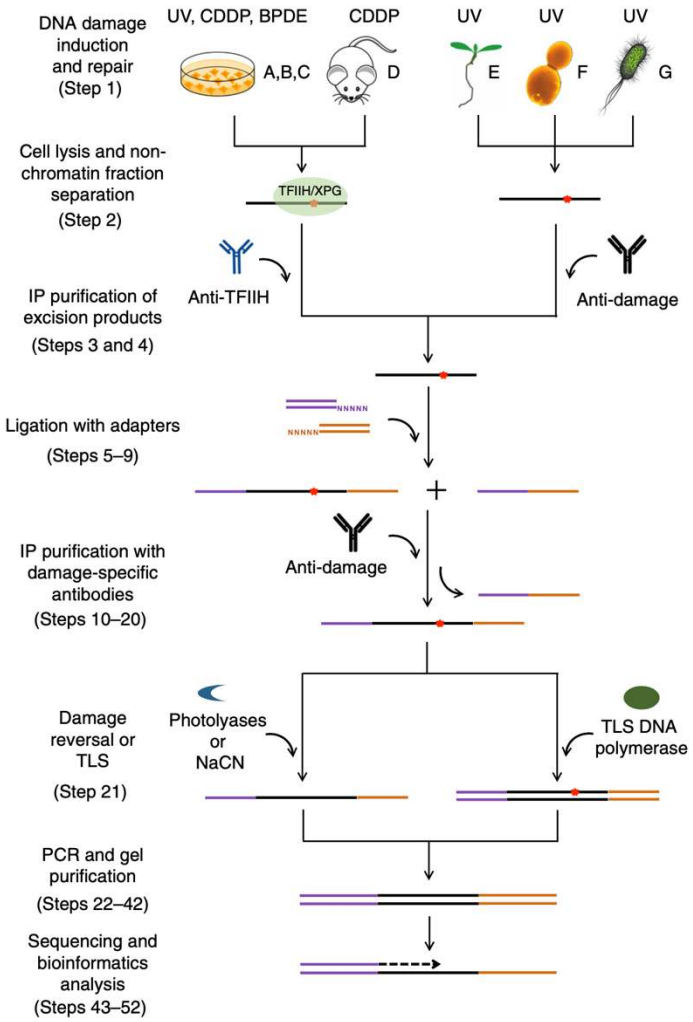

Figure 1.6: An overview of XR-seq protocol.

The experiment begins with the UV exposure of organism/tissue of interest. Next, cells are allowed to repair UV products, leading to the production of excision products in dark by photolyase enzyme, which utilizes light to function, is prevented to repair the

damage; otherwise, some excision products would be already repaired before biochemical capturing. Subsequently, cells are lysed based on the nature of outer cellular material specific to organism, which is followed by separation of non-chromatin fraction. Next, excision products are immunoprecipitated (IP) with damage-specific antibodies such as anti-CPD and anti-(6-4) PP. Then, the products are subjected to adapter ligation, which is necessary for sequences to be compatible for next-generation sequencing (NGS) platform. Next, a second IP step was necessary to eliminate adapter dimers and non-specific molecules. After completing the last IP step, immunoprecipitated photoproducts were subjected to photolyase treatment to reverse the damage. Then, PCR is performed to generate libraries of excision products, which is followed by gel purification. At this stage, the experimental protocol is completed, and the samples are subjected to NGS followed by bioinformatic analysis (J. Hu et al., 2019; Wu et al., 2022)

Advantages and Limitations of XR-seq

In XR-seq method, two IP steps are successful to exclude non-specific genomic DNA, which actually helps background noise of signal get reduced (J. Hu et al., 2019). On the other hand, since the method is based on the usage of antibody and requires gel purification, which might result in the loss of essential excision products. A new updated method of XR-seq called ATL-XR-seq utilizes one-side dA-tailing method, which eliminates adapter dimerization with no significant bias introduced to the data (Wu et al., 2022). Moreover, XR-seq is a powerful method as it is able to map a time-dependent dynamic process of nucleotide excision repair at single nucleotide resolution while some products are continuously degraded and others are getting produced. XR-seq, however, is not able to measure the absolute and true rates of repair on account of the differences in production and removal rates of excision products in different organisms (J. Hu et al., 2019). Lastly, one of the most critical advantage of XR-seq method lies in the power of unraveling TCR process in cells.

Future Aspects

In the future studies applying XR-seq, chromatin states of particular genomic regions and other repair specific factors in a variety of organisms could be widely studied to understand the effects of determinants of nucleotide excision repair mechanism. Moreover, the carcinogenesis might be investigated using damage-generating chemicals

and physical factors using XR-seq method. Furthermore, the consequences of drugs and toxins on DNA level might also be worked with this powerful method. Last but not the least, the essential players of excision repair could be determined by creating mutant or transgenic versions of corresponding genes to better understand their roles in the system (J. Hu et al., 2019).

This chapter of the thesis specifically focuses on the elucidation of nucleotide excision repair map across Arabidopsis tissues using XR-seq method. To do so, bioinformatic analysis was carried out to account for overall differences and specifically transcription-coupled repair using CPD and 6-4 photoproduct data on wild-type *Arabidopsis thaliana* exposed to UV light.

1.2 Methodology

According to XR-seq bioinformatics pipeline published by Hu. J. et al. in 2019, the following steps are required to complete the analysis of XR-seq data.

All of the analysis was conducted on High Performance Computing (HPC) cluster, which is based on Linux operating system.

Before starting, the following programs are required to perform proper command execution throughout the analysis: Bowtie2, cutadapt, sratoolkit, samtools, bedtools, ucsc tools, deeptools (Kodama, 2012; Langmead & Salzberg, 2012; Li et al., 2009; Martin, 2011; Quinlan, 2014; Ramirez, 2016). The programs were requested to be installed from Information Technology Center of the university.

First of all, after sequencing the samples, the sequencing service provides a FASTQ to the customer. The following lines represent the basic structure of a FASTQ file.

```
@K00269:126:HTK5LBBXX:7:1101:1468:1367 1:N:0:ACAGTG
CTCCAGATTAAATAGCTGGAATTCTCGGGTGCCAAGGAACTCCAGTCACA
+
AAAFFJAFJJFF-FJJJJJ<7AJJJJJJJJJJJ<FAJJFJJJFJJJ<F-F
```

The first line starting with @ symbol represents the sequence identifier of a particular read together with information related to the run and the cluster. The second line is a sequence of bases, that are measured and inferred from the samples while the third line is just a separator. The last line corresponds to the quality score of each base read, that is represented with ASCII characters.

In order to work a job on a cluster, it is conventional to create shell (.sh) file, that is comprised of the following information:

```
#!/bin/bash
#
#
# -= Resources =-
#SBATCH --job-name=
#SBATCH --nodes=
```

```
#SBATCH --ntasks-per-node=
#SBATCH --mem
#SBATCH --partition=
#SBATCH --qos=
#SBATCH --account =
#SBATCH --time =
#SBATCH --output =
#SBATCH --mail-type =
#SBATCH --mail-user =
# -Your code comes here-
# -End of the shell file-
```

Depending on the availability of nodes allocated to cluster users and the estimated time required to finish a particular job, the above lines are filled.
Moreover, one can easily type the following command to run a job on the cluster:

```
sbatch yourfile.sh
```

The job is queued right after the command execution until the queue finishes. Araport11 has been selected as a reference genome and downloaded to the home folder for downstream applications.

To start with processing fastq files, an sbatch file was created and the following commands of module were executed inside the file.

```
module load cutadapt/1.9.1
module load bowtie2/2.4.1
module load sratoolkit/2.10.8
module load bedtools/v2.29.1
module load samtools/1.10
module load deepTools/3.5.0
```

This step is essential as the commands to be given need those modules to be downloaded beforehand. It is necessary to emphasize that “\” symbol at the end of each

line of code is depicted that this is the end of line and it should not be included in the original code. Next, within the same shell file, the following lines are written:

```
   echo “If your fastq file is named as “your_sample.fastq”,
define a variable called “SAMPLE” as below.

SAMPLE= your_sample

   echo “Cut the adapters for each captured XR-seq read”

cutadapt -a \
TGGAATTCTCGGGTGCCAAGGAACTCCAGTNNNNNNACGATCTCGTATGCCGTCTTCTG
CT\TG  -o ${SAMPLE}_cutadapt.fastq your_sample.fastq

   echo “Index your reference genome using bowtie2’s build
command”

bowtie2-build Araport11_genome.fa Araport_11_index

echo "Align your reads to the reference genome Araport11,
that is previously index using Bowtie2 module."

bowtie2 -p 4 -x Araport11_index -U ${SAMPLE}_cutadapt.fastq
-S ${SAMPLE}_cutadapt.sam

echo "Convert the sam file to .bam format."

samtools view -q 20 -b -o ${SAMPLE}_cutadapt.bam
${SAMPLE}_cutadapt.sam

echo "Convert the bam file to .bed format."

bedtools bamtobed -i ${SAMPLE}_cutadapt.bam >
${SAMPLE}_cutadapt.bed

echo "Sort the bed file with a -u parameter to remove
duplicates."

sort -u -k1,1 -k2,2n -k3,3n ${SAMPLE}_cutadapt.bed
>${SAMPLE}_cutadapt_sorted.bed

echo "Count number of mapped reads and write it to a txt
file"

grep -c "^" ${SAMPLE}_cutadapt_sorted.bed
>${SAMPLE}_cutadapt_sorted_readCount.txt
```

```
echo "Generate the read length distribution by executing
the following line of command:"

awk '{print $3-$2}' >${SAMPLE}_cutadapt_sorted.bed | sort
-k1,1n | uniq -c | sed 's/\s\s*/ /g' | awk '{print
$2"\t"$1}'

echo "To generate nucleotide distribution of sequences at a
certain read length, first retrieve sequence read of a
certain length(n=27 for Arabidopsis)"

awk '{ if ($3-$2 == 27) { print } }' >
${SAMPLE}_cutadapt_sorted.bed > \
>${SAMPLE}_cutadapt_sorted_27.bed

echo "Convert the sequences from bed to FASTA format:"

bedtools getfasta -fi Araport11_genome.fa -bed
>${SAMPLE}_cutadapt_sorted_27.bed -fo
>${SAMPLE}_cutadapt_sorted_27.fa -s

echo "Split reads into plus and minus strands:"

awk '{if($6=="+"){print}}' >${SAMPLE}_cutadapt_sorted.bed >
>${SAMPLE}_cutadapt_sorted_plus.bed

awk '{if($6=="-"){print}}' >${SAMPLE}_cutadapt_sorted.bed >
>${SAMPLE}_cutadapt_sorted_minus.bed

echo "Generate BedGraph files using bed files:"

bedtools genomecov -i >${SAMPLE}_cutadapt_sorted_plus.bed -
g Araport11_genome.fa.fai -bg -scale  $(cat
>${SAMPLE}_cutadapt_sorted_readCount.txt |  awk '{print
1000000/$1}') > ${SAMPLE}_cutadapt_sorted_plus.bdg

bedtools genomecov -i ${SAMPLE}_cutadapt_sorted_minus.bed -
g Araport11_genome.fa.fai -bg -scale  $(cat
${SAMPLE}_cutadapt_sorted_readCount.txt |  awk '{print
-1000000/$1}') > ${SAMPLE}_cutadapt_sorted_minus.bdg

echo "Convert bedgraph to bigwig"

bedGraphToBigWig ${SAMPLE}_cutadapt_sorted_plus.bdg
Araport11_genome.fa.fai ${SAMPLE}_cutadapt_sorted_plus.bw

bedGraphToBigWig ${SAMPLE}_cutadapt_sorted_minus.bdg
Araport11_genome.fa.fai ${SAMPLE}_cutadapt_sorted_minus.bw
```

These bigwig files are used to visualize the signal from XR-seq data using Integrative Genomics Viewer (IGV) across the genome (Robinson et al., 2011).

To be able to use Araport11_GFF3_genes in bed format, the following python script called "`gff_to_bed.py`" is run within the py37 virtual environment using the following command on the command line of shell: `python gff_to_bed.py`.

```
script, infile1 = argv

outputname = '%s.bed' % infile1

outfile = open(outputname, "w")

file1 = open(infile1, "r")

lines1 = file1.readlines()

for line in lines1:

    if '#' not in line:

        content1 = line.split("\t")

        ID=content1[8].split(';')

        my_ID = str(ID[0]).split('=')

        if content1[2]=='mRNA':

            result = content1[0] + "\t" + content1[3] +
"\t" + content1[4] + "\t" + str(my_ID[1]) + "\t" +
content1[5] + "\t" + content1[6] + "\n"

            outfile.write(result)

file1.close()

print("Finished")
```

The output is called `Araport11_GFF3_genes.bed` to be used for bedtools intersect command next.

```
echo "Count read values for transcribed (TS) and
nontranscribed (NTS) strands"
```

```
bedtools intersect -sorted -a Araport11_GFF3_genes.bed -b\
${SAMPLE}_cutadapt_sorted.bed -wa -c -S -F 0.5 >\
${SAMPLE}_cutadapt_sorted_TScount.txt

bedtools intersect -sorted -a Araport11_GFF3_genes.bed -b\
${SAMPLE}_cutadapt_sorted.bed -wa -c -s -F 0.5 >\
${SAMPLE}_cutadapt_sorted_NTScount.txt
```

Next, TS and NTS count files are converted to reads per kilobase million (RPKM) values for both TS and NTS using the following python script called "`calculate_RPKM.py`", which is run within the py37 virtual environment using the following command on the command line of shell:

```
from sys import argv

script, infile1, infile2 = argv

outname = '%s_RPKMs' % infile1

outfile = open(outname, "w")

file1 = open(infile1, "r") #### Bedtools intersect output
file

lines1 = file1.readlines()

file2 = open(infile2, "r") #### XR-seq bed file

lines2 = file2.readlines()

total_read_count = 0

for i in lines2:

    total_read_count += 1

print(total_read_count)

for line in lines1:

    content = line.split("\t")

    gene_length = int(content[2]) - int(content[1])

    occurrence = int(content[6])

    ID = content[3]
```

```
    RPKM = float(occurrence) / (float(gene_length) / 1000)
/ (float(total_read_count) / 1000000)

    result = str(content[0]) + "\t" + str(content[1]) + "\t"
+ str(content[2])  + "\t" + str(ID) + "\t" + str(content[4])
+ "\t" + str(content[5]) + "\t" + str(RPKM) + "\n"

    outfile.write(result)

python calculate_RPKM.py
${SAMPLE}_cutadapt_sorted_TScount.txt\
${SAMPLE}_cutadapt_sorted.bed

python calculate_RPKM.py
${SAMPLE}_cutadapt_sorted_NTScount.txt\
${SAMPLE}_cutadapt_sorted.bed
```

In order to visualize the signals of XR-seq reads on gene body of all reads mapped to genes, deeptools were utilized. The following lines were written within the same shell file.

```
Out = genes_sorted_clustered5kb

echo "Sorting BED"

bedtools sort -i Araport11_GFF3_genes.bed > ${Out}.bed\

echo "Clustering genes"

bedtools cluster -i ${Out}.bed > ${Out}_clustered5kb.bed\

echo "Eliminating clustered genes"

python eliminate_clustered.py ${Out}_clustered5kb.bed\

echo "Separating strands"

awk '$6 == "+" {print
$1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7}'\
${Out}_clustered5kb_Mod_eliminated.bed > \
${Out}_clustered5kb_eliminated_plus.bed

awk '$6 == "-" {print
$1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7}'\
${Out}_clustered5kb_Mod_eliminated.bed > \
${Out}_clustered5kb_eliminated_minus.bed

echo "Extending TSS and TES by 2.5 kb"
```

```
awk '{print $1"\t"($2 - 2501)"\t"($2 + 2500)"\t"$4"\t"$5"\t
\ "$6"\t"$7}' ${Out}_clustered5kb_eliminated_plus.bed > \
${Out}_clustered5kb_eliminated_plus_TSS_2.5kbextended.bed

awk '{print $1"\t"($3 - 2500)"\t"($3 + 2501)"\t
"$4"\t"$5"\t \ "$6"\t"$7}'
${Out}_clustered5kb_eliminated_plus.bed > \
${Out}_clustered5kb_eliminated_plus_TES_2.5kbextended.bed

awk '{print $1"\t"($2 - 2501)"\t"($2 + 2500)"\t"$4"\t"$5"\t
\ "$6"\t"$7}' ${Out}_clustered5kb_eliminated_minus.bed > \
${Out}_clustered5kb_eliminated_minus_TES_2.5kbextended.bed

awk '{print $1"\t"($3 - 2500)"\t"($3 + 2501)"\t"$4"\t"$5"\t
\ "$6"\t"$7}' ${Out}_clustered5kb_eliminated_minus.bed > \
${Out}_clustered5kb_eliminated_minus_TSS_2.5kbextended.bed

echo "All done."
```

The python script called `eliminate_clustered.py` is the following:

```
from sys import argv

script, infile1 = argv

outname = 'genes_sorted_clustered5kb_eliminated.bed'

outfile1 = open(outname, "w")

file1 = open(infile1, "r")

lines1 = file1.readlines()




for i in range(0, len(lines1)-1):

    line = lines1[i]

    previous_line = lines1[i-1]

    next_line = lines1[i+1]

    elems = line.split("\t")

    elems_prev = previous_line.split("\t")

    elems_next = next_line.split("\t")
```

```
    if not elems[6] == elems_prev[6] and not elems[6] ==
elems_next[6]:

        results = line

        outfile1.write(results)

file1.close()

echo "Compute matrix for gene body"

computeMatrix scale-regions -S
${SAMPLE}_cutadapt_bam_plus_merged.bw ${SAMPLE}_cutadapt
_bam_minus_merged.bw -R
genes_sorted_clustered5kb_eliminated_plus.bed
genes_sorted_clustered5kb_eliminated_minus.bed -b 2500 -a
2500 -o XRseq_on_gene_bodies

echo “Sort your bam file of XR-seq data using samtools sort
command”

samtools sort ${SAMPLE}_cutadapt.bam -o \
${SAMPLE}_cutadapt_sorted.bam

echo “Index your sorted bam file of XR-seq data using
samtools index command”

samtools index ${SAMPLE}_cutadapt _sorted.bam

echo "Separate plus and minus reads"

samtools view -bh -F 16 ${SAMPLE}_cutadapt_sorted.bam >\
${SAMPLE}_cutadapt_bam_plus.bam

samtools view -bh -f 16 ${SAMPLE}_cutadapt_sorted.bam >\
${SAMPLE}_cutadapt_bam_minus.bam

echo "Merge bam files"

samtools merge ${SAMPLE}_cutadapt_bam_plus_merged.bam\
${SAMPLE}_cutadapt_bam_plus.bam

samtools merge ${SAMPLE}_cutadapt_bam_minus_merged.bam\
${SAMPLE}_cutadapt_bam_minus.bam

echo "Index bam files using samtools index command"

samtools index ${SAMPLE}_cutadapt_bam_plus.bam

samtools index ${SAMPLE}_cutadapt_bam_minus.bam
```

```
samtools index ${SAMPLE}_cutadapt_bam_plus_merged.bam

samtools index ${SAMPLE}_cutadapt_bam_minus_merged.bam

echo "Convert plus and minus merged BAM files to BW files"

bamCoverage -b ${SAMPLE}_cutadapt_bam_plus_merged.bam -o \
${SAMPLE}_cutadapt_bam_plus_merged.bw

bamCoverage -b ${SAMPLE}_cutadapt_bam_minus_merged.bam -o \
${SAMPLE}_cutadapt_bam_minus_merged.bw

echo "Plot profile"

plotProfile --matrixFile XRseq_on_gene_bodies --plotType
lines --plotTitle "XR-seq reads on gene bodies" --
regionsLabel "Plus strand" "Minus strand" --samplesLabel
"Plus XR-seq reads" "Minus XR-seq reads" --numPlotsPerRow 2
--outFileName XRseq_on_gene_bodies.png
```

Moreover, since XR-seq is also an indirect and good RNA data due to the presence of TCR in Arabidopsis, a conventional differential gene expression analysis method, DESeq2, was performed on XR-seq data (Love et al., 2014). For this purpose the following R script was run to understand significant XR-seq count differences of genes across different tissues.

```
library(DESeq2)

#countData is a dataframe composed of RPKM values for each
gene and each tissue.

countData <- read.csv("tissues_counts.csv", header = TRUE,
sep = ";")

#metaData is a dataframe composed of id and tissue columns
for each sample(identified in id column e.g shoot1 and
shoot2) and sample set(containing replicates in tissue
column e.g shoot, root, seedling etc.)

metaData <- read.csv("tissues_metadata.csv", header = TRUE,
sep = ";")

dds <- DESeqDataSetFromMatrix(countData=countData,

                              colData=metaData,

                              design=~tissue, tidy = TRUE)
```

```
dds$tissue <- relevel( dds$tissue, "tissue2" )

dds <- DESeq(dds)

res <- results(dds)

head(results(dds, tidy=TRUE))

summary(res)

mcols(res, use.names=TRUE)

res <- res[order(res$padj),]

head(res)

tiff("volcano_plot.tiff", units="in", width=8, height=8,
res=300)

par(mfrow=c(1,1))

# Generate a volcano plot

with(res, plot(log2FoldChange, -log10(pvalue), pch=20,
main="Volcano plot for tissue1 vs. tissue2", xlim=c(-
10,10)))

# Add colors to genes given as dots: color blue if padj <
0.01, color red if log2foldchange>1 and padj < 0.05)

with(subset(res, padj <0.01 ), points(log2FoldChange, -
log10(pvalue), pch=20, col="blue"))

with(subset(res, padj <0.01 & abs(log2FoldChange)>2),
points(log2FoldChange, -log10(pvalue), pch=20, col="red"))

dev.off()

write.csv(as.data.frame(res),
file="tissue1_vs_tissue2_DESeq2_results.csv")

install.packages("ggplot2")

vsdata <- vst(dds, blind=FALSE)

tiff("PCA_tissue1_vs_tissue2_CPD.tiff", width=8, height=8,
units="in",res=300)

# Plot PCA plot for each sample of each tissue
```

```
plotPCA(vsdata, intgroup="tissue")

library(ggplot2)

pcaData <- plotPCA(vsdata, intgroup=c("tissue"),
returnData=TRUE)

percentVar <- round(100 * attr(pcaData, "percentVar"))

f <- ggplot(pcaData, aes(PC1, PC2, color=tissue)) +

  geom_point(size=4) +

  xlim(-15, 15) +

  ylim(-2, 2) +

  xlab(paste0("PC1: ",percentVar[1],"% variance")) +

  ylab(paste0("PC2: ",percentVar[2],"% variance")) +

  geom_text(aes(label=name),vjust=2) + theme(plot.title =
element_text(hjust = 0.5))+scale_color_discrete(name="")

f

f + labs(title = "PCA for tissue1 vs. tissue2")

dev.off()

# Create a heatmap for samples

install.packages("pheatmap")

library(pheatmap)

tiff("heatmap_tissue1_vs_tissue2.tiff", width=8, height=8,
units="in",res=300)

dds <- DESeq(dds)

# log transform the dataset called dds, which is DESeqed.

mrw <- rlog(dds, blind=F)

topGenes <- head(order(rowVars(assay(rld)),
decreasing=T),10000)

matrix <- assay(mrw)[ topGenes, ]
```

```
matrix <- rowMeans(matrix) - matrix

annotation_data <- as.data.frame(colData(mrw)[c("id")])

pheatmap(matrix, annotation_legend = FALSE, show_rownames =
F, angle_col = 45)

head(colData(mrw))

tail(colData(mrw))

dev.off()
```

Moreover, TS/NTS values were calculated using the following python script called `calculate_TSoverNTS.py` to account for relative abundance of excision products originated from different strands by calling on the terminal:

```
python calculate_TSoverNTS.py\
${SAMPLE}_cutadapt_sorted_NTScount.txt_RPKMs.bed \
${SAMPLE}_cutadapt_sorted_TScount.txt_RPKMs.bed.

from sys import argv

script, infile1, infile2 = argv

file1 = open(infile1, "r") #### NTS bed file

lines1 = file1.readlines()

file2 = open(infile2, "r") #### TS bed file

lines2 = file2.readlines()

if 'plus_gene' in infile1:

    GeneStrand = 'plus_gene'

else:

    GeneStrand = 'minus_gene'

outfile = open(str(GeneStrand) + "_" + "TSoverNTS", "w")

for i in range(0, len(lines1)):

    line_NTS = lines1[i]

    line_TS = lines2[i]
```

```
    NTS_ID = line_NTS.split("\t")[3]

    TS_ID = line_TS.split("\t")[3]

    NTS_RPKM = line_NTS.split("\t")[6]

    TS_RPKM = line_TS.split("\t")[6]

    if 'chrC' and 'chrM' not in line_NTS.split("\t")[0]:

        if NTS_ID == TS_ID:

            if float(NTS_RPKM) != 0.0:

                TSoverNTS = float(TS_RPKM) /
float(NTS_RPKM)

                result = str(NTS_ID) + "\t" +
str(TSoverNTS) + "\n"

                outfile.write(result)
```

1.3 Results and Discussion

After completing the whole analysis, different aspects of the data need to be demonstrated to justify and infer the findings. To this end, first of all, read length distribution of XR-seq data for each sample was plotted in Figure 1.7.

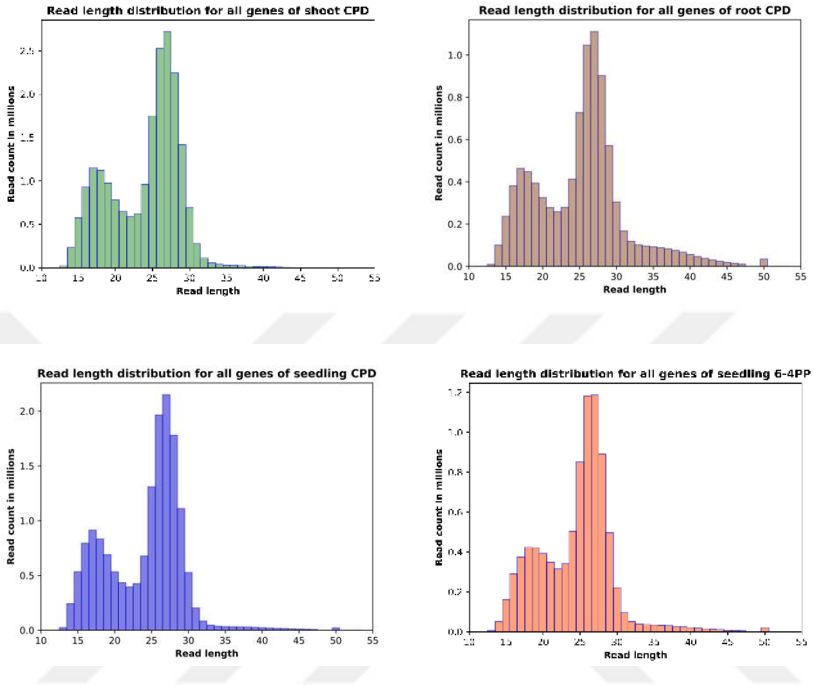

Figure 1.7: Read length distribution of shoot CPD, root CPD, seedling CPD, and seedling 6-4 photoproducts (from top left to bottom right).

All samples demonstrated a maximum read count in 27 nt length, which was previously validated (Canturk et al., 2016). Although the read count numbers in the y axis for each sample were not normalized, it is relatively possible to see the maximum number of read counts in shoot CPD and seedling CPD data.

Moreover, in order to justify that the origin of the data really comes from XR-seq reads, the following plots for each sample was plotted in Figure 1.8.

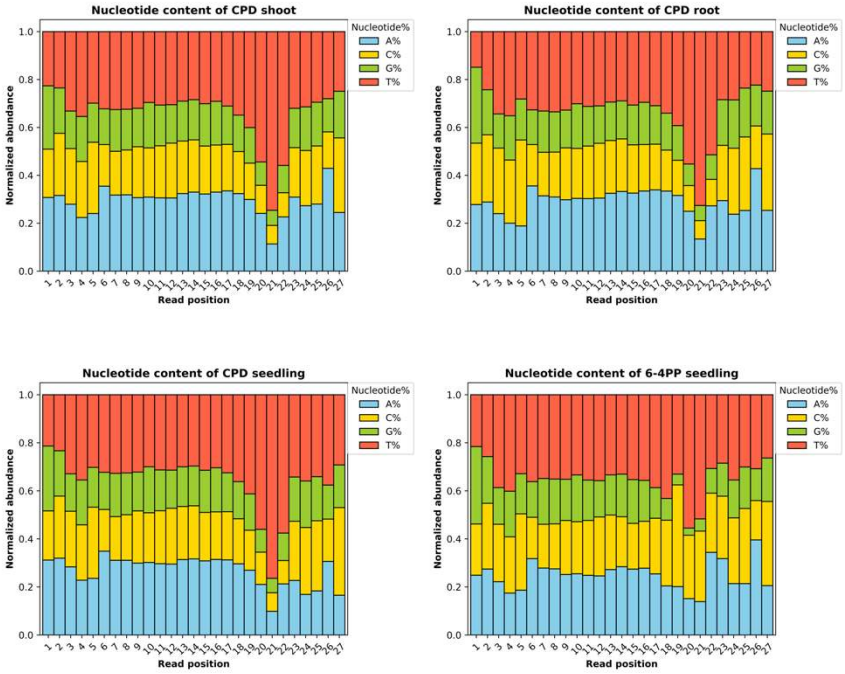

Figure 1.8: Nucleotide content of shoot CPD, root CPD, seedling CPD, seedling 6-4 photo-products (from top left to bottom right).

Here, since the maximum number of read counts was observed at maximum of 27 nt length, only these 27 nt length data were analyzed to reveal single nucleotide content of those photoproducts. It is clear from the figure that 5-6 nucleotide upstream of the reads were enriched in TT dinucleotide, which supports the fact that these are potential excision product sites due to the chemical nature of TT dinucleotide. For CPD data, TT frequency was reported to be the most at these sites as well as for seedling 6-4 PP. Interestingly, cytosine frequency was dominated by thymine frequency at these sites of CPD data compared to 6-4PP; in other words, 6-4 PP data was also enriched in CC sites (shown in yellow) in this position, which is in comparable level with TT dimers (shown in red) for this sample. As a result, the method was concluded to work well in collecting excision products for each sample.

Nucleotide excision repair in Arabidopsis was previously shown to be coupled to transcription (Oztas et al., 2018). Moreover, our study was also conducted to understand the dominance of TCR in different tissues of *Arabidopsis thaliana*. To investigate this, TS/NTS ratio of each gene associated with XR-seq read was quantitatively calculated and plotted in histograms in Figure 1.9 below.

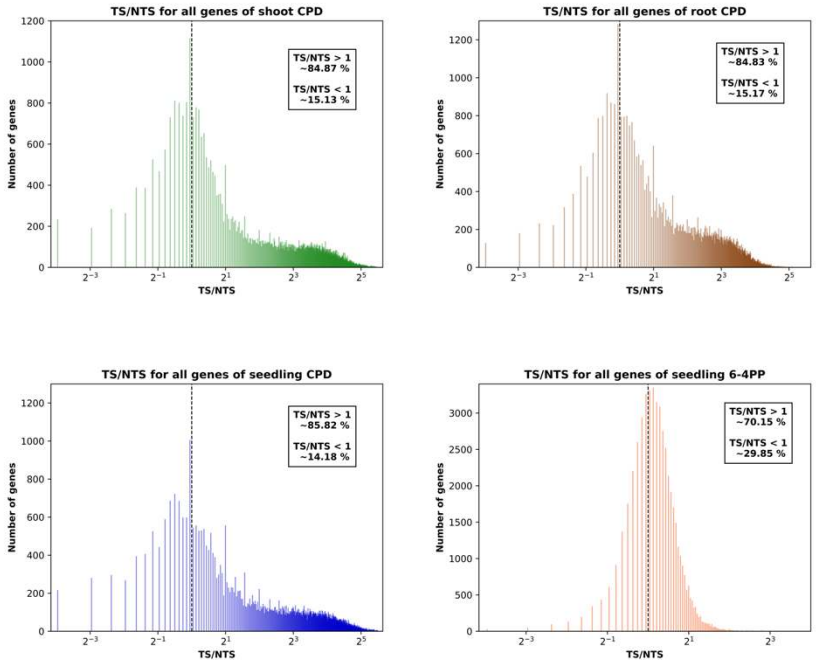

Figure 1.9: TS/NTS ratio of genes covered by shoot CPD, root CPD, seedling CPD, and seedling 6-4 photo-products (from top left to bottom right).

TS/NTS ratios shown in Figure 1.9 demonstrated the relative distributions where the reads of genes on the left side of the plots are assigned to show NTS abundance over TS whereas reads of genes on the right side of the plots exhibits TS is more pronounced compared to NTS, which may imply that TCR event is preferentially selected. Although there is no statistical test conducted, relative abundances of TS/NTS vs NTS/TS ratios might be able to show tissue-specific relative TCR abundance by comparing the relative TS/NTS ratio. For instance, genes from CPD samples were approximately 85% accumulated at the right side of the plots while the same percentage was 70% in the case of 6-4 PP. CPD data might be speculated that TCR is more pronounced in CPD photoproducts compared to 6-4 PP. The reason why this is the case could be due to chemical nature of photoproducts, which might differ in the rates of getting recognized by corresponding repair enzymes. Moreover, another critical finding was the high percentage of TS/NTS ratio in CPD root sample, which is interesting as no UV light was subjected to this tissue in nature. The presence of these photoproducts might be due to some chemicals present in soil, which might lead to formation of DNA damage products similar to ones captured by CPD and 6-4 PP antibodies.

In addition, DESeq2 method was recruited to account for significant differences of genes across tissues. Since XR-seq, due to its TCR reflection, is a good method to

have an idea of RNA profile of the system, DESeq2 method, which is commonly used to find out differential gene expression, was applied. Read count values for each gene of tissue was subjected to DESeq2, and the XR-seq profiles of the following genes were found to be significantly different across tissues compared. The following genes are the ones highly pronounced in root vs shoot samples.

AT5G04140 is a gene encoding for ferredoxin dependent glutamate synthase, whose expression was reported to be induced by light while AT4G16990 was reported to encode for a disease resistance protein (Cui et al., 2020; Staal et al., 2008). The excision products from these genes were reported to be significantly higher in shoot tissue compared to root.

Moreover, read count values for each gene of tissue was subjected to DESeq2, and the XR-seq profiles of the following genes were found to be significantly different across tissues compared; thus TS/NTS ratio of significant genes were demonstrated in Figure 1.10.

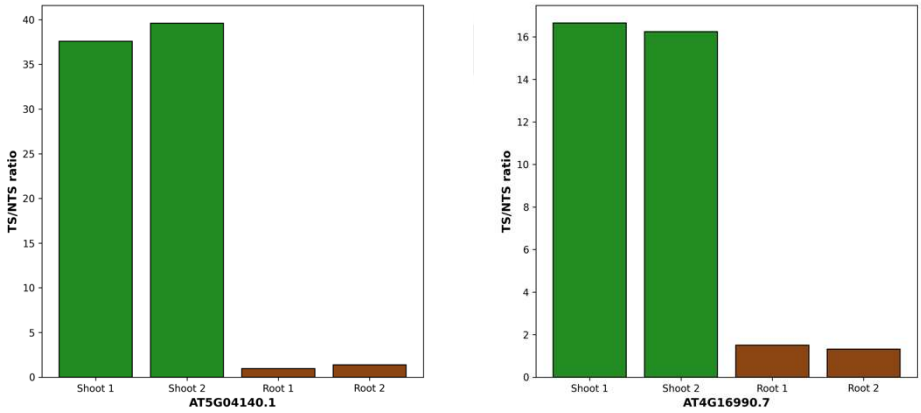

Figure 1.10: TS/NTS ratio of AT5G04140 and AT4G16990 genes across shoot CPD and root CPD photoproducts.

Figure 1.11 and 1.12 are the visual representation of read distribution across the corresponding genes.

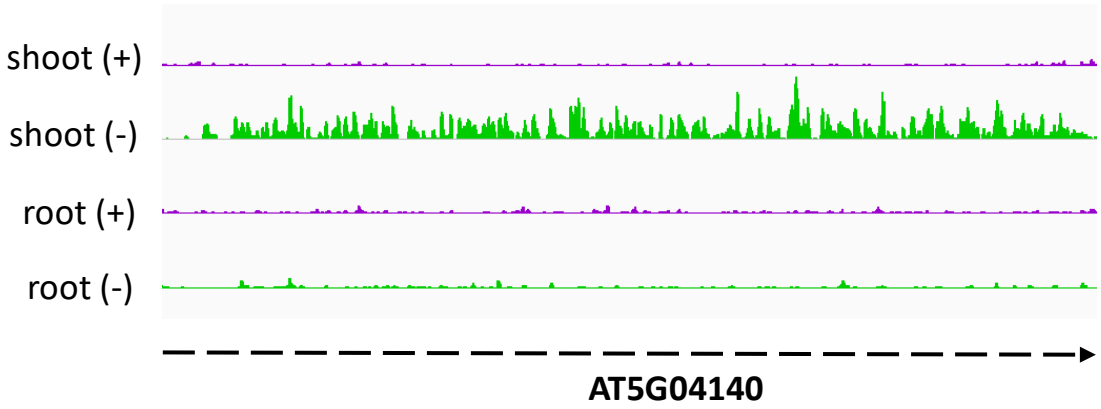

Figure 1.11: IGV screenshot of AT5G04140 XR-seq reads across the genome.

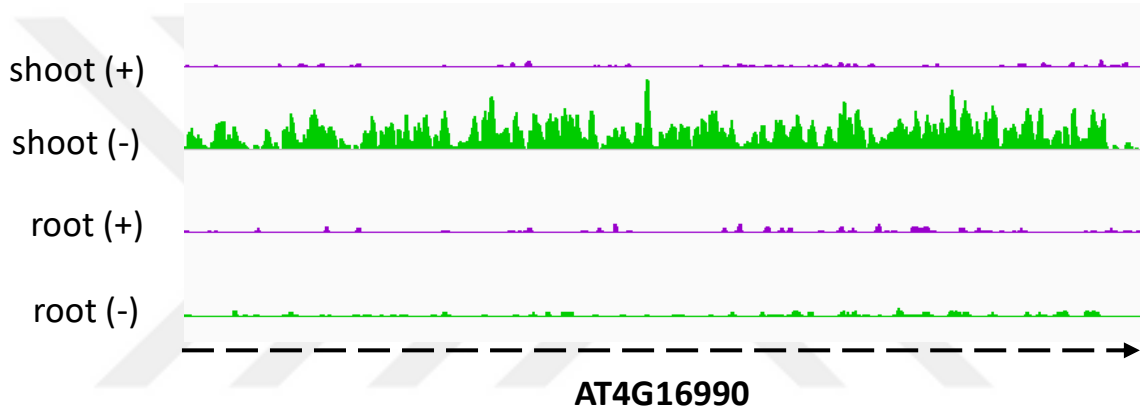

Figure 1.12: IGV screenshot of AT4G16990 XR-seq reads across the genome.

Figure 1.10 directly shows the TCR effect on these significantly different excision repair of genes across different tissues as a higher TS/NTS ratio strongly suggest the presence of TCR. Moreover, it is clear from Figure 1.11 and Figure 1.12 that the signal on the minus strand of genes imply the direction of gene expression that is shown in dashed arrows. However, it should not be understood from Figure 1.10 that any general trend of any sample might be concluded over the other as if shoot TCR is more pronounced compared to that of root, instead it is only a gene whose excision product count is significantly different across those samples.

Moreover, AT5G06839 is known to encode for a bZIP transcription factor family protein, whose XR-seq reads were significantly accumulated in root tissue compared to shoot whereas significantly important AT1G01320 (also known as REC1) in shoot compared to root encodes for a protein, which aids to adjust the size of chloroplast compartment in plant cells (Jakoby et al., 2002; Larkin et al., 2016).

Figure 1.13 shows TS/NTS ratio of significantly different excision repair of genes AT5G06839 and AT1G01320.

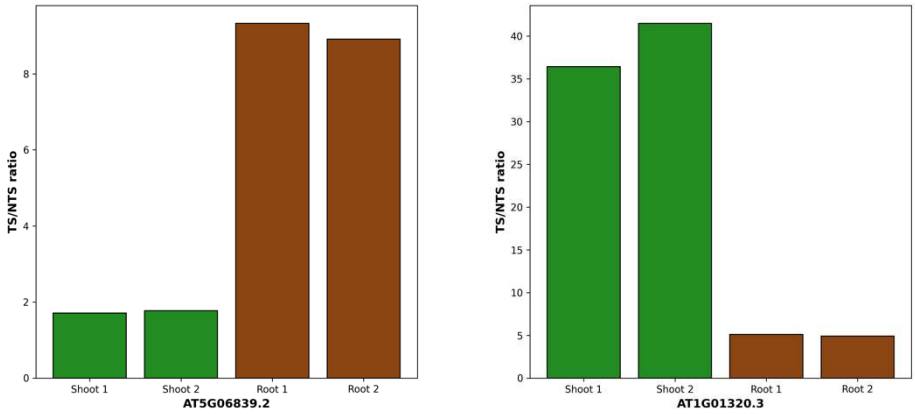

Figure 1.13: TS/NTS ratio of AT5G06839 and AT1G01320 genes across shoot CPD and root CPD photoproducts.

In addition, Figure 1.14 and 1.15 are the visual representation of read distribution across the corresponding genes.

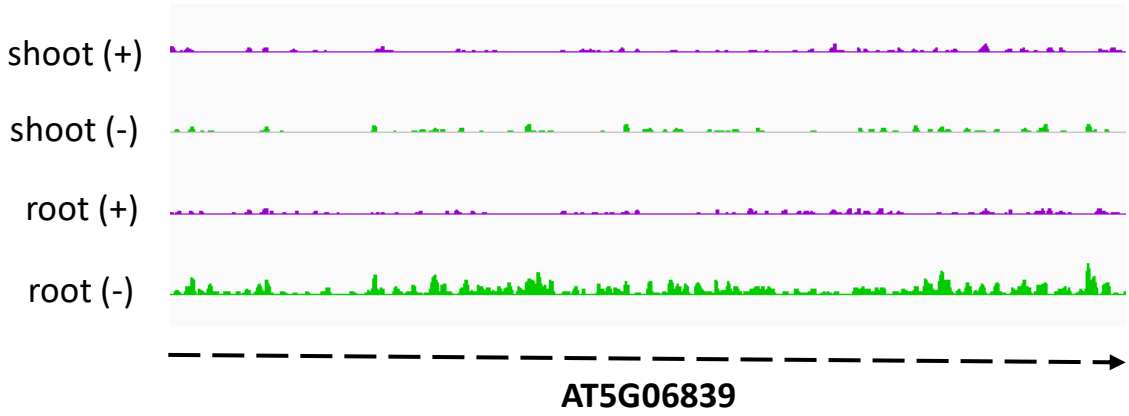

Figure 1.14: IGV screenshot of AT5G06839 XR-seq reads of shoot and root tissues across the genome.

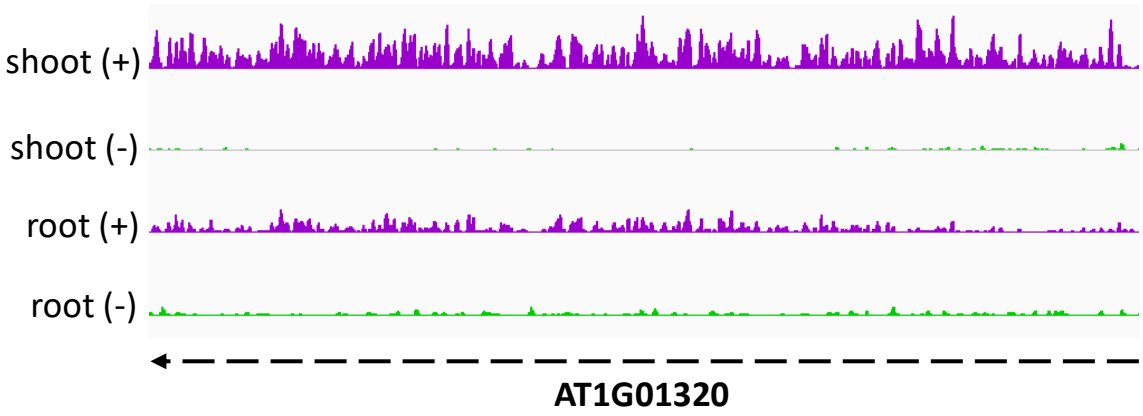

Figure 1.15: IGV screenshot of AT1G01320 XR-seq reads of shoot and root tissues across the genome.

Moreover, AT4G10060 is a gene encoding for glucosylceramidase enzyme, which is responsible for the hydrolysis of long acyl chain glucosylceramides while AT4G28080 (also known as REC2) works with REC1 and REC3 partners to tune the size of chloroplast compartment (Dai et al., 2020; Larkin et al., 2016).

Figure 1.16 shows TS/NTS ratio of significantly different excision repair of genes AT4G10060 and AT4G28080.

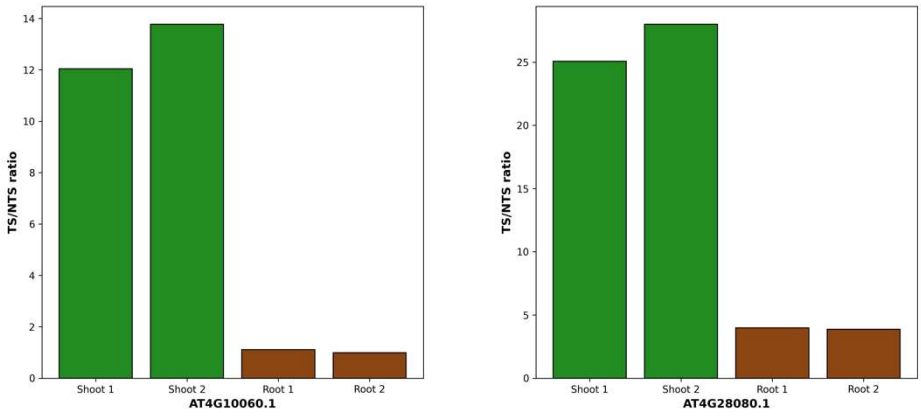

Figure 1.16: TS/NTS ratio of AT4G10060 and AT4G28080 genes across shoot CPD and root CPD photoproducts.

Figure 1.17 and 1.18 are the visual representation of read distribution across the corresponding genes.

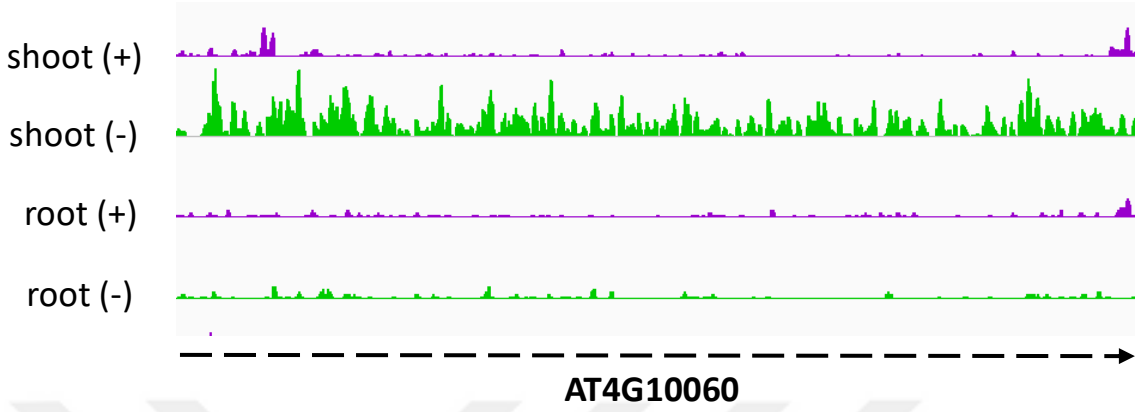

Figure 1.17: IGV screenshot of AT4G10060 XR-seq reads of shoot and root tissues across the genome.

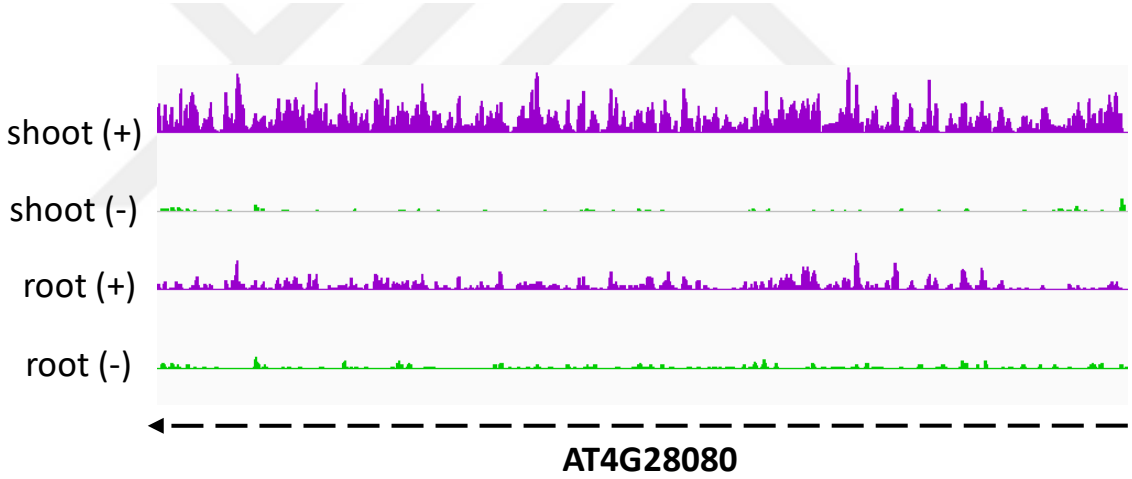

Figure 1.18: IGV screenshot of AT4G28080 XR-seq reads of shoot and root tissues across the genome.

Furthermore, AT4G19530, whose excision products were enriched in shoot, is a gene encoding for a TIR-NB-LRR resistance protein whereas AT4G34830, whose excision products were enriched in shoot, is responsible for stabilization of rbcL mRNA (Johnson et al., 2010; Swiderski et al., 2009).

Figure 1.19 demonstrates TS/NTS ratio of significantly different excision repair of genes AT4G19530 and AT4G34830.

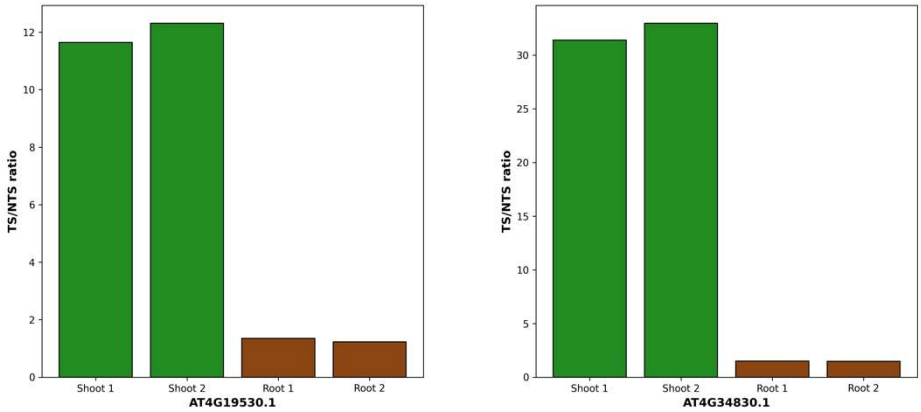

Figure 1.19: TS/NTS ratio of AT4G19530 and AT4G34830 genes across shoot CPD and root CPD photoproducts.

From Figure 1.19, TCR effect is observed in the repair of AT4G19530 *and* AT4G34830 genes. Moreover, Figure 1.20 and 1.21 are the visual representation of read distribution across the corresponding genes.

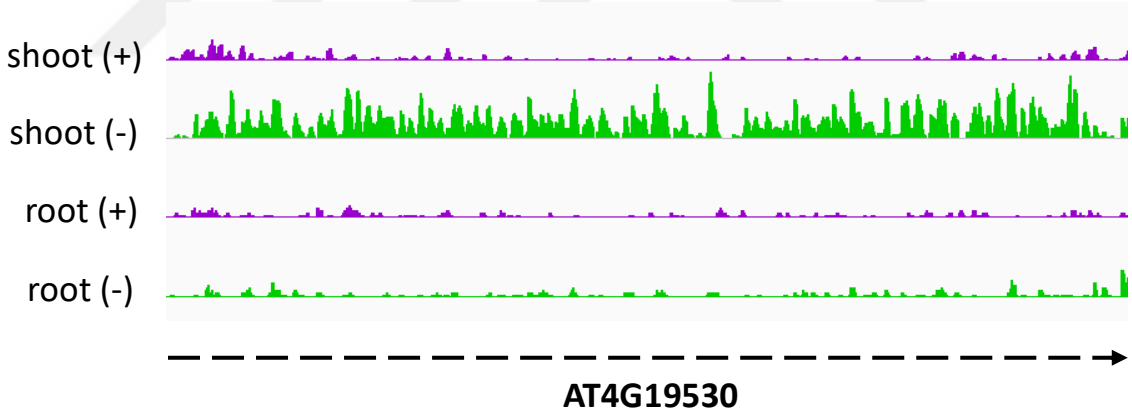

Figure 1.20: IGV screenshot of AT4G19530 XR-seq reads of shoot and root tissues across the genome.

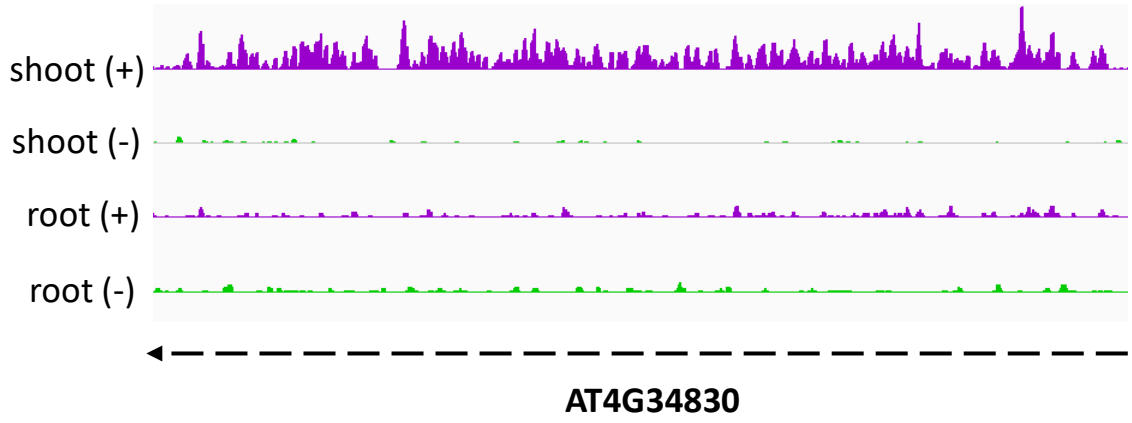

Figure 1.21: IGV screenshot of AT4G34830 XR-seq reads of shoot and root tissues across the genome.

In addition, a volcano plot from DESeq2 package was generated to see the distribution of statistically significant genes in terms of XR-seq reads. Figure 1.22 shows the positive fold change (significance: CPD shoot > CPD root) and negative fold change (significance: CPD root > CPD shoot) significant genes in the context of XR-seq.

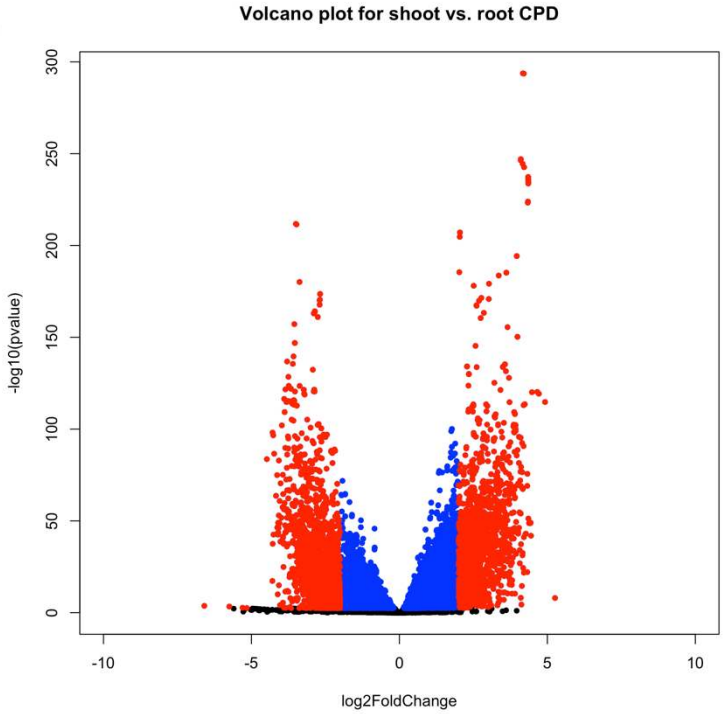

Figure 1.22: Volcano plot of shoot and root CPD XR-seq reads by DESeq2.

The dots representing the extreme log2foldchange and log(padj) values are the ones, which have been reported to be most significantly different genes affected by UV damage. Red dots correspond to genes whose log2foldchange is outside the interval of (-2, 2) while blue dots correspond to the interval of (-2, 2).

What is more, a heatmap of shoot and root CPD was created using significantly different top 5000 genes in terms of the accumulation of XR-seq reads and shown in Figure 1.23 below.

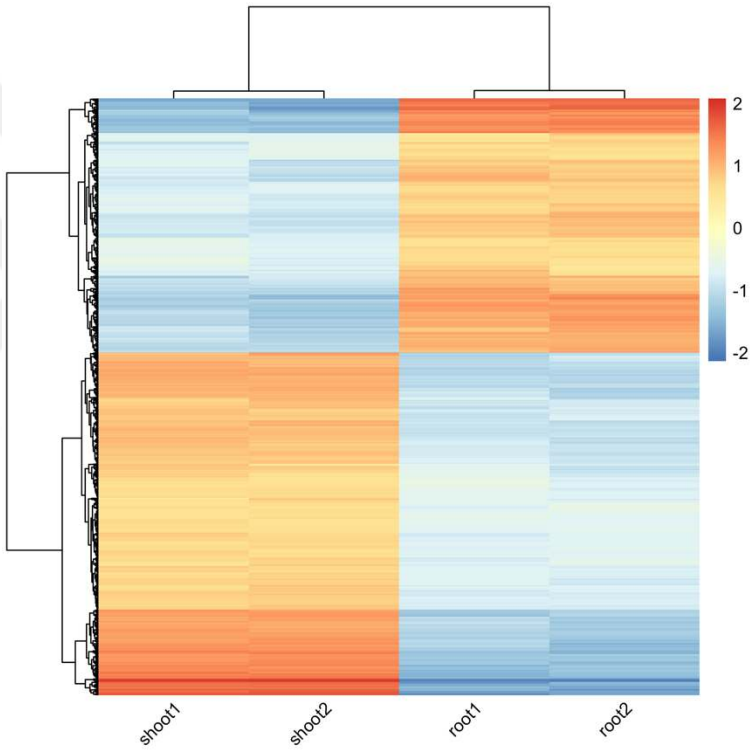

Figure 1.23: Heatmap of shoot and root CPD XR-seq reads by DESeq2.

The gradient palette emphasizes the relative abundance of XR-seq reads across different samples on columns for each gene in rows. Moreover, genes colored and clustered in red are the ones whose XR-seq reads are significantly enriched compared to the other samples while blue lines are the ones whose excision product read count is much less pronounced.

Read count values for each gene of tissue was subjected to DESeq2, and the XR-seq profiles of the following genes were found to be significantly different across tissues

compared. The repair of the following genes are the ones highly pronounced in shoot vs seedling samples.

XR-seq reads from gene AT2G14210, which encodes for a MADS box transcription factor acting as a repressor factor of seed germination event, and AT5G06839 were found to be significantly higher in seedling compared to shoot tissue (Jakoby et al., 2002; Lin et al., 2020).

Figure 1.24 exhibits TS/NTS ratio of significantly different excision repair of genes AT2G14210 and AT5G06839.

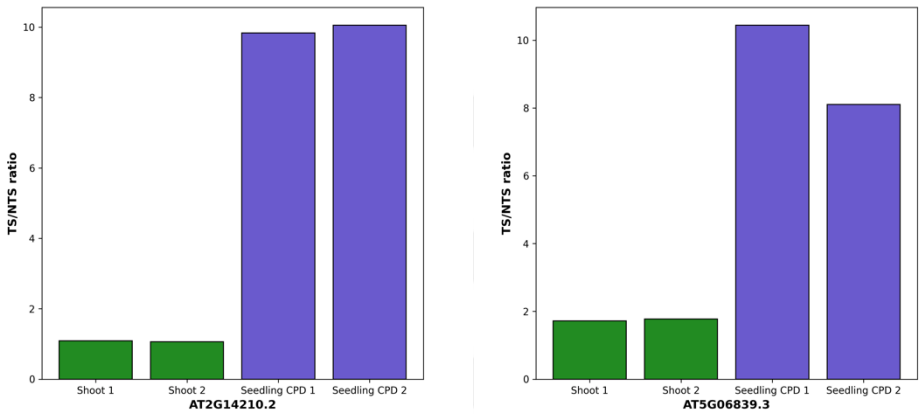

Figure 1.24: TS/NTS ratio of AT2G14210 and AT5G06839 genes across shoot CPD and seedling CPD photoproducts.

Figure 1.25 and 1.26 below are the visual representation of read distribution across the corresponding genes.

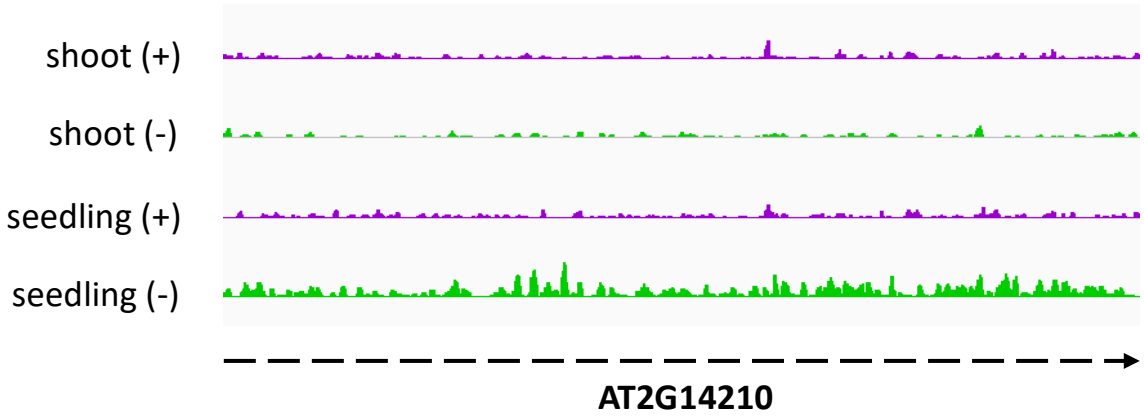

Figure 1.25: IGV screenshot of AT2G14210 XR-seq reads of shoot and seedling tissues across the genome.

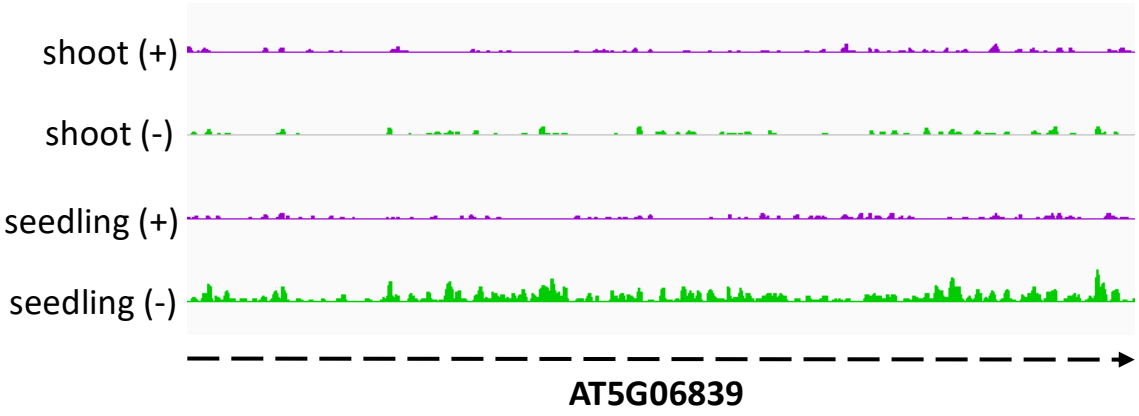

Figure 1.26: IGV screenshot of AT5G06839 XR-seq reads of shoot and seedling tissues across the genome.

Also, AT1G08320 is a gene encoding for a member of bZIP transcription factor family and AT2G47000 is a gene encoding for an auxin flux transporter protein controlling the cellular auxin concentration, whose XR-seq reads both were significantly enriched in seedling in comparison with shoot tissue (Jakoby et al., 2002; Kubeš et al., 2012).

Figure 1.27 shows TS/NTS ratio of significantly different excision repair of genes AT1G08320 and AT2G47000.

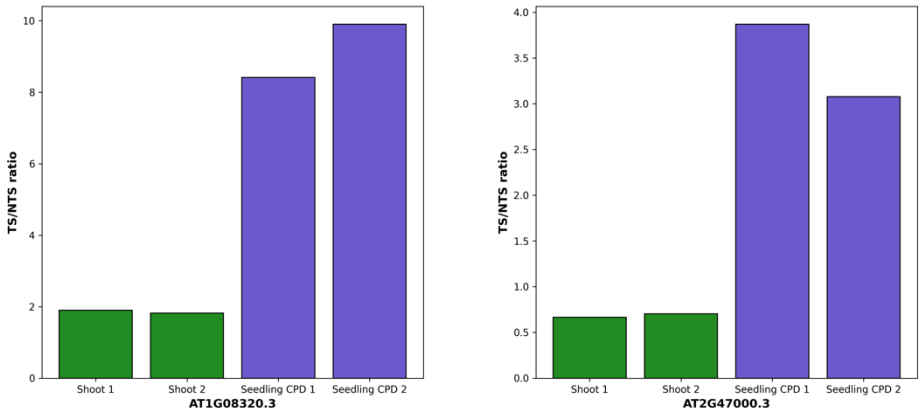

Figure 1.27: TS/NTS ratio of AT1G08320 and AT2G47000 genes across shoot CPD and seedling CPD photoproducts.

Furthermore, Figure 1.28 and 1.29 are the visual representation of read distribution across the corresponding genes.

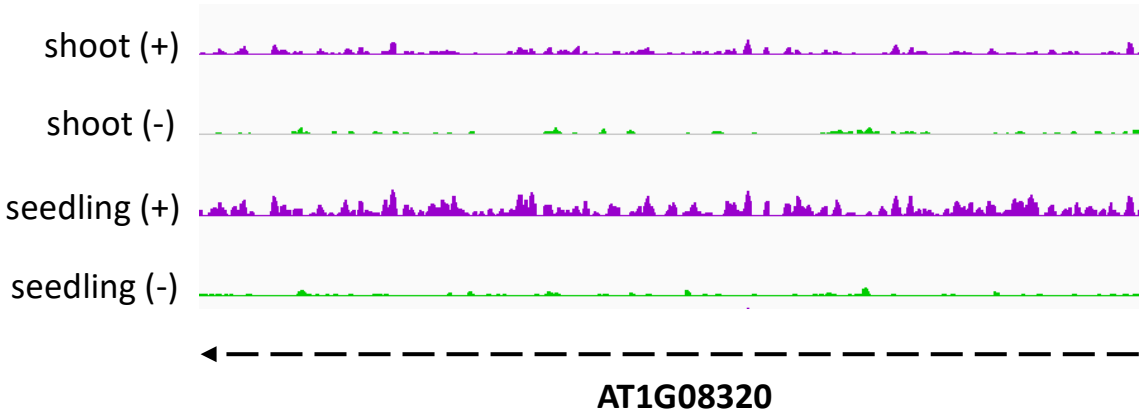

Figure 1.28: IGV screenshot of AT1G08320 XR-seq reads of shoot and seedling tissues across the genome.

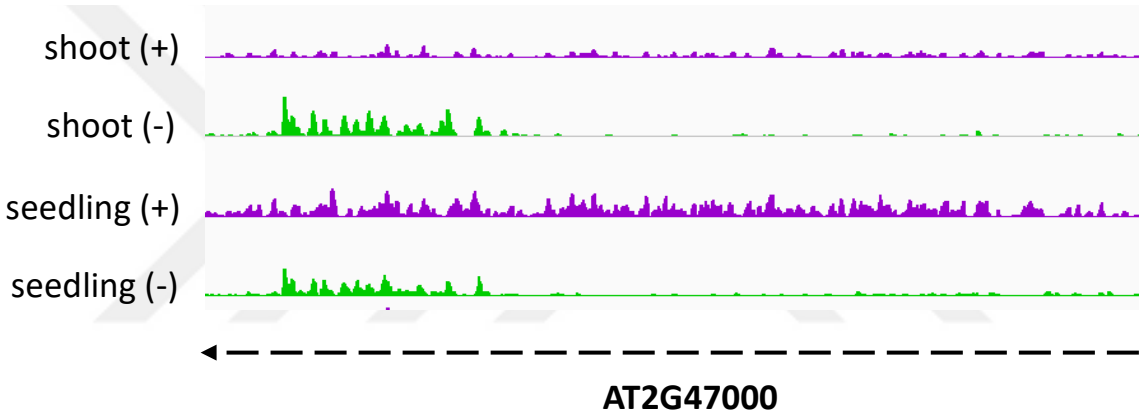

Figure 1.29: IGV screenshot of AT2G47000 XR-seq reads of shoot and seedling tissues across the genome.

Moreover, AT3G53480 is a gene encoding for an inhibitor responsible for negative regulation of auxin transport and AT1G30650 is a gene encoding for a member of WRKY transcription factor family, whose XR-seq reads both were significantly enriched in seedling in comparison with shoot tissue (Jang et al., 2010; Ru˚zic˘ka et al., 2010).

Figure 1.30 demonstrates TS/NTS ratio of significantly different excision repair of genes AT3G53480 and AT1G30650.

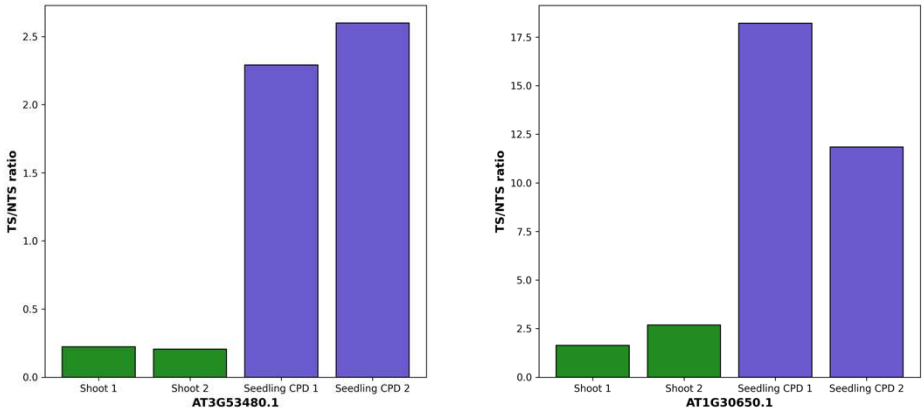

Figure 1.30: TS/NTS ratio of AT3G53480 and AT1G30650 genes across shoot CPD and seedling CPD photoproducts.

Figure 1.31 and 1.32 are the visual representations of read distribution across the corresponding genes.

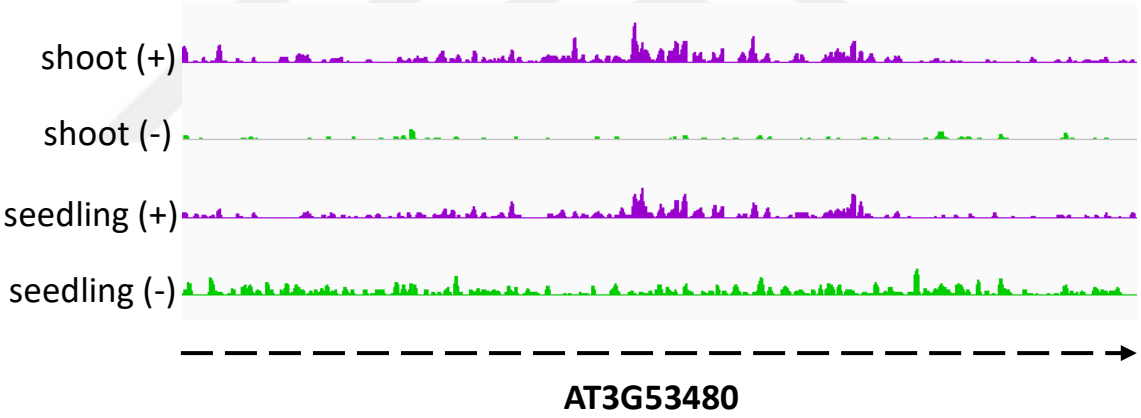

Figure 1.31: IGV screenshot of AT3G53480 XR-seq reads of shoot and seedling tissues across the genome.

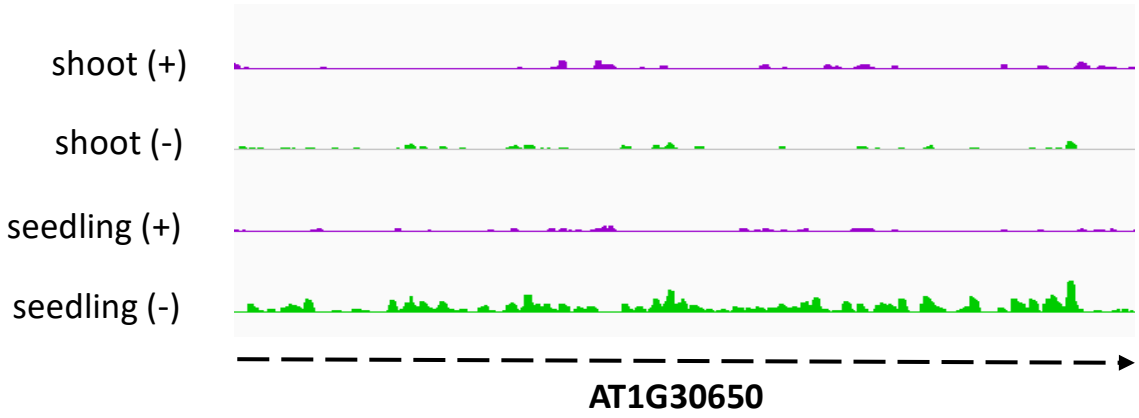

Figure 1.32: IGV screenshot of AT1G30650 XR-seq reads of shoot and seedling tissues across the genome.

Furthermore, AT4G05200 is a gene encoding for a cysteine-rich receptor like kinase protein and the protein product of AT4G35380 is responsible for the regulation of post-golgi trafficking (Suo et al., 2021; Wrzaczek et al., 2010).

Figure 1.33 exhibits TS/NTS ratio of significantly different excision repair of genes AT4G05200 and AT4G35380.

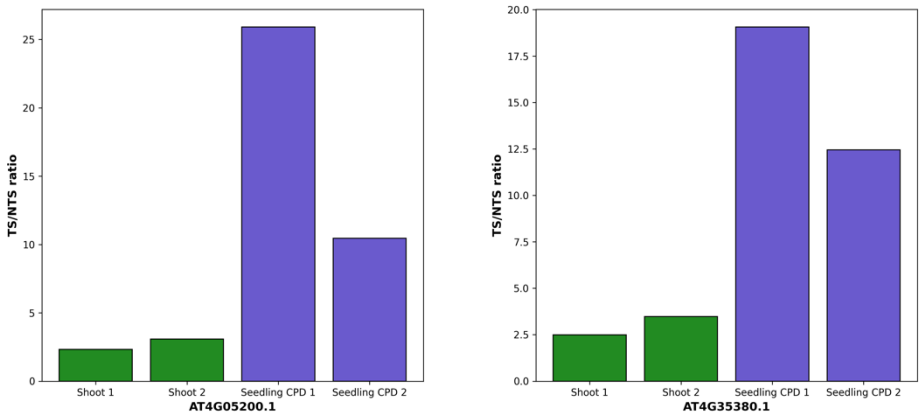

Figure 1.33: TS/NTS ratio of AT4G05200 and AT4G35380 genes across shoot CPD and seedling CPD photoproducts.

Moreover, Figure 1.34 and 1.35 help visualize the read distributions across the corresponding genes.

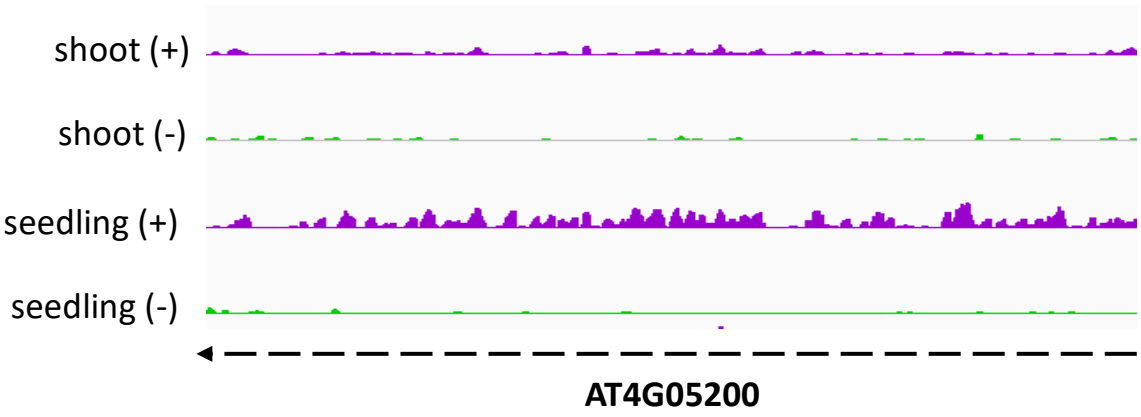

Figure 1.34: IGV screenshot of AT4G05200 XR-seq reads of shoot and seedling tissues across the genome.

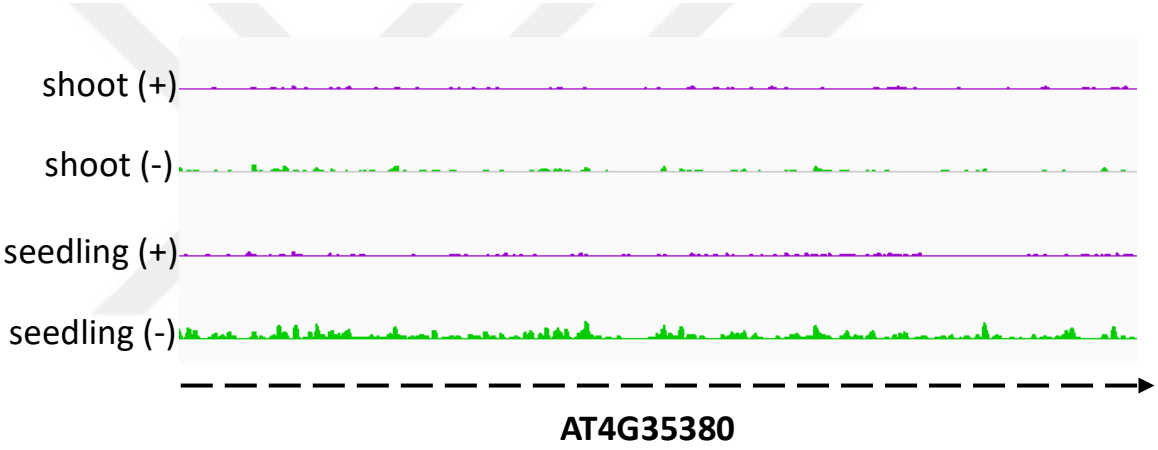

Figure 1.35: IGV screenshot of AT4G35380 XR-seq reads of shoot and seedling tissues across the genome.

In addition, a volcano plot from DESeq2 package was generated to see the distribution of statistically significant genes in terms of XR-seq reads. Figure 1.36 shows the positive fold change (significance: CPD shoot > CPD seedling) and negative fold change (significance: CPD seedling > CPD shoot) significant genes in the context of XR-seq.

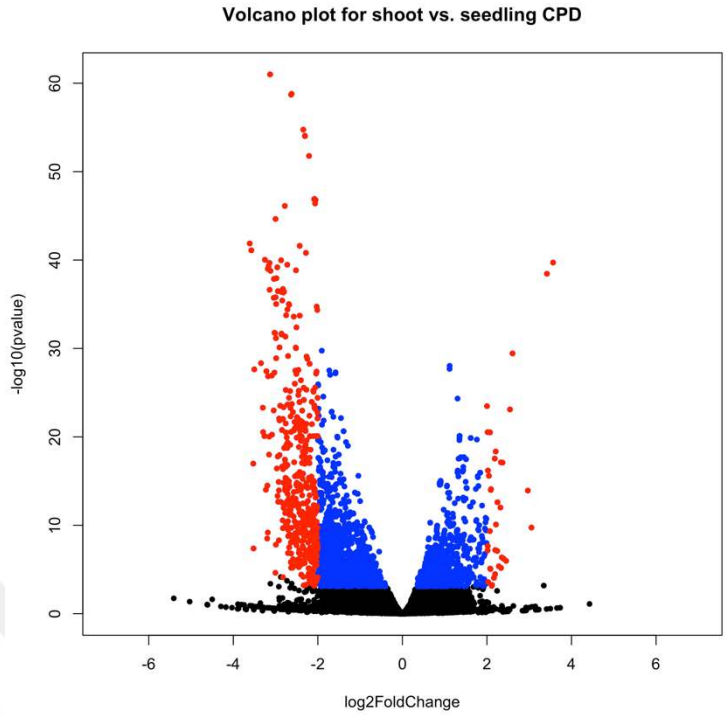

Figure 1.36: Volcano plot of shoot and seedling CPD XR-seq reads by DESeq2.

As previously mentioned, the dots in the plot accounts for the extreme log2foldchange and log(padj) values for each gene where all padj values smaller than 0.01 were selected for the sake of a better visual representation and understanding in this thesis. Red dots correspond to genes whose log2foldchange is outside the interval of (-2, 2) while blue dots correspond to the interval of (-2, 2) as the previous one

Another heatmap for was plotted for this comparison and top 5000 genes whose XR-seq profiles between different tissues were found to be significantly different in Figure 1.37. Again, for all heatmap plots in this thesis were generated by taking top 5000 genes into account but not the less significant ones.

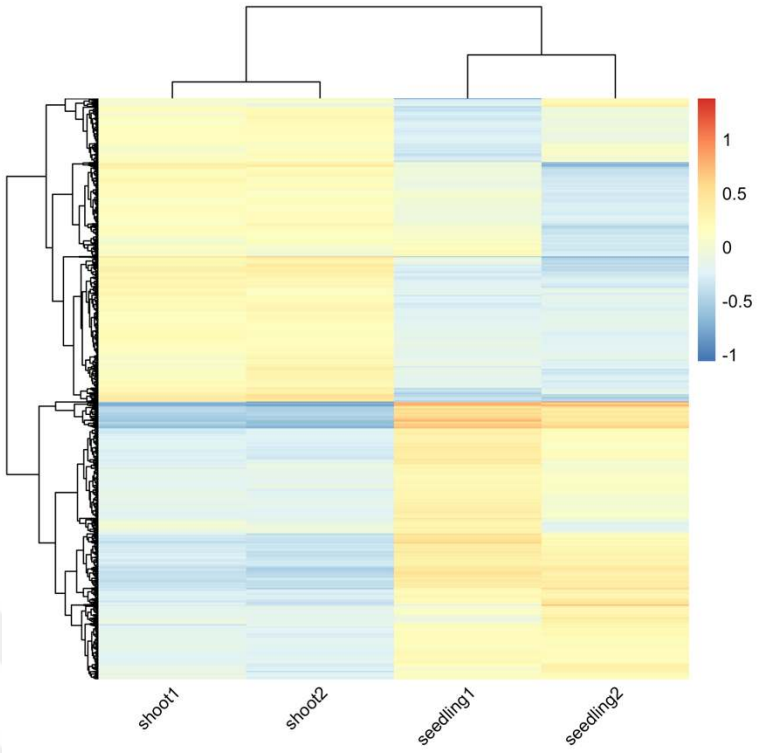

Figure 1.37: Heatmap of shoot and seedling CPD XR-seq reads by DESeq2.

Moreover, read count values for each gene of tissue was subjected to DESeq2, and the XR-seq profiles of the following genes were found to be significantly different across tissues compared. The following genes are the ones highly pronounced in root vs seedling samples.

AT5G04140 is a gene encoding for ferredoxin dependent glutamate synthase-like protein thought to be involved in photorespiration and nitrogen assimilation and AT1G58602 is a gene encoding for a protein specialized in the recognition of peronospora parasitica (L. Li et al., 2020). The XR-seq reads of corresponding genes were highly enriched in seedling tissue compared to root.

Figure 1.38 shows TS/NTS ratio of significantly different excision repair of genes AT5G04140 and AT1G58602 across root and seedling CPD tissue.

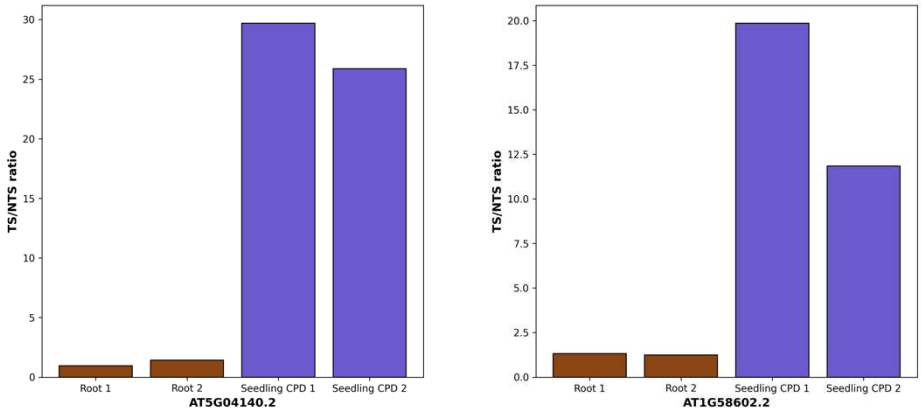

Figure 1.38: TS/NTS ratio of AT5G04140 and AT1G58602 genes across root CPD and seedling CPD photoproducts.

Figure 1.39 and 1.40 below are the visual representation of read distribution across the corresponding genes.

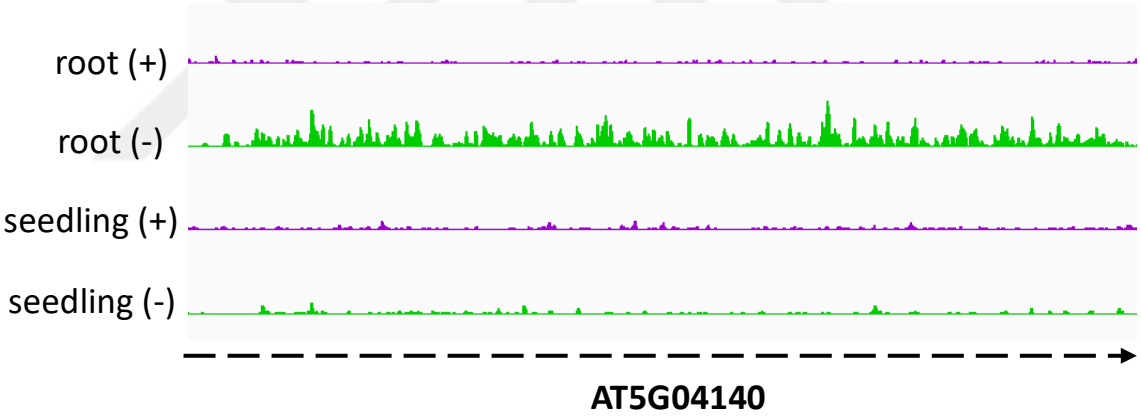

Figure 1.39: IGV screenshot of AT5G04140 XR-seq reads of root and seedling tissues across the genome.

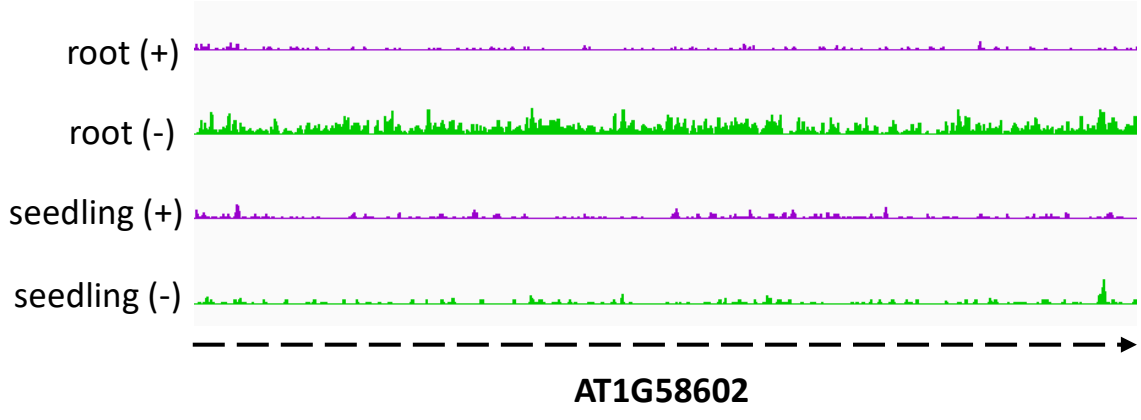

Figure 1.40: IGV screenshot of AT1G58602 XR-seq reads of root and seedling tissues across the genome.

Furthermore, AT4G16990 is reported to encode for a disease resistance protein while AT4G34830, as previously mentioned, is involved in stabilization of rbcL mRNA (Johnson et al., 2010). The XR-seq reads of corresponding genes were highly enriched in seedling tissue compared to root.

TS/NTS ratio of significantly different excision repair of genes AT4G16990 and AT4G34830 across root and seedling CPD tissue is demonstrated in Figure 1.41.

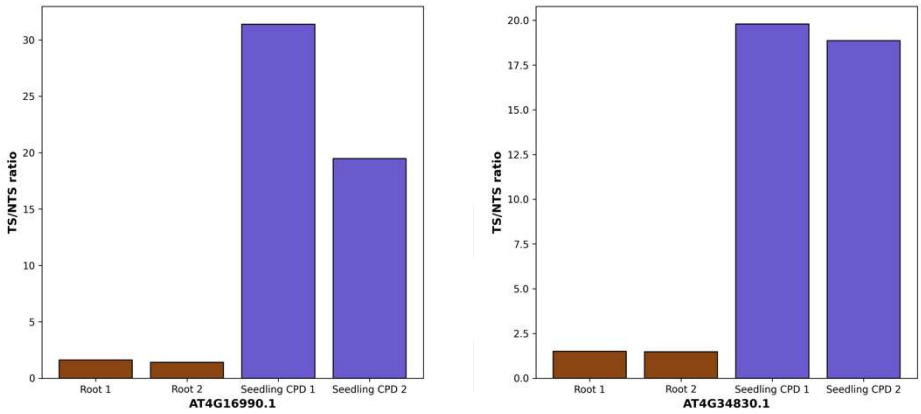

Figure 1.41: TS/NTS ratio of AT4G16990 and AT4G34830 genes across root CPD and seedling CPD photoproducts.

Nucleotide excision coupled to transcription probably occurs in the case of these genes based on the observation of higher TS/NTS ratio in seedling CPD samples.

In addition, Figure 1.42 and 1.43 are the visual representation of read distribution across the corresponding genes.

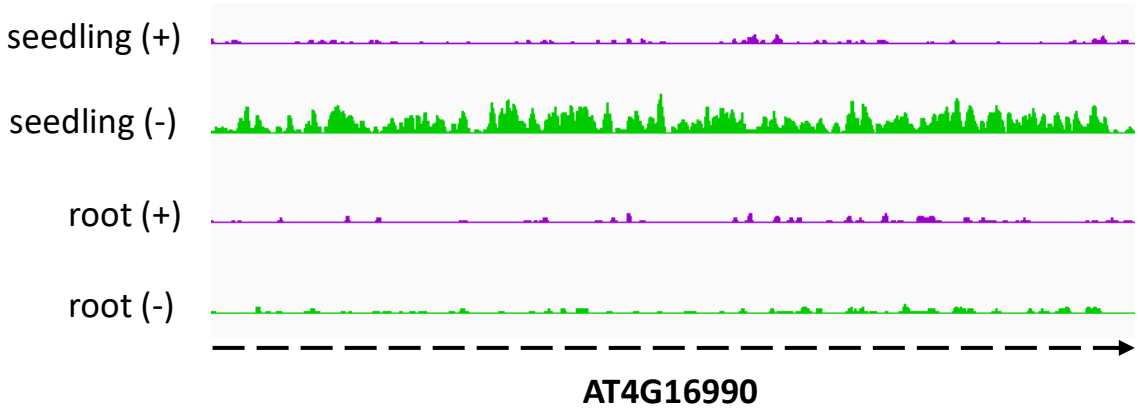

Figure 1.42: IGV screenshot of AT4G16990 XR-seq reads of root and seedling tissues across the genome.

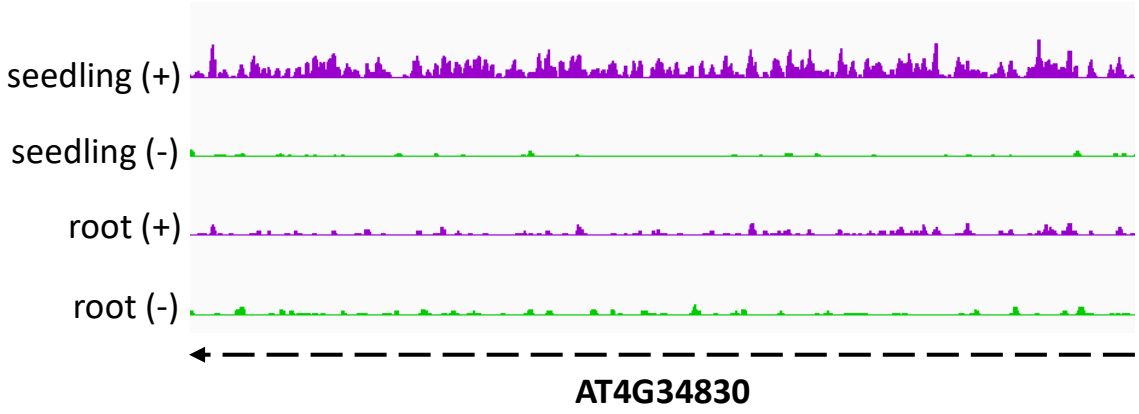

Figure 1.43: IGV screenshot of AT4G34830 XR-seq reads of root and seedling tissues across the genome.

As previously mentioned, AT4G10060 is a gene encoding for glucosylceramidase enzyme responsible for the hydrolysis of long acyl chain glucosylceramides (Dai et al., 2020). In addition, AT3G25690 is responsible for the production of acting binding protein, whose task is to position chloroplast in plant cell (Kitashova et al., 2021). The XR-seq reads of corresponding genes were highly enriched in seedling tissue compared to root.

TS/NTS ratio of significantly different excision repair of genes AT4G10060 and AT3G25690 across root and seedling CPD tissue is demonstrated in Figure 1.44.

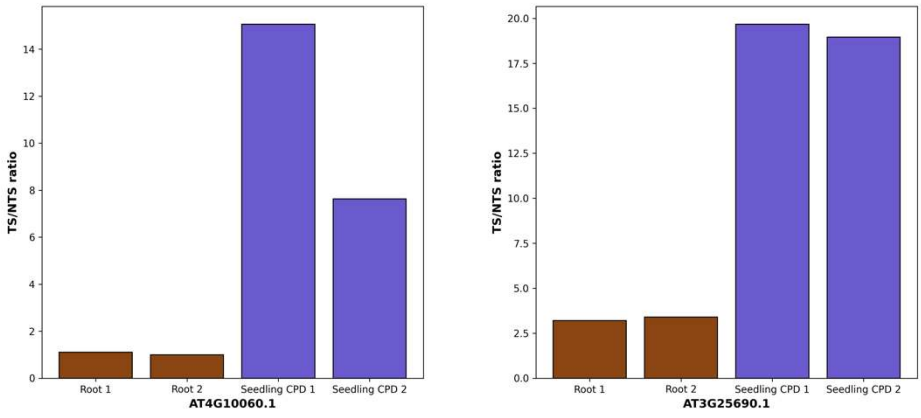

Figure 1.44: TS/NTS ratio of AT4G10060 and AT3G25690 genes across root CPD and seedling CPD photoproducts.

Thus, Figure 1.45 and 1.46 below are the visual representation of read distribution across the corresponding genes.

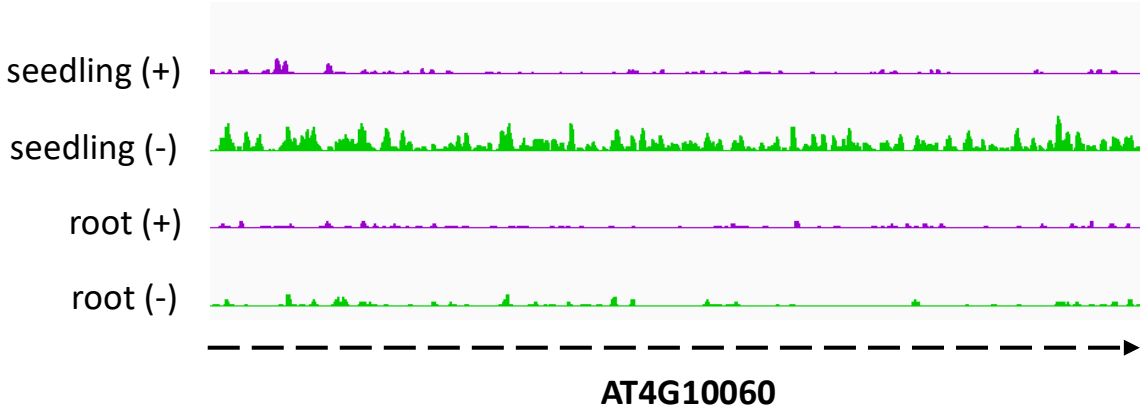

Figure 1.45: IGV screenshot of AT4G10060 XR-seq reads of root and seedling tissues across the genome.

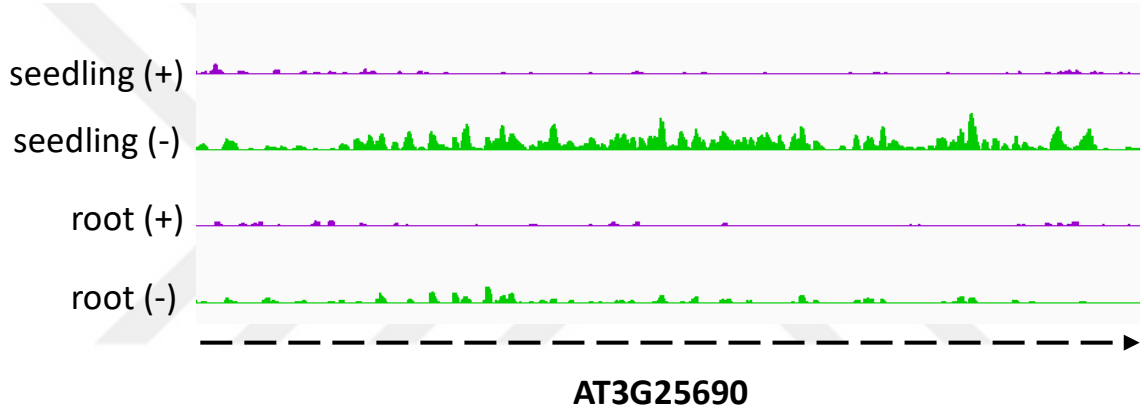

Figure 1.46: IGV screenshot of AT3G25690 XR-seq reads of root and seedling tissues across the genome.

In addition, AT1G66970 is a gene encoding for a member of the glycerophosphodiester phosphodiesterase like (GDPD-like) family. Furthermore, as previously mentioned, AT4G28080 (also known as REC2) works with REC1 and REC3 partners to tune the size of chloroplast compartment (Cheng et al., 2011; Larkin et al., 2016). The XR-seq reads of corresponding genes were highly enriched in seedling tissue compared to root.

TS/NTS ratio of significantly different excision repair of genes AT1G66970 and AT4G28080 across root and seedling CPD tissue is exhibited in Figure 1.47.

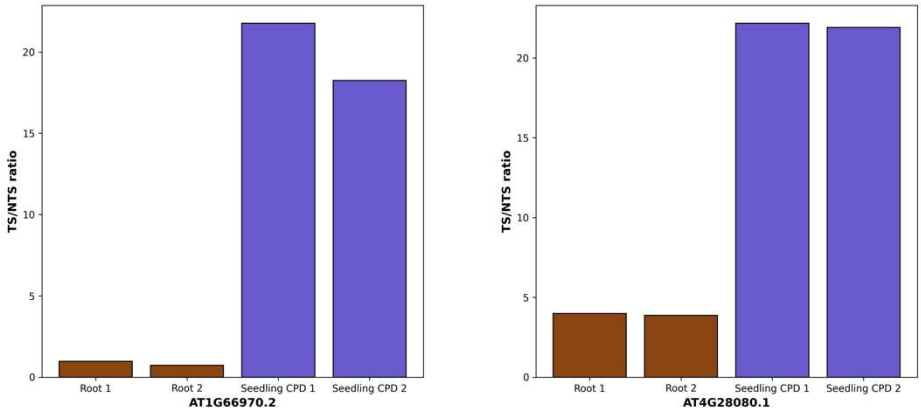

Figure 1.47: TS/NTS ratio of AT1G66970 and AT4G28080 genes across root CPD and seedling CPD photoproducts.

AT1G66970 and AT4G28080 genes might be suggested that they were being actively expressed and the indirect effect of transcription was measured collecting significant amount of excision products from the seedling samples in comparison to root samples, resulting in higher TS/NTS ratio as an indicator of TCR.

Thus, Figure 1.48 and 1.49 are the visual representation of read distribution across the corresponding genes.

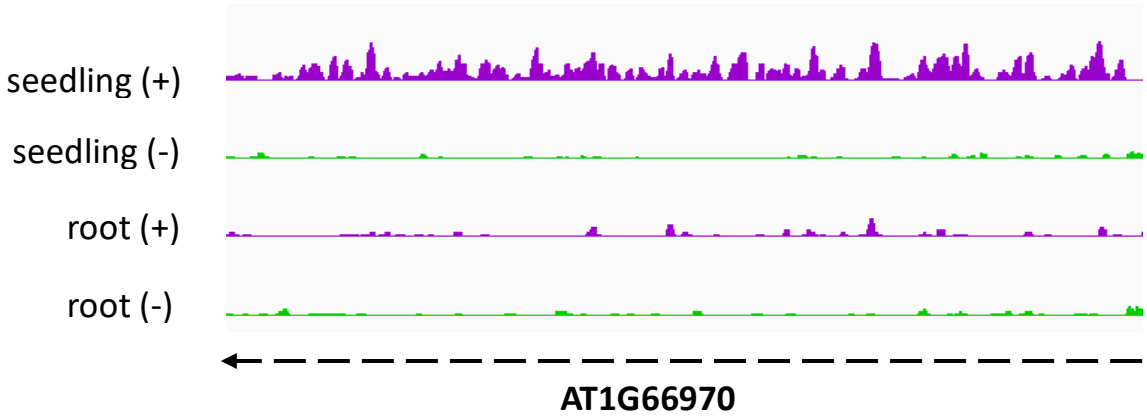

Figure 1.48: IGV screenshot of AT1G66970 XR-seq reads of root and seedling tissues across the genome.

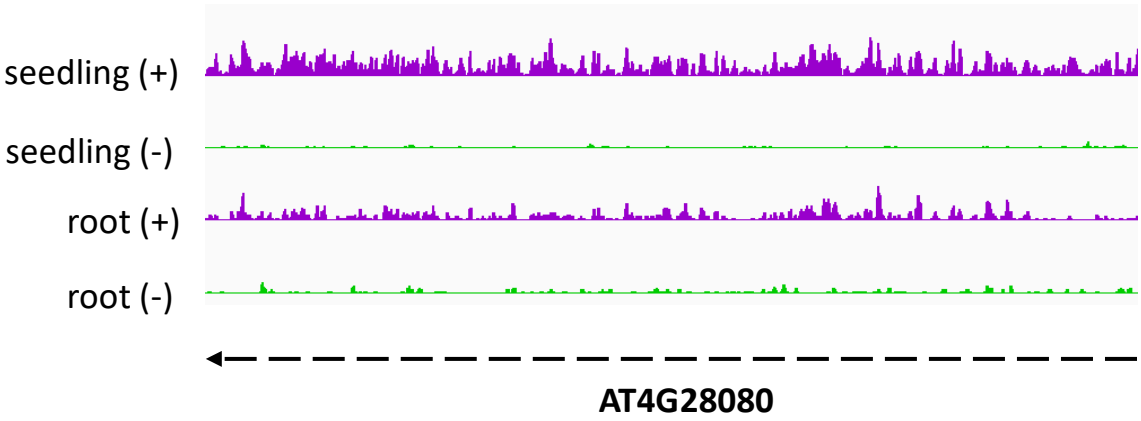

Figure 1.49: IGV screenshot of AT4G28080 XR-seq reads of root and seedling tissues across the genome.

Moreover, a volcano plot from DESeq2 package was generated to see the distribution of statistically significant genes in terms of XR-seq reads. Figure 1.50 shows the positive fold change (significance: CPD root > CPD seedling) and negative fold change (significance: CPD seedling > CPD root) significant genes in the context of XR-seq.

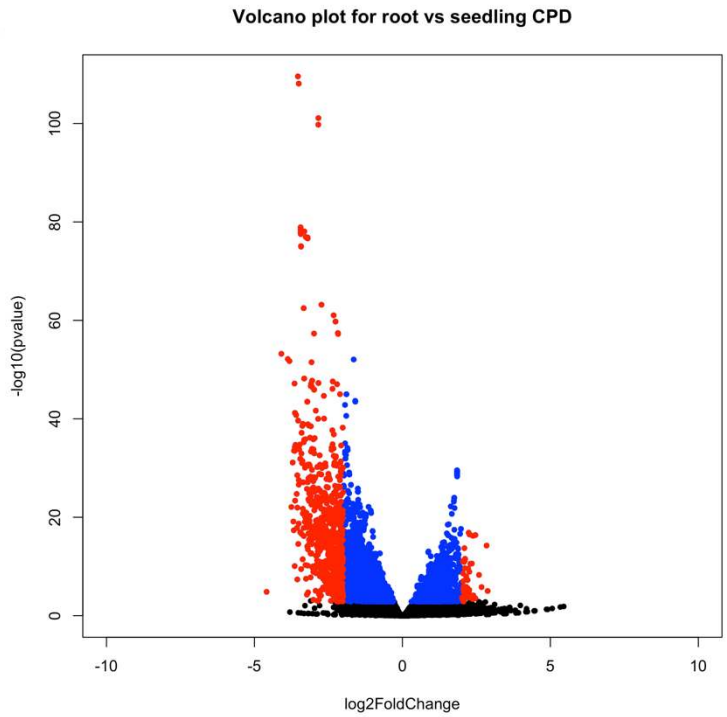

Figure 1.50: Volcano plot of root and seedling CPD XR-seq reads by DESeq2.

The powerful behavior of volcano plots is the ability of observing the significant changes on the plot. As the y axis values increase, the significance for that particular gene becomes more important compared to less y values. For instance, the most statistically significant gene is at the very top, given that the color is red, this gene in CPD seedling data is significantly different in containing XR-seq reads compared to CPD root condition as negative log2foldchange corresponds to the genes whose dominance is taken over by CPD seedling data.

Furthermore, a heatmap for root vs. seedling was plotted using pheatmap. The scaling on the upper right side of the plot correspond to z scores which are derived and calculated from read count values for each gene across all samples.

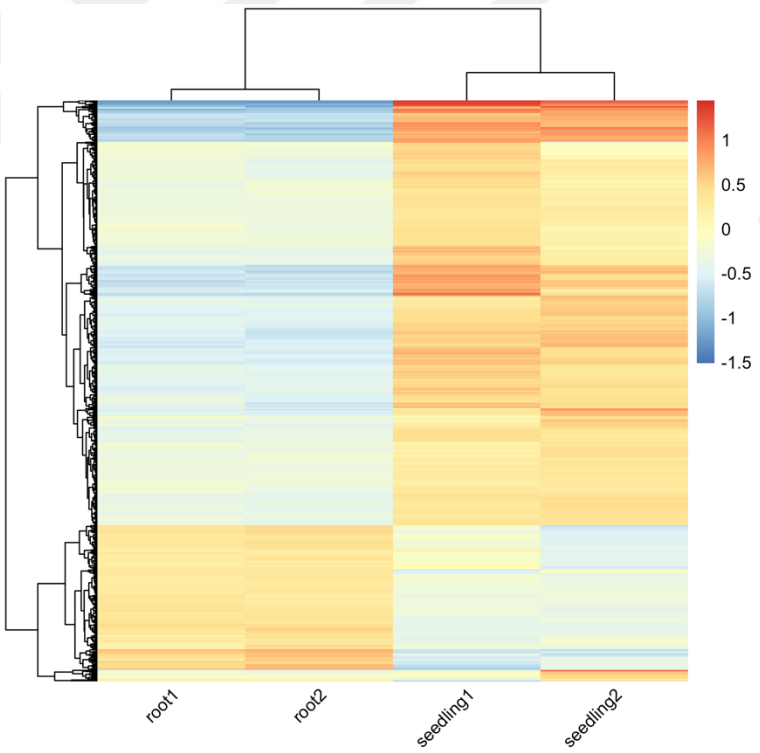

Figure 1.51: Heatmap of root and seedling CPD XR-seq reads by DESeq2.

As a final comparison of two Arabidopsis damage photoproduct, read count values for each gene of tissue was subjected to DESeq2, and the XR-seq profiles of the following genes were found to be significantly different across tissues. The following genes are the ones highly pronounced in 6-4 PP vs CPD seedlings samples.

AT5G20490 is a gene encoding for a member of myosin family protein and involved in root hair growth. Moreover, AT4G13370 is responsible for the production of a protein called cortical microtubule disordering1 (CORD1), which is essential for secondary cell wall patterning . The XR-seq reads of corresponding genes were highly enriched in CPD seedling tissue compared to 6-4 PP seedling (Park & Nebenführ, 2013; Sasaki et al., 2017).

TS/NTS ratio of significantly different excision repair of genes AT5G20490 and AT4G13370 across seedling CPD and seedling (6-4) PP photoproducts is shown in Figure 1.52.

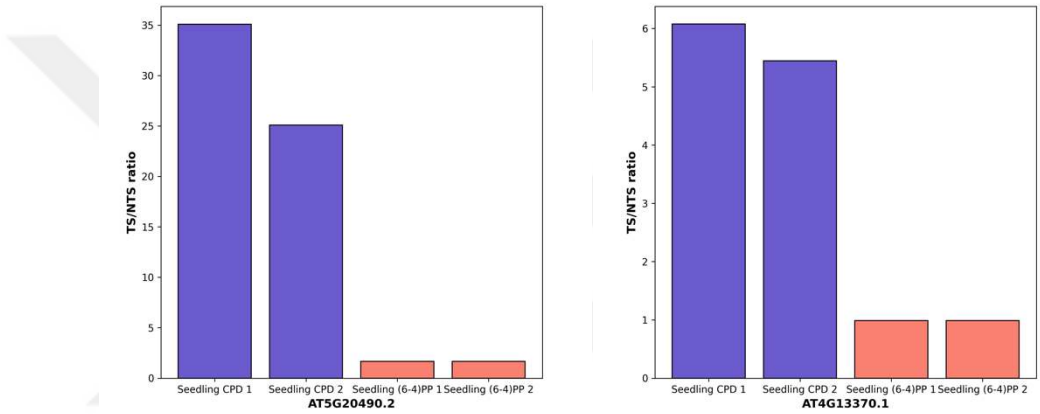

Figure 1.52: TS/NTS ratio of AT5G20490 and AT4G13370 genes across root CPD and seedling CPD photoproducts.

Although CPD repair was previously shown to be dominant over (6-4) PP in terms of transcription coupling, Figure 1.52 is also an additional evidence of how TCR of CPD photoproducts is superior to (6-4) PP.

Thus, Figure 1.53 and 1.54 are the visual representation of read distribution across the corresponding genes.

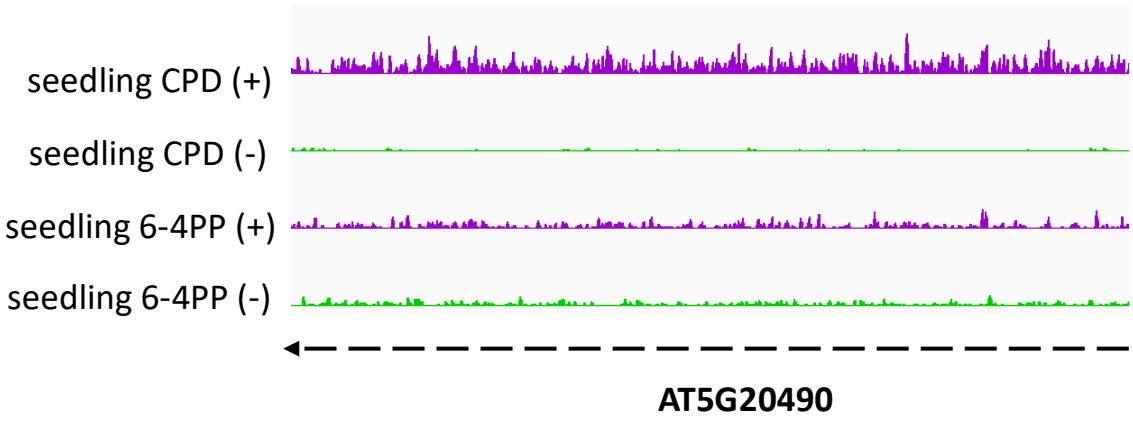

Figure 1.53: IGV screenshot of AT5G20490 XR-seq reads of CPD and 6-4 PP seedling tissues across the genome.

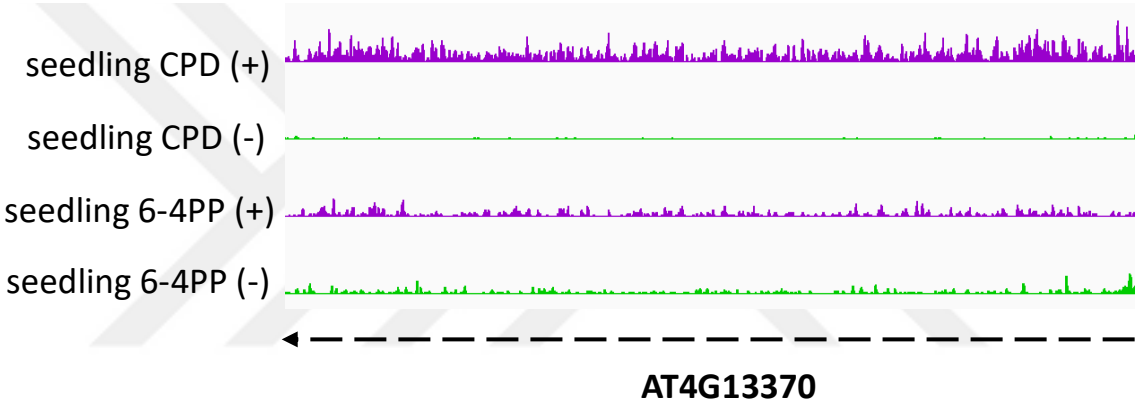

Figure 1.54: IGV screenshot of AT4G13370 XR-seq reads of CPD and 6-4 PP seedling tissues across the genome.

Moreover, AT1G17440 is a gene reported to operate in cytokinin signaling as a negative regulator. Moreover, AT3G43300 is a gene, which was reported to be associated with immunity and known to produce a member of adenosine diphosphate ribosylation factor guanine nucleotide exchange factor AFR-GEF protein (Kubo et al., 2011; Zhao et al., 2020). The XR-seq reads of corresponding genes were highly enriched in CPD seedling tissue compared to 6-4 PP seedling.

TS/NTS ratio of significantly different excision repair of genes AT1G17440 and AT3G43300 across seedling CPD and seedling (6-4) PP photoproducts is demonstrated in Figure 1.55.

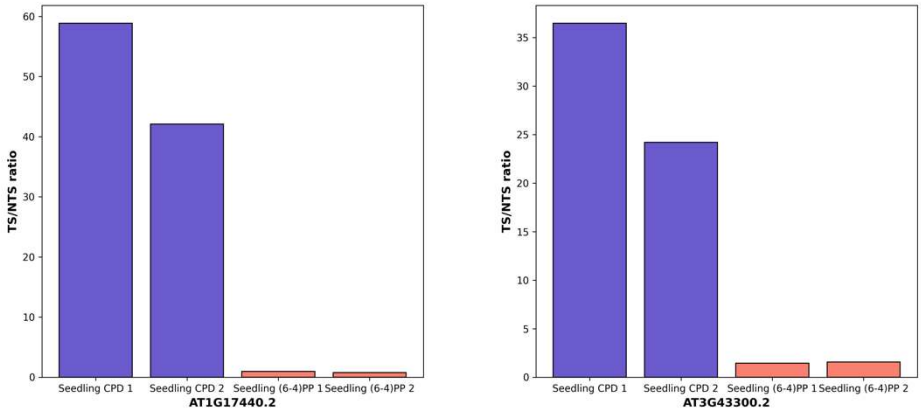

Figure 1.55: TS/NTS ratio of AT1G17440 and AT3G43300 genes across root CPD and seedling CPD photoproducts.

Thus, Figure 1.56 and 1.57 are the visual representation of read distribution across the corresponding genes.

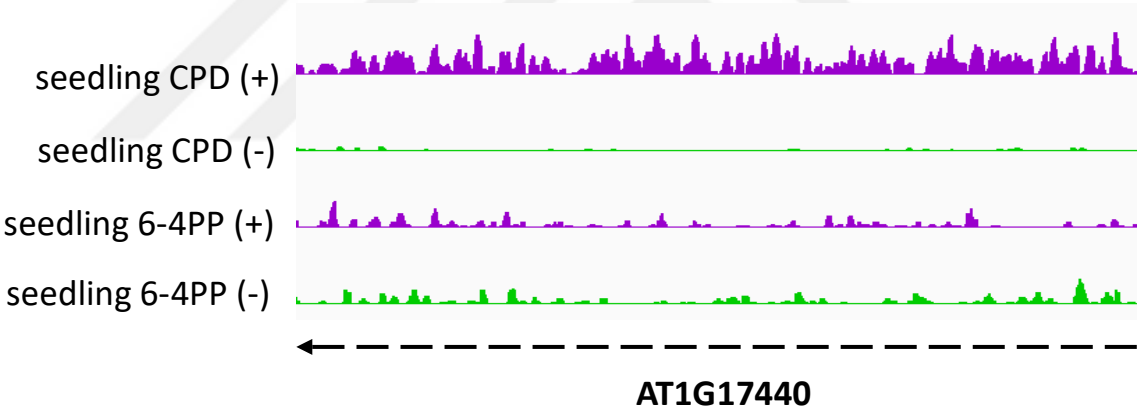

Figure 1.56: IGV screenshot of AT1G17440 XR-seq reads of CPD and 6-4 PP seedling tissues across the genome.

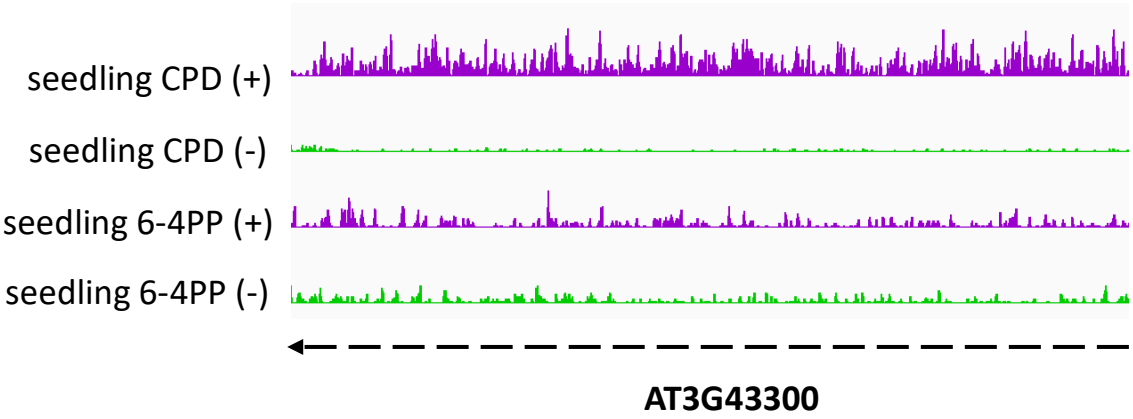

Figure 1.57: IGV screenshot of AT3G43300 XR-seq reads of CPD and 6-4 PP seedling tissues across the genome.

Furthermore, AT3G44340 is a gene encoding for a protein homologous to yeast Sec24. Furthermore, AT5G35980 is a protein kinase encoding gene, whose some of the substrates include annexin family proteins. The XR-seq reads of corresponding genes were highly enriched in CPD seedling tissue compared to 6-4 PP seedling.

TS/NTS ratio of significantly different excision repair of genes AT3G44340 and AT5G35980 across seedling CPD and seedling (6-4) PP photoproducts is demonstrated in Figure 1.58.

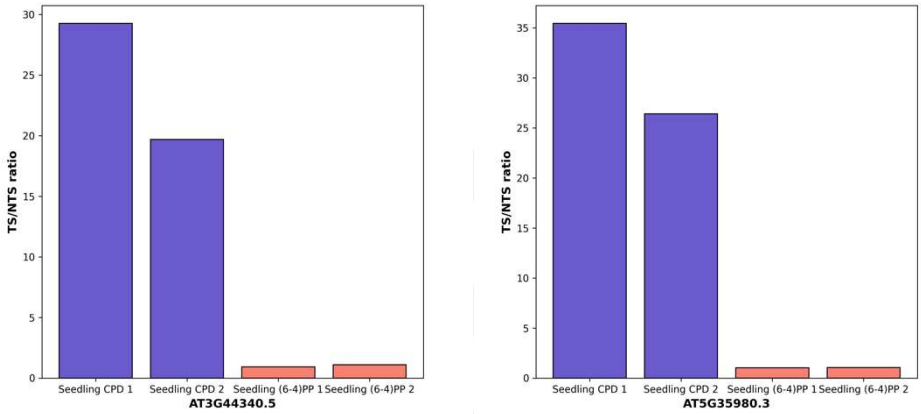

Figure 1.58: TS/NTS ratio of AT3G44340 and AT5G35980 genes across root CPD and seedling CPD photoproducts.

Thus, Figure 1.59 and 1.60 are the visual representations of read distribution across the corresponding genes.

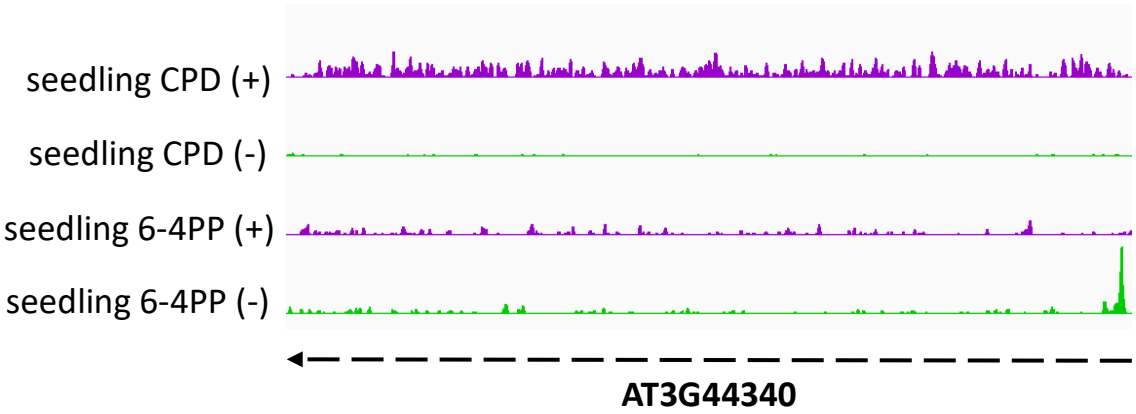

Figure 1.59: IGV screenshot of AT3G44340 XR-seq reads of CPD and 6-4 PP seedling tissues across the genome.

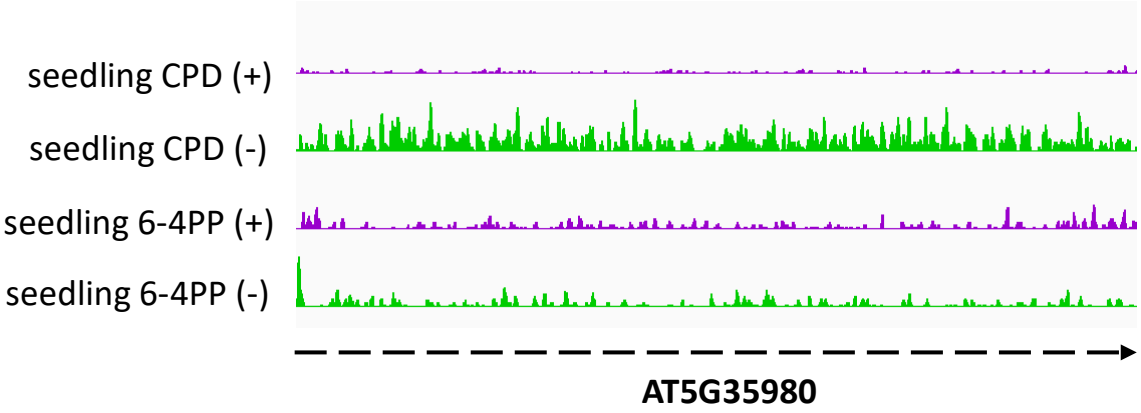

Figure 1.60: IGV screenshot of AT5G35980 XR-seq reads of CPD and 6-4 PP seedling tissues across the genome.

Finally, AT4G31900 is a gene, which encodes for chromatin remodeler 7 (chr7) in Arabidopsis. Moreover, AT1G72270 is a gene called harbinger transposon-derived protein 1 responsible for regulating active DNA demethylation; however, the paper explaining AT1G72270 has been retracted in the journal (Huang et al., 2017). Interestingly, the XR-seq reads of corresponding genes were highly enriched in 6-4 PP seedling tissue compared to CPD seedling unlike other previous comparisons.

TS/NTS ratio of significantly different excision repair of genes AT4G31900 and AT1G72270 across seedling CPD and seedling (6-4) PP photoproducts is demonstrated in Figure 1.61.

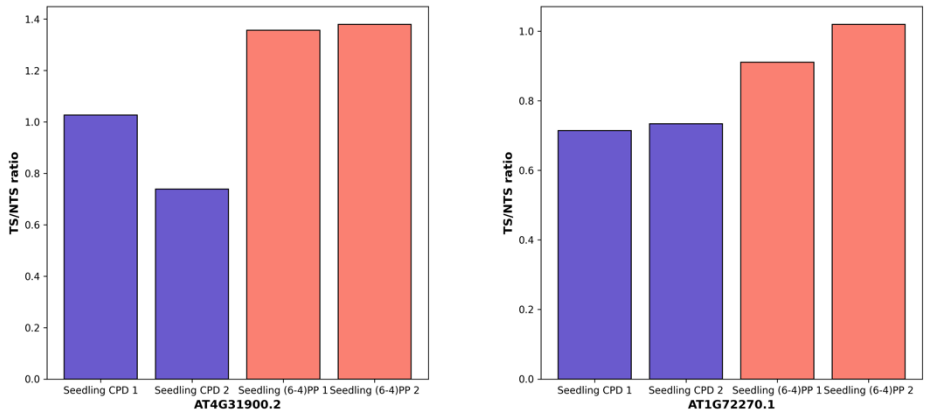

Figure 1.61: TS/NTS ratio of AT4G31900 and AT1G72270 genes across root CPD and seedling CPD photoproducts.

Although DESeq2 has found significance over tissues sample sets, it is hard to support this argument. What actually DESeq2 does in our analysis is to compare the read counts from both strands of a gene; therefore, not only TCR but also GGR are effective in determining the fate of DESeq2 results. These genes might be asserted that they are not actively transcribed, which explains the low effect of TCR such that the repair signal is indistinctive of whether it is obtained by TCR or GGR, but probably the effect of GGR was better pronounced in the DESEq2 analysis of these genes, given previous analyses supporting the superiority of TCR in CPD damage.

Thus, Figure 1.62 and 1.63 below are the visual representation of read distribution across the corresponding genes.

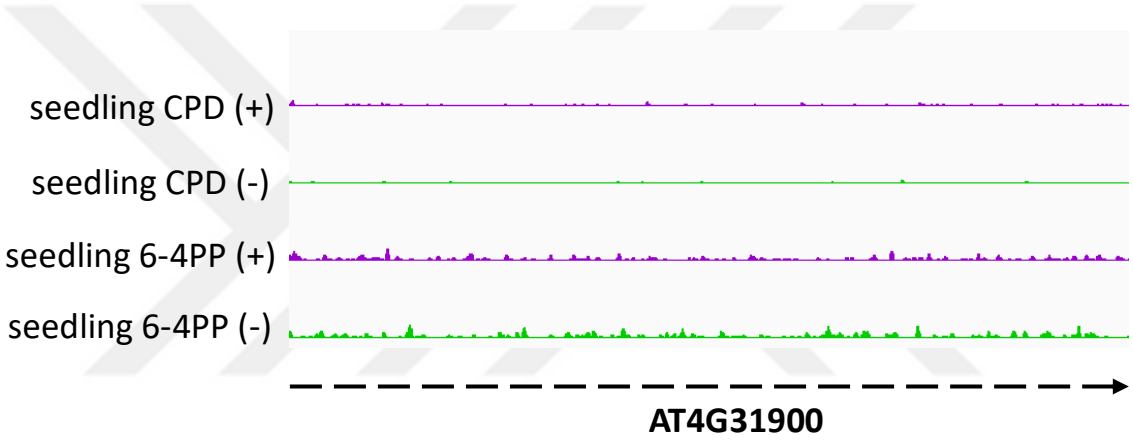

Figure 1.62: IGV screenshot of AT4G31900 XR-seq reads of CPD and 6-4 PP seedling tissues across the genome.

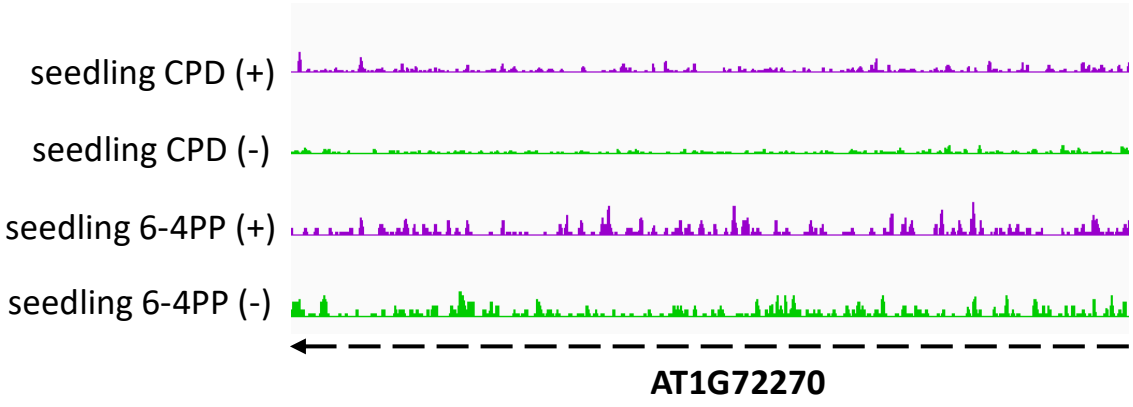

Figure 1.63: IGV screenshot of AT1G72270 XR-seq reads of CPD and 6-4 PP seedling tissues across the genome.

Moreover, a volcano plot from DESeq2 package was generated to see the distribution of statistically significant genes in terms of XR-seq reads. Figure 1.64 below shows the positive fold change (significance: CPD seedling > 6-4 PP seedling) and negative fold change (significance: 6-4 PP seedling > CPD seedling) significant genes in the context of XR-seq.

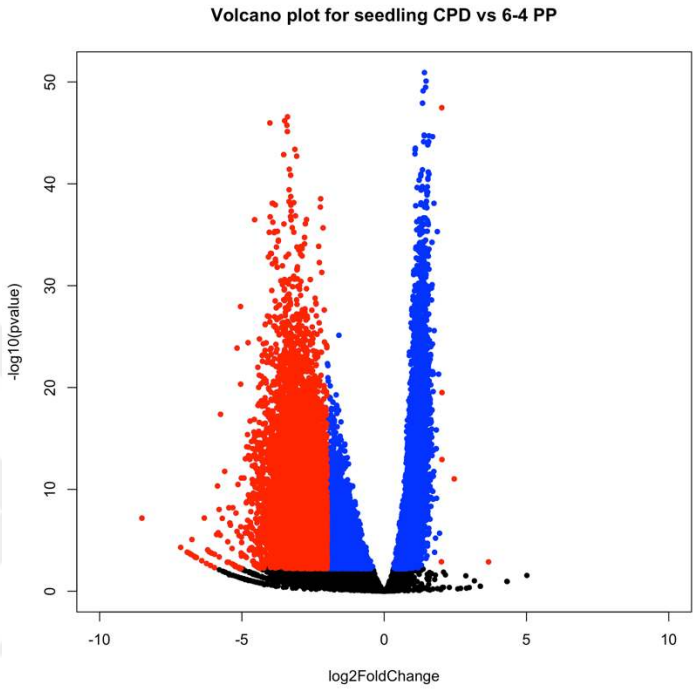

Figure 1.64: Volcano plot of seedling CPD and 6-4 PP XR-seq reads by DESeq2.

From Figure 1.64 above, it was inferred that 6-4 PP seedling data experienced a more dramatic log2fold change compared to CPD seedling data; in other words, more genes were abundant in 6-4 PP seedling sample in terms of XR-seq reads because of the overall dominance of red dots in the 6-4 PP seedling data zone on the left side. Moreover, the reason why the plot is called volcano is because smaller log2foldchange values are not included and the biological data is well correlated with log2foldchange and minus logpadj values in an exponential manner, which together make the characteristic shape of the volcano plot.

Lastly, the heatmap for 6-4 PP vs. CPD seedling was generated to have a general understanding of tissue specific behavior of DNA excision repair in two different photoproducts.

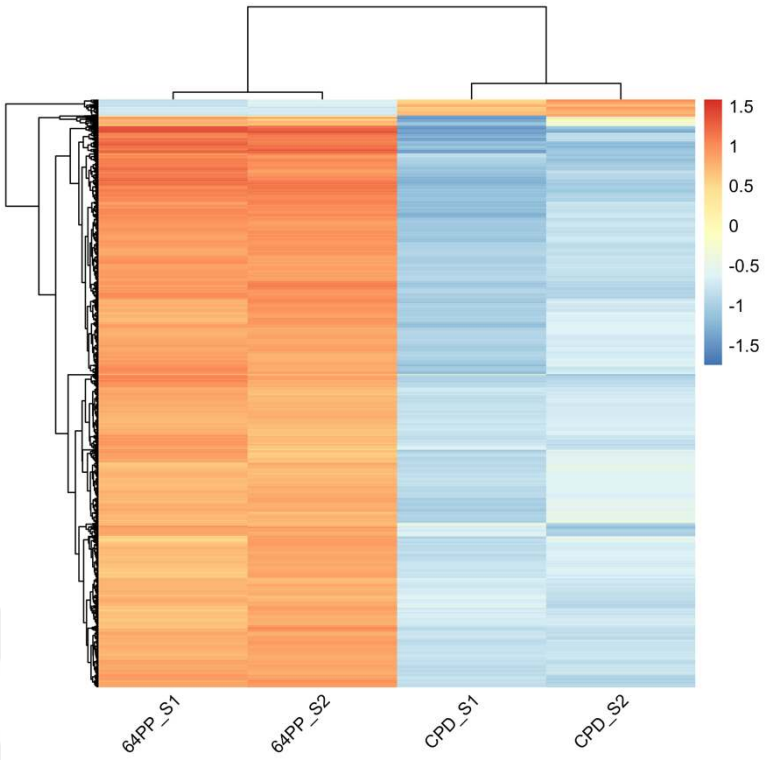

Figure 1.65: Heatmap of seedling CPD and 6-4 PP XR-seq reads by DESeq2.

6-4 PP XR-seq data from Figure 1.49 is especially intense compared to CPD seedling data. There might be some reasons beyond this significant difference as the repair of 6-4 PP starts before that of CPD, which could have resulted in the collection of more excision products from 6-4 PP condition.

Lastly, all samples were subjected to PCA to have a general picture of variance across tissues and the replicates in Figure 1.66.

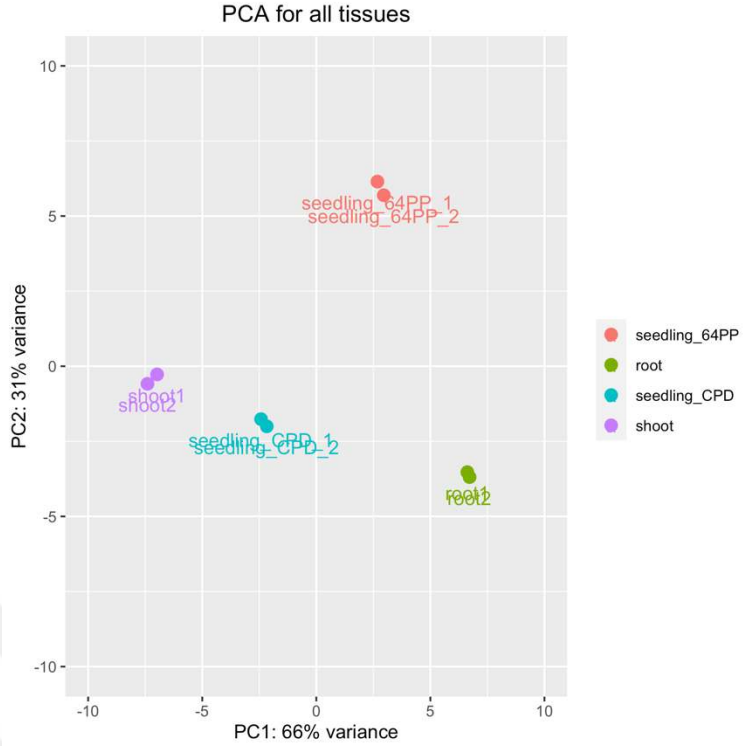

Figure 1.66: PCA plot for all samples: shoot, root, seedling CPD and seedling 6-4 PP XR-seq reads by DESeq2.

It is successfully represented in Figure 1.66 that all tissue and photoproduct type were clustered together, which support the idea of reliability on the replicates of experiment . Moreover, shoot CPD and seedling CPD data were clustered a bit closer to each other compared to root CPD and seedling 6-4 PP, which might be expected when considering the similarity of seedling and shoot tissues, which are together highly diverse from root and also the type of damage even if the tissue is seedling again.

As a general check of main principal components, a scree plot was generated below in Figure 1.67.

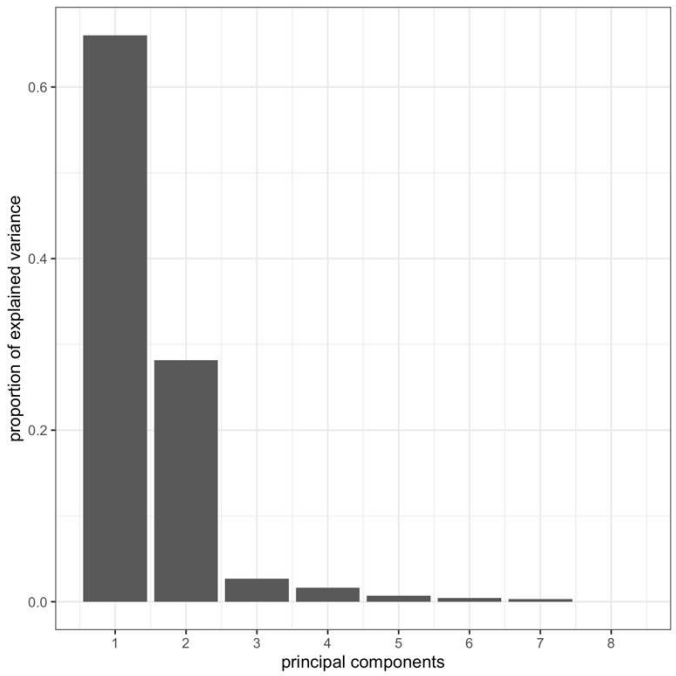

Figure 1.67: Distribution of proportion of principal components according to explained variance across all samples.

It is clear from Figure 1.67 that PC1 and PC2 are the principal components, which are the most effective and significant ones in explaining the variance across all samples, where PC1 explains the 66% of all the variance and PC2 31% of all the variance, respectively.

Moreover, XR-seq reads mapped to genes were also visualized in regions of gene bodies across for all genes. Figure 1.68, 1.69, and 1.70 below demonstrates gene bodies from all CPD XR-seq reads mapped to genes for each tissue.

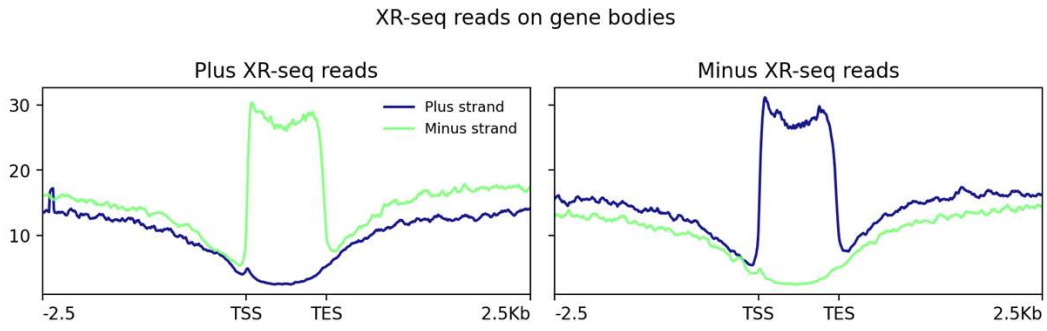

Figure 1.68: Shoot CPD XR-seq reads on gene body/length.

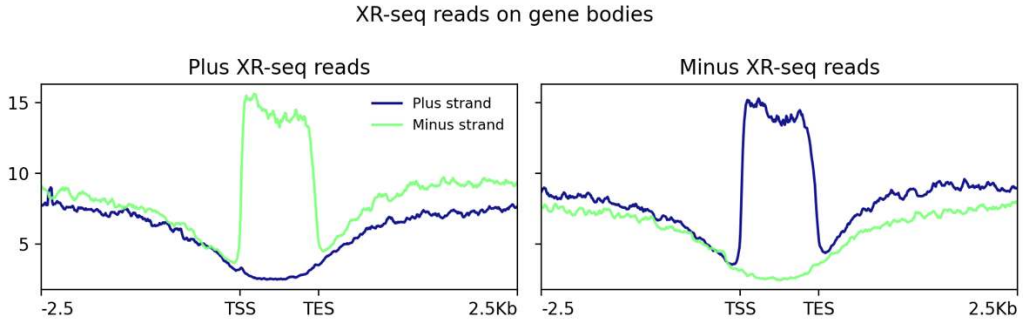

Figure 1.69: Root CPD XR-seq reads on gene body/length.

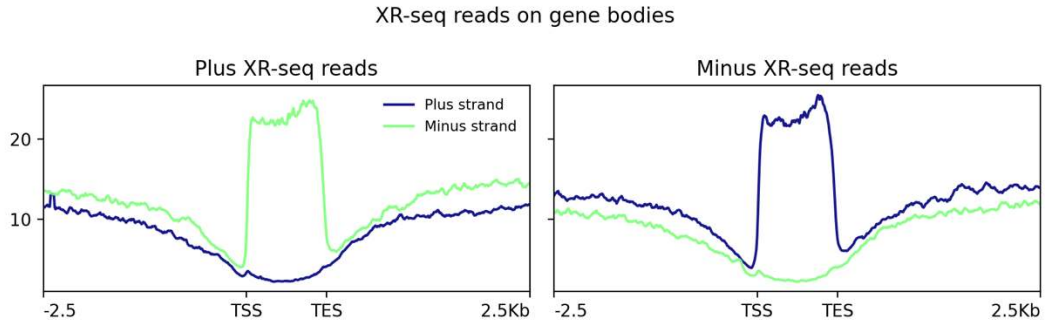

Figure 1.70: Seedling CPD XR-seq reads on gene body/length.

The figures represent the XR-seq read distribution over plus and minus strands of each gene. Moreover, non-template strand corresponds to coding and or plus strand of a gene while template strand accounts for minus or non-coding strand. TCR only occurs on the transcribed/template/non-coding strand while GGR could handle both strands. For instance, if a hypothetical plus XR-seq read is found to be mapped to minus or template strand, then it is an important indicator of TCR event for that gene and vice versa. In other words, if NER is observed in template strand of a gene, then it is reported that TCR mechanism might be on the action. The same trend has been observed for all tissues of CPD photoproducts.

1.4 Conclusion

As a conclusion, all XR-seq data from each tissue and damage type were reported to be reliable considering PCA plot, read length distribution, and nucleotide content frequency plots. On the other hand, TCR is still thought to be a key mechanism in driving nucleotide excision repair when TS/NTS distribution plots and XR-seq read distribution on gene body are taken into consideration. However, the origin of excision repair observed in root needs to be further investigated and presence of specific chemicals such as aflatoxin in soil have been thought so far as one of the potential factor to justify this phenomena observed in root tissue of *Arabidopsis thaliana*. Moreover, CPD seedling data demonstrated the superiority over 6-4 PP in terms of TCR as 6-4 PP seedling has shown that the number of genes whose TS/NTS ratio higher than 1 is less in comparison to CPD seedling.

Chapter 2:

THE FATE OF LYM1/2 PROTEINS DURING NODULATION IN *MEDICAGO TRUNCATULA*

2.1 Introduction

2.1.1 An Overview of Symbiotic Nitrogen Fixation

One of the main essential element in the nature, nitrogen, is crucial for organisms as they require and utilize a variety of nitrogen forms. Except the gaseous form N_2, nitrogen is utilized to synthesize building blocks of life such as DNA, RNA, and proteins (Pankievicz et al., 2019). The challenge of using atmospheric nitrogen gas is that the compound is quite stable and inert to be subjected to chemical reactions and it requires high amount of energy to break the triple bond between the two nitrogen atoms (Pankievicz et al., 2019). Figure 2.1 demonstrates the representative flow nitrogen in the circulating system.

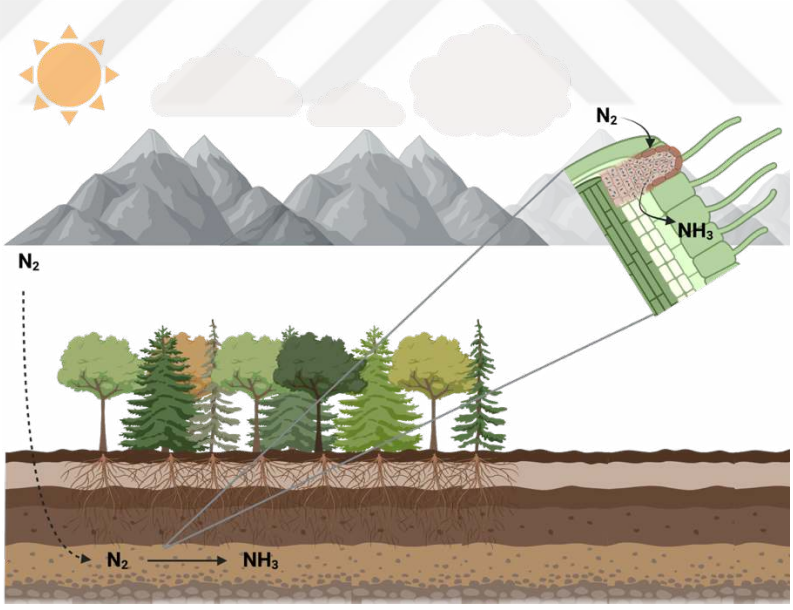

Figure 2.1: Atmospheric nitrogen fixation by rhizobia invaginated inside the leguminous hairy roots into ammonia by the use of plants (Created with BioRender.com)

In the context of industrial chemistry, Haber-Bosch process, named after two German scientists, has allowed the conversion of nitrogen gas into the form of ammonia and has

led to significant industrial advancements in various fields whereas nitrogenase enzyme had already been conducting the reaction of interest in nature. This enzyme catalyzes the reduction reaction, which serves as the most demanded available form of nitrogen and has been widely utilized in many sectors such as agriculture (Oztas, 2017; Roy et al., 2020). Nitrogenase is a metalloenzyme complex consisting of two subunits: dinitrogenase, with catalytic function, and dinitrogenase reductase, with iron-chelated properties (Pankievicz et al., 2019). This advancement in the industry has dramatically revolutionized the delivery of nitrogen to plants by adding it into soils. However, the way of nitrogen harvesting has significant economic consequences for the society (Mus et al., 2016; Roy et al., 2020; Rutten et al., 2020). Moreover, the fact that plants depend on fertilizers holds a significant problem when considering the dramatic rise in human population (Lopez et al., 2022; Mus et al., 2016). To address this urgent problem, symbiotic nitrogen fixation (SNF), a mutualistic relationship between leguminous plants and nitrogen fixing bacteria, has been extensively studied (Mus et al., 2016; Oztas, 2017; Rutten et al., 2020).

Medicago truncatula and *Lotus japonicus* are the leguminous plants widely investigated for their best nodulation traits (Rutten et al., 2020). Bacteria and archaea are able to fix atmospheric nitrogen and the process has been reported to be limited only for those domains of life (Mus et al., 2016; Pankievicz et al., 2019). SNF is based on the chemical share of some essential compounds between the interacting species. Nitrogen-fixing bacteria receive sugars from the plant, which is supplied fixed nitrogen by the bacteria (Luu et al., 2021; Mus et al., 2016; Roy et al., 2020). The process they take part is called nodulation where nodules are the compartments of differentiating bacteria in the roots of plants (Gautrat et al., 2019; Mus et al., 2016; Rutten et al., 2020). However, several efforts are ongoing to establish the SNF also in non-legume plants as well (Mus et al., 2016). In nature, nearly 20 nitrogen fixation (nif) genes have been reported to play crucial role in SNF and the main events for the process include the perception of nodulation factors, intracellular signaling in both symbiosis and immunity pathways, and also essential role of calcium signaling (Pankievicz et al., 2019; Zipfel & Oldroyd, 2017). In this symbiotic relationship, partners are able to alter, monitor, and regulate their metabolic pathways to better serve their individual needs. For instance, legumes can tightly control and halt any possible bacteria that lack in expressing traits necessary for nitrogen fixation. Moreover, bacteria in soil are invited by the attractant nature of several

plant exudates thanks to bacteria's chemotaxis capabilities (Mus et al., 2016; Roy et al., 2020). Moreover, they can improve a variety of plant traits to support the fixation process and help them coping with the environmental stresses (Pankievicz et al., 2019).

In conclusion, the effort of adapting the SNF process into cereal crops has already been put to satisfy the basic food needs (Zipfel & Oldroyd, 2017). Moreover, scientists are aware of the need of efficient biological nitrogen fixation process. To do so, they have been playing key roles in advancing and engineering the system to better utilize from a more sustainable process and also meet the food needs of humanity dramatically increasing in population all around the world.

2.1.2 Nodulation Signaling

When rhizobium arrives at the interface of legume interaction, a variety of phenolic molecules such as flavonoids are released by legumes for the uptake of rhizobium to trigger the chemotaxis under limited nitrogen availability (Gautrat et al., 2019; Mus et al., 2016; Oztas, 2017; Roy et al., 2020). These molecules are indeed effective on nodule development within the legume roots even if no rhizobium is present in the environment (Roy et al., 2020). The interaction between the two species leads to morphological changes in the host such as root hair curling and establishment of nodule primordium (Oztas, 2017). Upon the uptake of flavonoids by the bacteria, these molecules trigger the molecular pathways that make them express corresponding nodulation genes in rhizobia as it gives the first signal of the presence of a potential host ready to be connected with (Radutoiu et al., 2007). The downstream effects of the changes in transcriptional network of the bacteria are observed with the production of lipochitooligosaccharides (LCOs), also called Nod factors (Fliegmann et al., 2013; Luu et al., 2021; Rutten et al., 2020). In addition to LCOs, bacterial surface molecules such as exopolysaccharides, lipopolysaccharides, capsular polysaccharides, and cyclic β-glucan are also known to be essential for the construction of the symbiosis (Zipfel & Oldroyd, 2017). Nod factors operate in low concentration and are responsible for initiating and attaining the interaction between the rhizobia and the legume. Moreover, these Nod factors have been reported to be specific for host-rhizobium pairs such that interaction, infection, and nodulation processes between them become exclusive and strain-specific (Roy et al., 2020). Furthermore, all Nod factor types share the same N-acetylglucosamine backbone

structure with a fatty-acyl chain on the non-reducing end; however, they may show dissimilarities in backbone length, fatty-acyl chain and also the size (Mus et al., 2016; Oztas, 2017; Pankievicz et al., 2019). In addition to these differences, chemical modifications such as acetylation, sulfation, and fucosylation might be performed on these chains to promote and support legume-rhizobia specificity. (Mus et al., 2016). Upon the release of Nod factors, Nod factor receptors (NFRs), present in the cell membrane of the plant, shows affinity to those molecules, capture them, and then trigger the pathways associated with symbiosis in the host (Kelly et al., 2017). What might be interesting is the fact that the downstream cleavage of Nod factors on its backbone by the host Nod-factor hydrolase enzymes is believed to be supporting the further infection, colonization, and nodulation processes (Roy et al., 2020). The interaction between root hair and rhizobia for symbiotic relationship is shown in Figure 2.2 below.

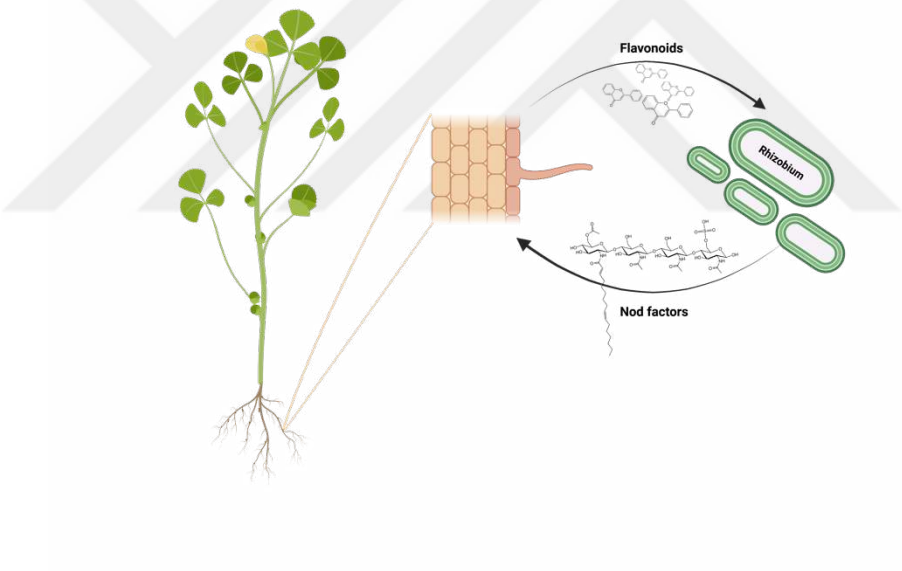

Figure 2.2: Release of Nod factor by rhizobia upon the arrival of flavonoids secreted by root hairs (Created with BioRender.com).

Moreover, pathogen-associated molecular pattern (PAMP) recognition occurs through the corresponding receptors, that are indeed receptor kinases or receptor-like proteins owned by the host (Zipfel & Oldroyd, 2017). Receptor kinases possess an extracellular domain responsible for ligand binding, and a single pass transmembrane domain. On the other hand, receptor-like proteins have only an extracellular domain but no intracellular domain. Besides its recognition by the receptors, PAMPs could also give

rise to production of reactive oxygen species (ROS) whose release into the infection thread is prevented by the presence of bacterial lipopolysaccharides and an increase in the intracellular calcium ion concentration (Oztas, 2017; Zipfel & Oldroyd, 2017). The latter may have a direct role in regulating symbiotic nitrogen fixation associated gene expression. Moreover, they are able to activate MAP kinases and Ca^{+2}-dependent protein kinases. From an evolutionary perspective, PAMP receptors are considered to have evolved via duplication, expansion, and divergence events as these receptors seem to be descended from only a limited number of taxonomic groups (Zipfel & Oldroyd, 2017). The main fluctuations in calcium concentration in the nucleus have been reported to be essential as it triggers a series of events associated with SNF (Zipfel & Oldroyd, 2017). Calmodulin-dependent kinase CCAMK in the presence of optimal Ca^{+2} concentration can phosphorylate a protein called IPD3, that has been determined to be a crucial player acting on the signaling pathway of SNF (Horváth et al., 2011; Oztas, 2017; Zipfel & Oldroyd, 2017).

2.1.3 Bacterial Invasion

Nod factors initiate cell division in the host, resulting in the curling shape of root hair and the formation of infection pocket. These factors give rise to calcium flux into the root of epidermal cells; therefore, membrane becomes depolarized and oriented towards high concentration of Nod factors by the action of alterations in cytoskeletal dynamics of the epidermal cells (Roy et al., 2020). Moreover, the operation of infection exerted by the rhizobium has different mode of actions: either triggering the infection thread formation (also known as root hair entry), or colonization in between epidermal cells (also known as crack entry) (Mendoza-Suárez et al., 2021; Mus et al., 2016). Subsequently, rhizobia start their journey so as to reach the nodule primordium through the infection thread, in which they are welcomed. The infection thread is separated from the host cytoplasm; the former is specially constructed to consist of only potential bacteria to be competent enough to acquire nitrogen fixing capabilities at the primordium. Here, the bacteria undergo a series of cell division event at the tips of infection threads and also get trapped and drifted by the flow of cell wall components (Roy et al., 2020). A visual representation of nodulation event is demonstrated in Figure 2.3 below.

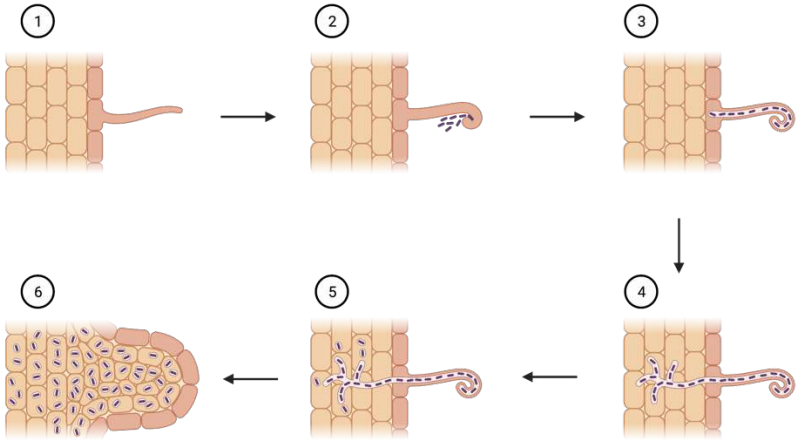

Figure 2.3: A step-by-step overview of nodule formation: 1) root-hair structure, 2) rhizobia trap by root, 3) infection thread formation, 4) extension of infection thread, 5) infection, 6) nodule formation (Created with BioRender.com).

Moreover, the inhibition of ROS production by the host has been reported due to Nod factors in the very earlier steps of the whole process. Nevertheless, ROS activity has been detected in a spatiotemporal fashion such that presence of ROS in cell walls and intercellular space of cortex, but not nitrogen-fixing zone, has been highlighted in different stages of SNF. It is believed that ROS efflux at later stages might contribute to the cell wall re-construction and mechanically support the infection threads via cross-linking nature of ROS (Chang et al., 2009). After traveling through epidermal and cortical tissue, rhizobia arrive nodule primordia, in which they begin to develop nodule tissues.

2.1.4 Nodule Development

At this stage of the process, bacteria are ultimately surrounded by the host cell, undergo a series of differentiation events and are now assumed to be an intracellular compartment of the host, called as symbiosome (Oztas, 2017). Each root nodule contains a few of bacteria, which undergo a series of morphological alterations from rod-shaped structure to branch-like-shaped one in order to differentiate and become compatible inside symbiosomes (Oztas, 2017; Zipfel & Oldroyd, 2017). In the case of rhizobium, they form mature indeterminate nodules, which has several distinct developmental zones: zone I, II, III, and IV. The first two zones correspond to meristem and infection zones,

respectively. Moreover, the last two zones are the ones where nitrogen fixation (zone-III), and senescence (zone-IV) take place, respectively (Farkas et al., 2014; Oztas, 2017; Roy et al., 2020). Those developmental zones are schematically represented in Figure 2.4 below.

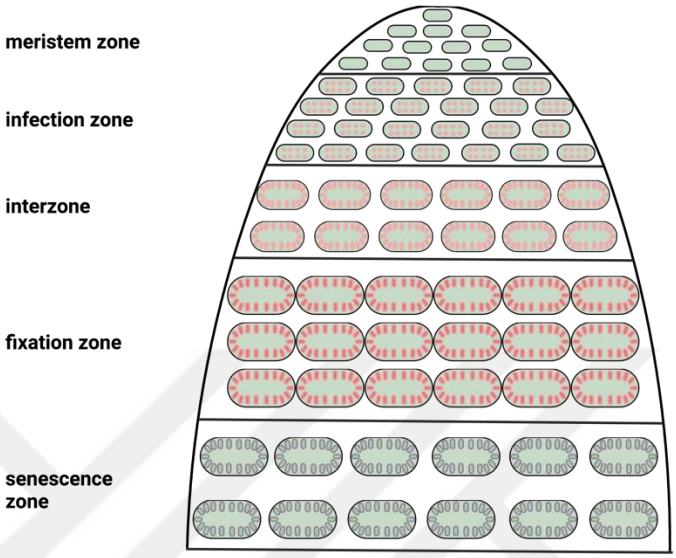

Figure 2.4: Developmental zones of rhizobium within the root of plants: meristem, infection, inter, fixation, and senescence zones, respectively (from top to bottom) (Created with BioRender.com).

During these physiological and morphological changes, multiple genomic endoreduplications as well as genome replications occur; thereby preserving the number of cells constant without any cytoplasmic divisions (Oztas, 2017; Roy et al., 2020). What has been suggested so far as a driving source of such differentiation events in the nodule is the production of nodule-specific cysteine rich (NCR) peptides, which results in the induction of the differentiation. Moreover, cysteine-rich peptides are reported to be an essential interference factor of the cell cycle, giving rise to the events such as endoreduplication and nodule differentiation (Roy et al., 2020). The morphology; however, is different in determinate legumes where no separate zone is observable; instead, the environment is homogenously distributed over the nodule tissue; resulting in the appearance of tissues corresponding only to nitrogen fixing and senescence events under the microscope (Mus et al., 2016; Oztas, 2017; Pankievicz et al., 2019). It is significant to emphasize that during the course of nodule development, metabolic processes are manipulated in a way that both species can benefit from their partner (Mus

et al., 2016; Roy et al., 2020). For instance, carbon metabolism of the host is modified to provide enough carbon source in the form dicarboxylates such as malate for bacteria's survival and symbiotic activities in the nodule (diCenzo et al., 2020). On the other hand, metabolic nitrogen flux of the guest is altered to take the advantage of nitrogen fixation by the plant via directly and or indirectly interfering with the amino acid metabolism of bacteria (Mus et al., 2016).

2.1.5 Nitrogen-Fixing Bacterium and Senescence

After completing a series of differentiation events experienced by bacteria in the symbiosome, at zone III, the bacteria is assumed compatible enough to possess the ability of fixing nitrogen. The key molecule exerting the unique chemistry on the nitrogen molecule and acting as a catalyst is called nitrogenase enzyme (Oztas, 2017). It is composed of two subunits, which include a heterotetrameric iron-molybdenum (Fe-Mo) unit acting as an electron acceptor and iron-only homodimeric unit responsible for donating electrons to the former subunit to perform the redox reaction of nitrogen (Oztas, 2017; Roy et al., 2020). The representative chemical equation of the conversion is shown in the Figure 2.5.

$$N_2 + 8\,H^+ + 8\,e^- + 16\,ATP \xrightarrow{\textbf{Nitrogenase}} 2\,NH_3 + H_2 + 16\,ADP + P_i$$

Figure 2.5: Chemical conversion of N_2 into NH_3 in the presence of ATP.

The above equation also depicts the change in oxidation state of nitrogen atom, where nitrogen gas itself has an oxidation number of 0 whereas resulting nitrogen from ammonia molecule has a -3 oxidation state.

Despite the essential role of the nitrogenase in converting the nitrogen into usable forms by the plants, the trade-off of the enzyme lies in its sensitivity to oxygen molecule, so called oxygen-labile nitrogenase. The fact that nitrogenase enzyme requires ATP, that is harbored by using oxygen for the respiration, to properly execute its function is actually not a barrier (Oztas, 2017). Another protein, called leghemoglobin, comes into play to overcome this barrier by acting as a suppressor of oxygen concentration in the nitrogen-fixing zone via directly transferring oxygen to the infected bacteroids (Mus et al., 2016).

Moreover, a variety of essential elements are required for a proper nitrogen fixation. These mainly include phosphate, iron, copper, cobalt, sulfur, and molybdenum. Phosphate is indispensable for nodule development, growth, and metabolism. The cofactor nature of iron chelated in the nitrogenase enzyme is also essential for SNF to properly function. Copper, on the other hand, joins in the structure of cytochrome oxidase, taking part in oxidative phosphorylation. Cobalt, also known as the core element for B12 production, is useful for SNF since B12 operates in the production exopolysaccharides (Mus et al., 2016). Moreover, sulfur atoms can be integrated into the structure of Fe-S nitrogenases, that is what makes sulfur one of the major micronutrients. Finally, molybdenum is incorporated in the synthesis of Fe-Mo type nitrogenases (Oztas, 2017).

Ultimately, bacteria can convert nitrogen into ammonia in the fixation zone and present it for the plant usage. The bacteria, which successfully complete the fixation task and their lifetime, travel deeper tissues called senescence zone. The zone is under the surveillance of multiple processes occurring during SNF and the fate of the bacteria transiting to senescence zone can be determined by multiple biological and environmental effects such as developmental processes, drought and nitrate levels in the soil (Roy et al., 2020). Furthermore, ROS has been put as an important evidence indicating the senescence event because of the iron atoms releasing from leghemoglobins, resulting in the loss of oxygen capture capability (Roy et al., 2020).

2.1.6 LysM Domain Containing Proteins

Innate immunity of the plant against potential invaders is a crucial adaptation to maintain the well-being of the host. However, the bacteria, although the bacterial species might be rhizobia to interact with the host, could be perceived as an enemy and the immunity acts on that particular pathogen. To enable rhizobia to safely travel through the plant tissues, the plant immunity against that species is attenuated step-by-step throughout SNF process (Oztas, 2017). The way that plants operate the perception of bacteria depends on the presence of plant pattern recognition receptors (PRRs) to make sure that the immunity acts specifically; thereby enabling the passage of the particular species but not the others (Oztas, 2017). PRRs are found in two distinct types depending on the recognition of the molecule or structure they bind: carbohydrates and proteins. Leucine-rich receptor proteins (LRR-RPs) and leucine-rich receptor proteins (LRR-RKs) are

integral proteins expressed to recognize protein-based structures such as bacterial flagellins (Chakraborty et al., 2019; Radutoiu et al., 2003; Zipfel & Oldroyd, 2017). On the other hand, lysin motif containing receptor (LysM) and receptor kinases, which were firstly unraveled in *Arabidopsis thaliana* and rice, are responsible for the perception of carbohydrate-based structures of the potential pathogens such as N-acetylglucosamine (GlcNAc) type of glycans while other types interact with patterns coming from rhizobia and fungi, which are associated with symbiotic processes (Bono et al., 2019). These LysM domain proteins are abundant in a variety of organisms including prokaryotes and eukaryotes but not in archaea domain of life. Moreover, the fact that glycans has conserved a specific GlcNAc pattern nearly all over the microbial species might indicate the presence of a common ancestor of LysM domain proteins associated with the binding of this pattern (Oztas, 2017). In the literature, three distinct types of LysM proteins have been identified so far, which include LysM receptor-like kinases such as CERK1 in rice, LYMs such as CEBIP (chitin elicitor binding protein), and also F-box LysM-type proteins (Oztas, 2017). CERK protein has the ability to recognize a variety of structures such as chitin, peptidoglycan, and also LCOs, which might indicate the use of CERK as co-receptor in targeting all of those ligands (Zipfel & Oldroyd, 2017).

Furthermore, *Medicago truncatula* is reported to possess two types of symbiotic receptor genes with LysM domains necessary for SNF. These include NFP (Nod factor perception) and LYK3 (Lysin Motif Receptor-like Kinase) type receptors (Oztas, 2017). These proteins have evolved to perform nodulation event together with NF recognition, and evolutionarily demonstrates an event of neofunctionalization history where tandem replications took place (Bono et al., 2019; Gautrat et al., 2019). The significant difference between these two protein in *Medicago truncatula* is the ability of LYK-type demonstrating kinase activity compared to the NFP one bearing a kinase domain, that is not functional (Oztas, 2017). Related to LYM type proteins, the subtype of LysM domain protein, they generally consist of a signal peptide in three LysM pattern length, and a GPI-attachment site, the latter of which helps the protein get anchored to the cellular membrane. In addition to those, unlike other LysM domain patterns, they do not have any intracellular domain responsible for downstream signaling (Zipfel & Oldroyd, 2017). CEBIP has different target molecules, for instance CEBIP, when dimerized with CERK in rice, is triggered by chitin in response to a fungal infection while *Arabidopsis thaliana* homolog acts against the presence of peptidoglycans although their structural homology

is conserved throughout the species in the line. In *Medicago truncatula*, LYM1 and LYM2 proteins have been reported to be essential for the perception of peptidoglycan and also activation of the host innate immune system; thereby inhibiting entry of the potential invasion attempt (Oztas, 2017). These LYM proteins are processed by post-translational modifications; thereby, acquiring N-glycosylation structures (Fliegmann et al., 2011). However, during the course of SNF, nodules have been revealed to contain LYM1 and LYM2 proteins, that are cleaved from their signal sequences, by the presence of an enzyme called DNF2 (defective in nitrogen fixation) so that the host defense against the pathogen is suppressed. The reason why it is named so is because mutant DNF genes halt SNF in different stages; showing the potential importance in their presence (Oztas, 2017). Figure 2.6 below shows a general overview of action of mechanism by which LYM proteins are processed.

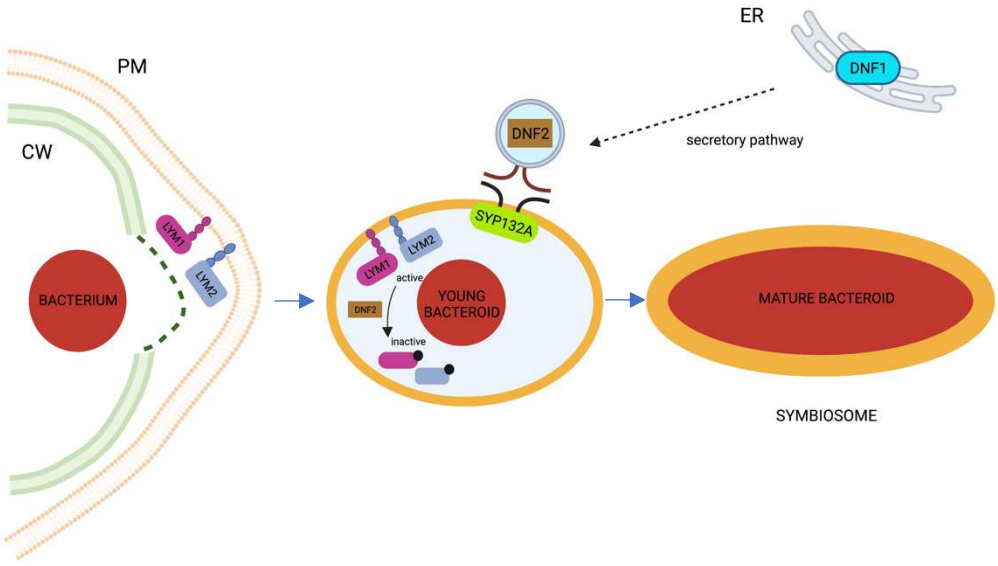

Figure 2.6: The processing and or inactivation of GPI-anchored LYM proteins by DNF2 protein invaginated into pre-symbiosomal structure with the help of SYP132A snare protein (Created with BioRender.com).

In this part of the thesis, LYM1/2 proteins were the center of our interest and their interactions with DNF2. To this end, these genes were cloned in their corresponding plant expression vectors and genotyping was performed to determine homozygous mutants of LYM genes.

2.2 Materials and Methods

2.2.1 RNA Isolation from Medicago truncatula Leaves

Leaves of wild-type A17 *Medicago truncatula* were cut and then subjected to RNA isolation by EURX's GeneMATRIX Universal RNA Purification Kit (Cat. No. E3599). In this protocol, 30 µl of activation buffer A was applied to homogenization spin column to activate the spin prior to material loading and kept at room temperature. For plants, protocol suggests a maximum of 100 µg of plant tissue to begin with. The tissue together with mortar and pestle was frozen under liquid nitrogen then grinded to have a powder-like material. Subsequently, it was placed in a RNAse-free eppendorf tube before adding 200 µl Lyse ALL and 100 µl RL buffers into the powder by mixing. The sample was centrifuged 4 minutes at maximum speed. After that, resulting supernatant was transferred into another clean RNAse-free eppendorf tube. Upon the transfer, 0.7 volume of buffer RL to the supernatant was added and mixed. Subsequently, the supernatant was transferred to the previously activated spin column, then centrifuged at 12,000 x g for 2 minutes in order to remove DNA. Then, 1.2 volume of absolute ethanol was added to the flow-through and it was mixed. Afterwards, the flow-through was transferred onto RNA binding spin-column and centrifuged at 11,000 x g for 1 minute and the resulting flow through was discarded. The procedure continues with the washing step. To do that, 500 µl of miRNA buffer was added to the RNA binding spin-column, then centrifuged at 11,000 x g for 1 minute. At this step, supernatant was discarded and this washing step including the miRNA buffer addition was repeated one more time. Lastly, to remove additional residues on the spin column, the sample was centrifuged at 11,000 x g for 1 minute. The spin column was placed into a new RNAse-free Eppendorf tube and 60 µl of RNAse free-water was added to elute DNA from the spin column to the Eppendorf. This was followed by the last centrifuge step applied at 11,000 x g for 2 minutes. At this stage, RNA eluted in the Eppendorf was ready to be stored at -20°C.

2.2.2 cDNA Synthesis

RNA isolated from the leaves was subjected to cDNA synthesis reaction at this step according to WISENT ADVANCED cDNA SYNTHESIS KIT (Cat. No. 801-100-XR). Before starting, 5 x cDNA mix was let to be thawed in the room temperature. Firstly,

4 µl of 5 x cDNA synthesis mix was added to the Eppendorf, then 1 µl of 20x RTase was added to reverse transcribe RNA template. Lastly, the RNA sample in amount of 4 pg - 4 µg was added to the sample, thereby completing the 20 µl final volume with molecular biology grade water. This reaction was incubated at 42°C for 30 minutes to enable the reaction to occur. Finally, the reaction was incubated at 85°C for 10 minutes to denature the reverse transcriptase enzyme.

2.2.3 PCR amplification of LYM1 coding sequence (CDS) and purification

Using Q5® High Fidelity Mastermix, LYM1 CDS was amplified using the following protocol in Table Y below.

Table 2.1: PCR protocol of LYM1 CDS amplification by Q5® High Fidelity Mastermix.

Components	25 µl PCR reaction
10 µM LYM1_F3	1.25 µl
10 µM LYM1_R3	1.25 µl
Template DNA (A17 gDNA)	5 µl
Q5® High Fidelity 2X Mastermix	12.5 µl
Nucleas-free Water	5 µl

LYM1 CDS was PCR amplified using the following PCR mixture and conditions.

Table 2.2: Thermocycling conditions for LYM1 CDS by Q5® High Fidelity 2XMastermix (Cat. No. M0492S).

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
30 Cycles	98°C	10 seconds
	54°C	30 seconds
	72°C	1 minute
Final Extension	72°C	2 minutes
Infinite hold	4°C	-

After checking the band length on the agarose gel, the corresponding band of LYM1 CDS was purified using Intron's MEGAquick-spin™ Plus Total Fragment DNA Purification Kit (Cat. No. 17290). To do so, the band was excised from the gel using a sterile scalpel. 20 µl PCR product loaded in the gel was transferred and then mixed with 100 µl BNL buffer by vortexing in a tube. The tube was incubated at 55°C for 10 minutes while making sure the gel is properly dissolved during this time. The tube was then incubated at room temperature and a column was placed into a collection tube. The sample mix was added to the collection tube through the column and then centrifuged at 11,000 x g for 30 seconds, thereafter discarding the flow through. Subsequently, 750 µl of Washing Buffer / Plus was added to the tube through the column and then centrifuged at 11,000 x g for 30 seconds thereafter discarding the flow through, again. In order to remove residual of buffers, an additional centrifugation step at 18,000 x g for 3 minutes was applied. Lastly, a new Eppendorf was placed for the column and then 40 µl of Elution Buffer was added at the center of the membrane to the Eppendorf through the column, followed by the final centrifugation 18,000 x g for 1 minute before storing at -20°C.

2.2.4 TOPO Cloning of LYM1 CDS into pCR®2.1 vector

Freshly obtained PCR product from the PCR reaction was transferred into pCR®2.1 vector. To do so, taq-treatment of the insert was performed to enable the insert to be compatible with pCR®2.1 vector insertion. 5 µl of PCR purified LYM1 CDS was mixed with 5 µl of Biomatik mastermix, then the resulting mixture was incubated at 72°C for 5 minutes. Afterwards, the following reaction was prepared and then incubated at 25°C for 30 minutes.

Table 2.3: TOPO Cloning recipe using pCR®2.1.

TOPO Cloning - pCR®2.1 vector
3.5 µl fresh taq-PCR product
1 µl salt solution
1µl TOPO vector
0.5 µl miliQ

The resulting reaction was transformed into DH5α. 1 vial of shot competent DH5α was thawed and then mixed with 2 µl of TOPO cloning reaction. The reaction was firstly incubated on ice for 20 minutes, then subjected to heat-shock for 30 seconds at 42°C without shaking. After completing the 30 seconds, the tube was immediately placed on ice. At this step, 250 µl of S.O.C medium was added to the tube and then the tube was incubated at 37°C for 1 hour with shaking. Once the incubation was completed, the sample was centrifuged at maximum for 1 minute. The resulting supernatant was discarded and the pellet was dissolved with 50 µl LB. The dissolved media was spread over LB-ampicilin, that is pre-treated with 120 µl Xgal and 40 µl IPTG for blue-white screening, and incubated at 37°C for overnight. Next day, colonies in white were selected and subjected to colony PCR reaction in table below to confirm the presence of LYM1 CDS.

Table 2.4: Colony PCR protocol of LYM1 CDS amplification by Biomatik 2X Mastermix.

Components	**25 µl PCR reaction**
10 µM LYM1_F3	1 µl
10 µM LYM1_R3	1 µl
Template DNA (A17 gDNA)	8 µl
Biomatik Taq 2X Mastermix	12.5 µl
Nuclease-free Water	2.5 µl

LYM1 CDS was PCR amplified using the following PCR mixture and conditions.

Table 2.5: Thermocycling conditions for LYM1 CDS by Biomatik 2X Mastermix.

Step	Temperature	Time
Initial Denaturation	94°C	3 minutes
30 Cycles	94°C	30 seconds
	57°C	1.5 minutes
	72°C	1 minute
Final Extension	72°C	10 minutes
Infinite hold	4°C	-

The positive colonies were selected to be grown in LB-ampicilin liquid medium overnight.

2.2.5 Plasmid Isolation from positive colonies

Overnight grown bacteria culture were subjected to plasmid isolation using Monarch® Plasmid Miniprep Kit (Cat. No. T1010L). To do so, 3 ml of the overnight culture was centrifuged for 30 seconds, then the supernatant was discarded. The resulting pellet was resuspended using 200 µl Plasmid Resuspension Buffer (B1) until no visible clumps can be observed. Afterwards, 200 µl of Plasmid Lysis Buffer was added, then gently inverted 5 times and incubated at room temperature for 1 minute. At this stage, the solution became viscous and the color change was observed. Subsequently, 400 µl of Plasmid Neutralization Buffer was added, gently inverted until the solution was neutralized by observing the precipitation, and then incubated for 2 minutes. Afterwards, the lysate was centrifuged at 16,000 x g for 5 minutes to make sure efficient removal of RNA. The resulting supernatant was transferred onto the spin column and then centrifuged at 16,000 x g for 1 minute; thereafter discarding the flow-through. The column was then reinserted in the collection tube and 200 µl of Plasmid Wash Buffer 1 was added for the centrifugation at 16,000 x g for 1 minute. Next, 400 µl of Plasmid Wash Buffer 2 was added and then centrifuged at 16,000 x g for 1 minute, again. The column was transferred onto another 1.5 ml Eppendorf tube and 40 µl of DNA elution buffer was

added to the center of the column. The tube was subjected to the spin centrifugation at 16,000 x g for 1 minute to elute DNA. The eluted DNA was stored at -20°C.

2.2.6 Double digestion and Ligation of pGWB12 with LYM1 – CDS

In this part of the study, pGWB12 expression vector and pCR®2.1 vector were double digested with XbaI and SacI. To do this, the following reactions were set and then incubated at 37°C for 3 hours.

Table 2.6: Double digestion of pGWB12 and pCR®2.1-LYM1 with XbaI and SacI.

pGWB12 digestion	**pCR®2.1-LYM1 digestion**
30 µl pGWB12	3 µl pCR2.1-LYM1
5 µl 10x rCutsmart Buffer	5 µl 10x rCutsmart Buffer
1 µl XbaI	1 µl XbaI
1 µl SacI	1 µl SacI
13 µl nuclease-free water	40 µl nuclease-free water

After 1st hour of the 3-hours incubation, alkaline phosphatase was added to the reaction containing pGWB12 vector to prevent it from self-ligating. Afterwards, the reactions were run on gel and corresponding gels were purified from the gel using Intron's MEGAquick-spin™ Plus Total Fragment DNA Purification Kit (Cat. No. 17290). The ligation reaction was set 3:1 ratio of insert to vector using the following ingredients with overnight incubation at 16°C (Cat. No. M0202L).

Table 2.7: Ligation of pGWB12 and LYM1-CDS using T4 Ligase.

Ligation reaction of pGWB12 and LYM1 – CDS
9 µl LYM1-CDS (3X)
1 µl pGWB12 (X)
1 µl T4 enzyme
2 µl T4 buffer
7 µl MB grade water

Subsequently, the resulting ligation reaction was transformed into DH5α. Next day, the colonies were screened by PCR in the following conditions and then positive colonies were grown overnight in a liquid culture containing LB-kanamycin. After all, overnight-grown bacteria were subjected to plasmid isolation using the Monarch® Plasmid Miniprep Kit (Cat. No. T1010L).

To confirm the identity of the construct, double digestion for pGWB12 having LYM1-CDS was performed using the following protocol.

Table 2.8: Recipe for double digestion of pGWB12:LYM1-CDS.

Double digestion of pGWB12:LYM1-CDS
12 µl Purified plasmid
5 µl 10X rCutSmart Buffer
1 µl *EcorI-HF*
1 µl *HindIII-HF*
31 µl nuclease-free water

A cloning summary of LYM1 into pGWB12 is given in Figure 2.7 below.

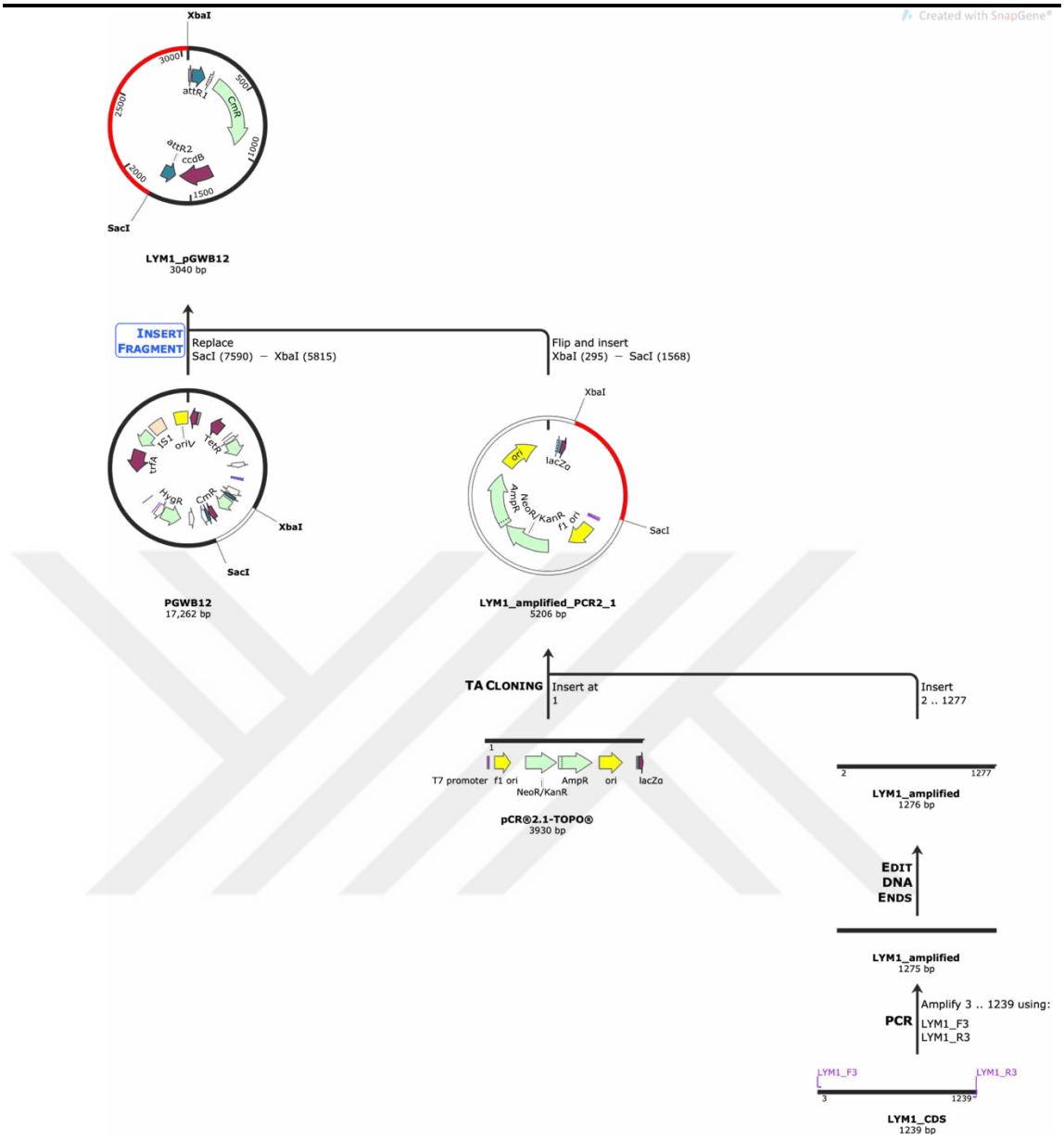

Figure 2.7: Cloning steps of LYM1 into pGWB12 expression vector (Created with SnapGene).

2.2.7 PCR amplification of LYM2 coding sequence (CDS) and purification

Using Q5® High Fidelity Mastermix, LYM2 CDS was amplified using the following protocol in Table 2.9 below.

Table 2.9: PCR protocol of LYM1 CDS amplification by Q5® High Fidelity Mastermix.

Components	25 µl PCR reaction
10 µM LYM2_F2	1.25 µl
10 µM LYM2_R2	1.25 µl
Template DNA (A17 gDNA)	5 µl
Q5® High Fidelity 2X Mastermix	12.5 µl
Nuclease-free Water	5 µl

LYM1 CDS was PCR amplified using the following PCR mixture and conditions.

Table 2.10: Thermocycling conditions for LYM2 CDS by Q5® High Fidelity 2X Mastermix (Cat. No. M0492S).

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
30 Cycles	98°C	10 seconds
	54°C	30 seconds
	72°C	1 minute
Final Extension	72°C	2 minutes
Infinite hold	4°C	-

2.2.8 Cloning of LYM2 CDS into pCR®8 vector (TOPO cloning)

Freshly obtained PCR product from the PCR reaction was transferred into pCR®8 vector (Cat. No. 12535-029). Initially, taq-treatment was applied to the PCR product in the ratio of (1:1) PCR product: Taq polymerase mixture at 72°C for 5 minutes, which can be stored at 4°C. Then, the following recipe was prepared to allow TOPO cloning to take place at room temperature for 30 minutes.

Table 2.11: TOPO Cloning recipe using pCR®8.

TOPO Cloning - pCR®8 vector
4 µl fresh taq-PCR product
1 µl salt solution
1 µl miliQ
0.5 µl pCR®8 TOPO vector

After 30 minutes incubation, 1 vial of shot competent DH5α was thawed and then mixed with 2 µl of TOPO cloning reaction. The reaction was firstly incubated on ice for 20 minutes, then subjected to heat-shock for 30 seconds at 42°C without shaking. After completing the 30 seconds, the tube was immediately placed on ice. At this step, 250 µl of S.O.C medium was added to the tube and then the tube was incubated at 37°C for 1 hour with shaking. Once the incubation was completed, the sample was centrifuged at maximum for 1 minute. The resulting supernatant was discarded and the pellet was dissolved with 50 µl LB. The dissolved media was spread over LB-spectinomycin and incubated at 37°C for overnight.

2.2.9 Colony Screening with PCR and Plasmid Isolation

Next day, the colonies were screened by PCR in the following conditions and then positive colonies were grown overnight in a liquid culture containing LB-kanamycin.

Table 2.12: Colony PCR protocol of LYM2 CDS amplification by Biomatik 2X Mastermix.

Components	25 µl PCR reaction
10 µM LYM2_F2	1 µl
10 µM LYM2_R2	1 µl
Template DNA (A17 gDNA)	8 µl
Biomatik Taq 2X Mastermix	12.5 µl
Nuclease-free Water	2.5 µl

LYM1 CDS was PCR amplified using the following PCR mixture and conditions.

Table 2.13: Thermocycling conditions for LYM2 CDS by Biomatik 2X Mastermix.

Step	Temperature	Time
Initial Denaturation	94°C	3 minutes
30 Cycles	94°C	30 seconds
	57°C	1.5 minutes
	72°C	1 minute
Final Extension	72°C	10 minutes
Infinite hold	4°C	-

After all, overnight-grown bacteria were subjected to plasmid isolation using the Monarch® Plasmid Miniprep Kit (Cat. No. T1010L).

2.2.10 Restriction Enzyme Digestion with EcorI

Next day, purified DNA was subjected to plasmid digestion with *EcorI* for 2 hours at 37°C , that is followed by 65°C heat inactivation using the following recipe.

Table 2.14: Digestion of pGWB12-LYM2 CDS with EcorI.

Restriction Enzyme Digestion of pGWB12-LYM2 with *EcorI*
2.5 µl fresh taq-PCR product
5 µl 10X buffer
1 µl *EcorI*
0.5 µl TOPO vector

Next, the digested products were run on gel and confirmed to be properly digested.

2.2.11 Cloning of the insert into pGWB12 plant expression vector via LR reaction.

The digested product of pCR®8 was subjected to another cloning into plant expression vector called pGWB12. The procedure was applied according to Gateway® Technology with Clonase® II (Cat. No. 12535-029). To do so, entry clone (50 - 150 ng in 1-7 µl) was added to a sterile eppendorf tube. Then, destination vector, pGWB12 (150 ng/µl) was added and then total volume was completed to 8 µl using TE buffer (pH = 8.0). Afterwards, 2 µl of Gateway® LR Clonase® II enzyme mix, after briefly vortexed,

was added to the Eppendorf tube containing entry (PCR8) and destination clones (pGWB12). The components were mixed then the reaction was subjected to the incubation at 25°C for 1 hour. Then, this reaction was further treated with 1 µl of 2 µg/µl Proteinase K solution at 37°C for 10 minutes. The resulting reaction was transformed into DH5α. Next day, positive colonies were screened using the following PCR protocol and were grown in LB-kanamycin overnight.

Next day, positive colonies were subjected to plasmid isolation using Monarch® Plasmid Miniprep Kit (Cat. No. T1010L). Next, the construct was sent to standard sequencing facility of Macrogen Facility in Netherlands.

A cloning summary of LYM2 into pGWB12 is given in Figure below.

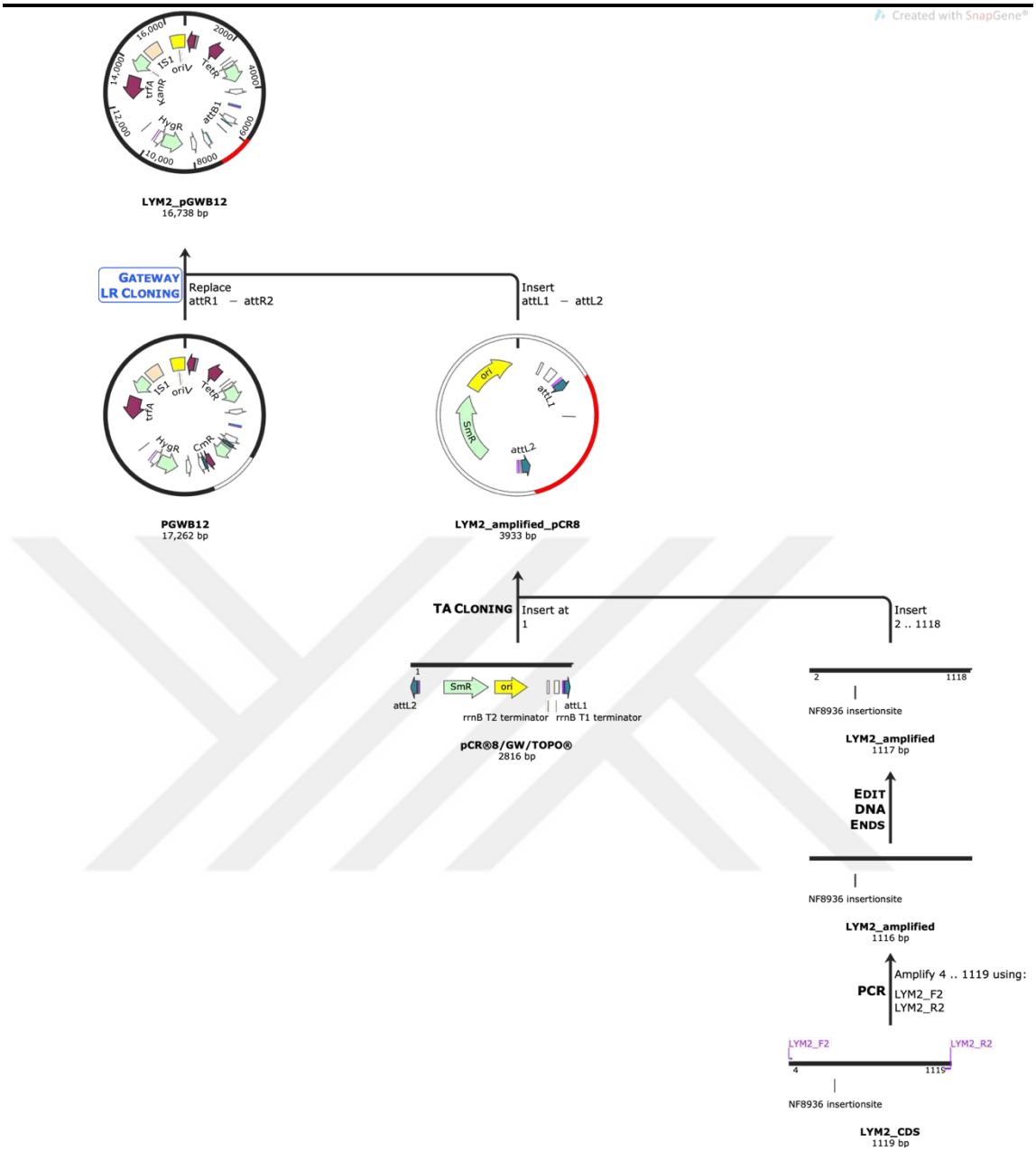

Figure 2.8: Cloning steps of LYM1 into pGWB12 expression vector (Created with SnapGene).

2.2.12 Transformation of expression vector into plant-transformable agrobacterium strain, ARqua1 bacteria.

The full construct was transformed into ARqua1, that can be directly introduced to plants. To do that, 2 µl of the construct was added into a one-shot ARqua1 competent cell stored at -80°C and mixed. After that, the bacteria together with the construct was put in a pre-cooled electroporation cuvette, and then the cuvette was placed in electroporation device to pulse the bacteria for a quick moment. Next, electroshocked

bacteria was diluted with 900 µl S.O.C medium, and then by pipetting both of them, the solution was transferred into a new Eppendorf tube. This tube was rotated at 30°C for 1.5-2 hours. Upon completing the incubation diluted and non-diluted bacteria medium was put into LB-kanamycin agar plates and then incubated at 30°C for 2 days. After two days, the colonies in diluted plate was screened by PCR to confirm the presence of the construct and also glycerol stock from this positive colonies containing our construct of interest was prepared.

2.2.13 Hairy root transformation

ARqua1, an agrobacterium strain, was subjected to electroporation using the construct of interest. To do so, 1 µl of construct was introduced to a shot of Arqua1 for 20 minutes then transferred into a electroporation cuvette. Subsequently, the sample was subjected to pulse to enable the construct to be electroporated into bacteria. Then, the bacteria in the cuvette was diluted in 900 µl S.O.C medium and pipetted up and down. After that, the resulting sample was rotated in a 30°C incubator for 1.5 hours, which was followed by the plating of grown bacteria for 2 days at the same temperature. Upon 2 days of growth, single colonies were screened to determine if coding sequence within construct is present. These bacteria were then grown in liquid culture, stored in glycerol and then used for hair root transformation. Moreover, the seeds of Jemalong A17 ecotype of *Medicago truncatula* legume plant were scraped and then disinfected with 3% bleach solution. After a few washing steps, they were rotated in distilled water overnight for germination. In the following day, they became germinated and are cut from their radicle site with a blazer so that they can be rubbed onto the bacterial colonies containing the construct of interest. They were transferred into a Fahräeus - agar medium for 1 week. Afterwards, plants were transferred into another Fahräeus - agar medium with the antibiotics that the construct is resistant with, that is kanamycin for 2 weeks. Following this incubation, plants were placed in a 1:1 turface:sand medium and were then treated with *Sinorhizobium medicae* strain ABS7 in ½ X SNM medium for nodulation and the nodule organs were expected to be arising from the 15^{th} day. The growth condition for the plants consists of two cycles: 25°C for 16 hours in light and 18°C for 8 hours in dark. The following formula corresponds to liquid, nitrogen free Fahräeus medium used in growth pouches. Combined nitrogen (NH_4NO_3 or KNO_3) can be added to the desired

concentration. For Fahräeus agar, 15 g/l plant agar was added before medium sterilization.

Table 2.15: Recipe for Fahräeus medium.

Stock Solutions	**Stock concentration**		**Volume (ml) (for *1 L* of 1X medium)**	**Final concentration**
Macronutrients	**g/l**	**M**		
$MgSO_4 \cdot 7H_2O$	123.2	0.5	1.0	0.5 mM
KH_2PO_4	95.3	0.7	1.0	0.7 mM
$Na_2HPO_4 \cdot 2H_2O$	71.2	0.4	2.0	0.8 mM
Fe-EDTA[2]	20 mM		2.5	50 µM
Micronutrients				
$MnSO_4$, $CuSO_4$, $ZnSO_4$, H_3BO_3, Na_2MoO_4	1 mg/ml each		0.1 / each	0.1 µg/ml each

Fe-EDTA solution was prepared using solutions of 5.6 g/l FeSO4 and 7.4 g/l Na_2EDTA by heating at 50°C and mixed. pH of the solution was adjusted to 6.5 using H_2SO_4. The medium was aliquoted into 250 or 500 ml flasks and sterilized at 120°C for 20 minutes by autoclaving. Finally, $CaCl_2$ (1 M stock solution) must be added to the medium (1 mM final concentration) after autoclaving and just before use.

2.2.14 Plant Growth Conditions and Inoculation

In this study, Jemalong A17 ecotype of *Medicago truncatula* legume plant was used. Plants were treated with *Sinorhizobium medicae* strain ABS7 in 1/2X BNM medium for nodulation and the nodule organs were expected to be arising from 15 days. The growth condition for the plants consists of two cycles: 25°C for 16 hours in light and 18°C for 8 hours in dark. ABS7 was grown in TY medium at 30°C for 2 days and the bacteria were then subcultured, and subjected to dilution ratio of OD_{600}=0.01/ml to be transferred into the roots of A17 seedling.

2.2.15 Genotyping of lym1 and lym2 mutants

In order to determine the genotype of seeds if they are homozygous mutant or not, first of all, randomly mutated R108 *Medicago truncatula* seeds were germinated and grown in Fahräeus medium for 1 week in the growth chamber. Next, gDNA isolation from the leaves of Medicago was performed following the CTAB method. The protocol starts with the disruption of leaf tissue using CTAB buffer itself, magnetic beads and Analytik Jena SpeedMill PLUS Homogenizer. The recipe for the CTAB is given below.

Table 2.16: Recipe for CTAB buffer.

CTAB buffer
0.2 g Hexadecyltrimethylammonium bromide
1.44 g NaCl
4 ml, 50 mM EDTA
1 ml, 1M Tris-HCl (pH=8.0)

The resulting lysate was incubated at 65°C for 30 minutes, followed by 25°C - 10 minutes incubation. The lysate was transferred into an Eppendorf, then the same volume of chloroform was added, inverted and subjected to cold centrifugation at 9,500 x g for 15 minutes at 4°C,. Next, the upper phase was transferred to a new Eppendorf and then it was mixed with 0.7 volume of absolute isopropanol, mixed and kept at room temperature for 10 minutes incubation. Subsequently, the solution was centrifuged at 4°C, 7,500 x g for 20 minutes. The supernatant was discarded and the pellet was dissolved in 1X TE buffer, then stored at -20°C for further downstream application, that is genotyping.

Genotyping PCR was designed in a way that each sample was amplified using both wild type and mutant primers as two separate reactions.

PCR protocol for LYM1 genotyping was given in Table 2.17 and 2.18 below for both wild type and mutant reactions.

Table 2.17: PCR protocol for LYM1 genotyping check for wild-type primers by Biomatik 2X Mastermix.

Components	25 µl PCR reaction
10 µM LYM1_GF	1 µl
10 µM LYM1_GR2	1 µl
gDNA from potential A17 mutant	1 µl
Biomatik Taq 2X Mastermix	12.5 µl
Nuclease-free Water	9.5 µl

Table 2.18: PCR protocol for LYM1 genotyping check for mutant primers by Biomatik 2X Mastermix.

Components	25 µl PCR reaction
10 µM FST_F	1 µl
10 µM LYM1_GR2	1 µl
gDNA from potential A17 mutant	1 µl
Biomatik Taq 2X Mastermix	12.5 µl
Nuclease-free Water	9.5 µl

Thermocycling conditions for LYM1 genotyping was listed in Table 2.19.

Table 2.19: Thermocycling conditions for LYM1 and LYM2 CDS by Q5□ High Fidelity 2XMastermix (Cat. No. M0492S).

Step	Temperature	Time
Initial Denaturation	94°C	3 minutes
30 Cycles	94°C	30 seconds
	54°C	1.5 minutes
	72°C	1 minute
Final Extension	72°C	10 minutes
Infinite hold	4°C	-

Moreover, PCR protocol for LYM2 genotyping was given in Table 2.20 and 2.21 below for both wild type and mutant reactions.

Table 2.20: PCR protocol for LYM2 genotyping check for wild-type primers by Biomatik 2X Mastermix.

Components	25 µl PCR reaction
10 µM LYM2_GR	1 µl
10 µM LYM2_F2	1 µl
gDNA from potential A17 mutant	1 µl
Biomatik Taq 2X Mastermix	12.5 µl
Nuclease-free Water	9.5 µl

Table 2.21: PCR protocol for LYM2 genotyping check for mutant primers by Biomatik 2X Mastermix.

Components	25 µl PCR reaction
10 µM FST,-F	1 µl
10 µM LYM2_F2	1 µl
gDNA from potential A17 mutant	1 µl
Biomatik Taq 2X Mastermix	12.5 µl
Nuclease-free Water	9.5 µl

Thermocycling conditions for LYM2 genotyping was listed in Table 2.22.

Table 2.22: Thermocycling conditions for LYM1 and LYM2 CDS by Q5□ High Fidelity 2XMastermix (Cat. No. M0492S).

Step	Temperature	Time
Initial Denaturation	94°C	3 minutes
30 Cycles	94°C	30 seconds
	54°C	1.5 minutes
	72°C	1 minute
Final Extension	72°C	10 minutes
Infinite hold	4°C	-

Primers, which were utilized in this study are given in Table 2.23 below.

Table 2.23: Primers used in this thesis.

Primer Name	Target	Sequence (5'->3')
LYM2_F2	LYM2 – CDS, Genotyping	AGTGGAGGGAACAAGATGGTTACTG
LYM2_R2	LYM2 - CDS	CTACAAAAGATAGACAAATAAGATGAGAAAGTGAATGAG
M13_F	pCR®8 Colony Screening	TGTAAAACGACGGCCAGT
LYM1_F3	LYM1 - CDS	TCTAGAATGGATTACAAGGATGACGACGATAAGCAGAACAAAAACATTCTATTATT GTTCCTCC
LYM1_R3	LYM1 - CDS	GAGCTCTTACAATGCTACAGGGATCATCAAC
LYM2_GR	Genotyping	CTGGAAGGGGAACATCAAGA
LYM1_GF	Genotyping	TGCACAACCTCCGACACATG
LYM1_GR2	Genotyping	CAGTGTCCTGCTGTAATA
FST_F	Genotyping	AATGAGATGGACATCATCA
FST_R	Genotyping	TGATGATGTCCATCTCATT
pG_F	Custom DNA sequencing	ACAAGTTTGTACAAAAAAGCA

2.3 Results and Discussion

2.3.1 LYM1 and LYM2 CDS PCR amplification

Firstly, total RNA extraction from the leaves of WT A17 Medicago truncatula was performed and run on the gel given in Figure 2.9.

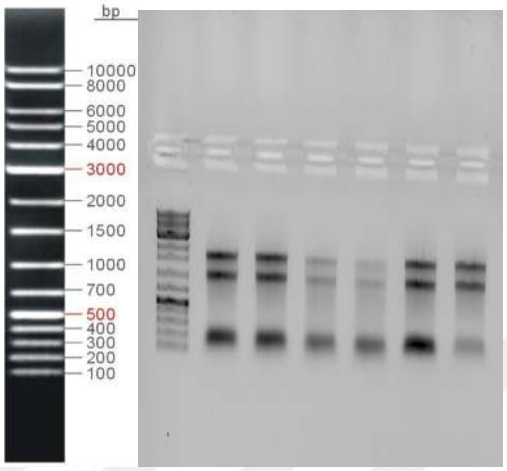

Figure 2.9: RNA isolated and run on a 1% agarose gel.

After isolation, RNA was converted to complementary DNA (cDNA), which does not include any intron sequences but exon sequences in order to amplify LYM1 and LYM2 coding sequence. Using forward and reverse primers, LYM1 and LYM2 CDSs were amplified and run on a gel demonstrated in Figure 2.10.

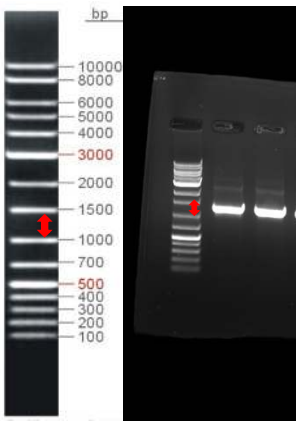

Figure 2.10: LYM1 CDS (a) and LYM2 (b) CDS run on a 1% agarose gel and ladder by Biomatik 1kb plus ladder.

The size of CDSs are 1237 and 1116 bp for LYM1 and LYM2, respectively. After confirming the band size of LYM1, that is 1247 bp in length, the resulting band was extracted from the gel and stored at -20°C for downstream applications

2.3.2 Cloning of LYM1_CDS into pGWB12

Before transferring the insert of interest, the insert was subjected to Taq-treatment as the Taq-polymerase adds A overhangs to the blunt-ended LYM1-CDS and then was directly inserted into a vector called pCR2.1 via TOPO cloning. The orientation of the insert into the vector is random and it does not affect the downstream applications. The reaction was transformed into DH5α and positive colonies were screened using blue-white screening. Next, PCR2.1:LYM1-CDS construct and pGWB12 plasmids were individually double digested with XbaI and SacI restriction enzymes, incubated at 37°C for 3 hours. One of the reasons why pGWB12 is used is because N-terminal FLAG sequence attached to LYM proteins were reported to be intact but not other conformations (Oztas, 2017). In the beginning of the first hour, alkaline phosphatase was added to the reaction containing pGWB12 vector to prevent self-ligation. The resulting reaction was ligated using T4 ligase overnight at 16°C and then transformed into DH5α and the colonies were screened to continue with the downstream applications. A negative control reaction for ligation without the insert was also designed. No colonies were concluded in negative control plate while the reaction was successfully completed. Colonies were grown, plasmid isolated. To confirm if transformation occurred or not, the plasmid was double digested for two hours. The resulting digestion was run on a gel and visualized in Figure 2.11 below.

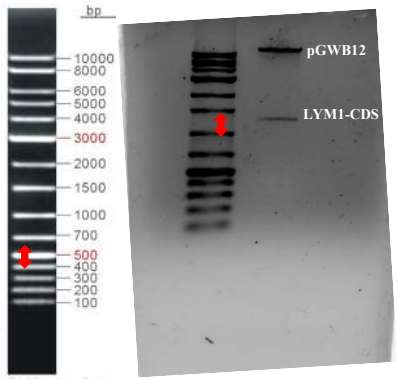

Figure 2.11: Double digestion of pGWB12:LYM1-CDS construct and ladder by Biomatik 1kb plus ladder.

With this confirmation, it was concluded that LYM1-CDS was successfully cloned into pGWB12 plant expression vector.

2.3.3 Cloning of LYM2_CDS into pGWB12

After amplifying LYM2 CDS, the insert was cloned into pCR8 vector. The resulting reaction was transformed into DH5α, grown in LB-kanamycin medium and then colonies were screened with PCR. Positive colonies were subjected to plasmid isolation, followed by a digestion reaction by EcoRI, resulting the formation of 1134 and 2800 bp digested regions. The resulting reaction was run on a gel given in Figure 2.12.

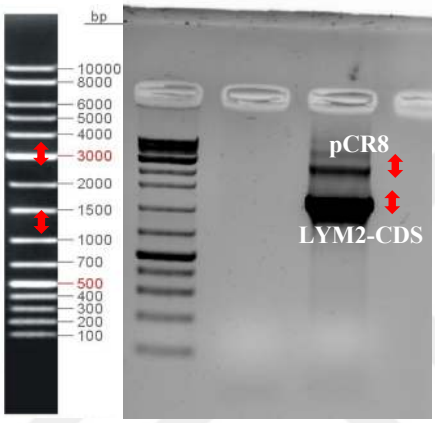

Figure 2.12: pCR8:LYM2-CDS construct after digestion with EcoRI in reference with the ladder by Biomatik 1kb plus ladder.

After confirming the successful cloning into pCR8, LYM2-pCR8 construct was subjected to LR reaction with pGWB12. The reaction takes place in a way that the reactants were cleaved at their attR1 attR2 and attL1 attL2 sites for entry clone and the destination vector respectively. The insert in the entry clone was incorporated into the site of cleaved region in the destination vector. The resulting clone is LYM2-CDS:pGWB12 and the byproduct is a toxic construct. The following figure demonstrates the attL1, attL2 sites in pCR8-LYM2-CDS entry clone and attR1, attR2 sites in pGWB12 destination vector.

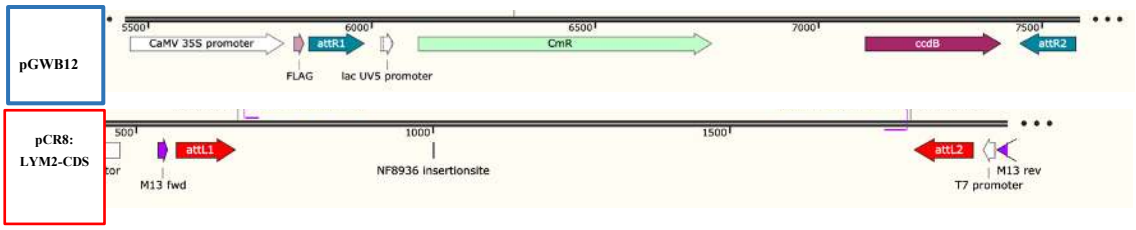

Figure 2.13: attR1, attR2 sites of pGWB12 destination vector and attL1, attL2 sites of pCR8-LYM2-CDS entry clone compatible for LR reaction.

The main product from this reaction was transformed into DH5α and the colonies were screened to be subjected with plasmid isolation. After plasmid isolation, the plasmid was double digested with SacI and XbaI to demonstrate the success of the cloning. The following figure was generated as a result of the double digestion.

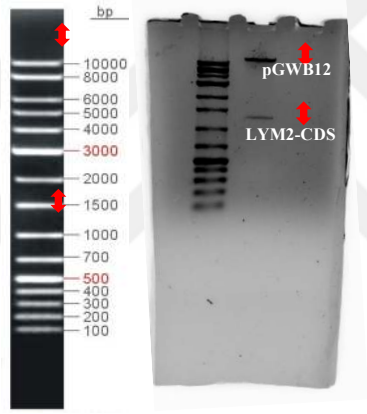

Figure 2.14: pGWB12:LYM2-CDS double digested with SacI and XbaI in reference with the ladder by Biomatik 1kb plus ladder.

Moreover, the construct was sequenced and found to be properly formed. However, the disadvantage of using LR reaction is the random orientation of the insert within destination vector. LYM2-CDS was found to be in the proper orientation; however, LYM1-CDS was incorporated inversely in the destination vector. That is why a new cloning method, that is conventional cloning, was designed in order to clone LYM1-CDS into pGWB12 expression vector as mentioned before

2.3.4 Transformation of expression vectors with Arqua1 into hairy root

Seeds were initially disinfected and let to be germinated. Afterwards, they were cut at their root-originating tissue to enable agrobacterium to enter the plant. Firstly,

Fahreäus medium with no antibiotics were used to grow seeds on for 1 week. Next, the roots of seedlings were transferred into another Fahreäus medium for 2 weeks, which was prepared to contain the antibiotic kanamycin, whose resistance gene was present in the construct to be introduced so that only the seeds that have received the bacteria together with the construct, would grow. The following figure shows the seedlings grown in Fahreäus medium.

Figure 2.15: Agrobacterium-infected seedlings grown in Fahreäus medium.

2.3.5 Genotyping for mutant LYM1 and LYM2 plants

Firstly, LYM1 seedling were subjected to DNA isolation, which was then followed genotyping experiment by PCR. FST is an insert incorporated into the gene of interest. To decide whethere the corresponding seedling is genotypically homozygous mutant or not, the following primers were used.

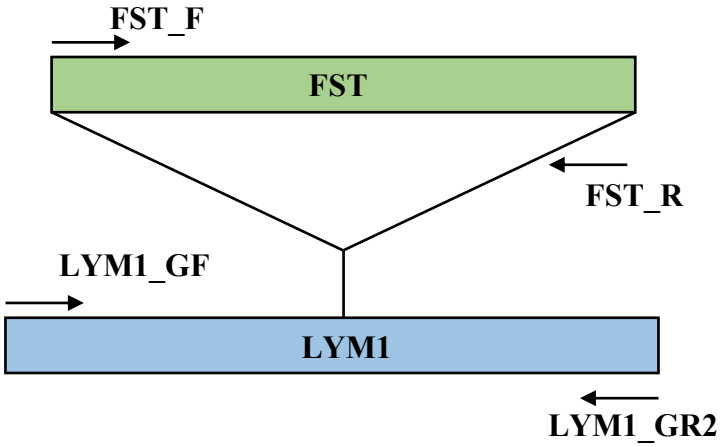

Figure 2.16: A representation of FST insertion into LYM1 together with corresponding primers.

The band which would be in length of size amplified by LYM1_GF and LYM1_GR2 primers corresponds to the WT genotype whereas the region amplified by FST_F & LYM1_GR2 or LYM1_GF & FST_R would be the indicator of mutant genotype. The primers that were used include LYM1_GF, LYM1_GR2, and FST_F. The following gel was obtained as a result of PCR reaction.

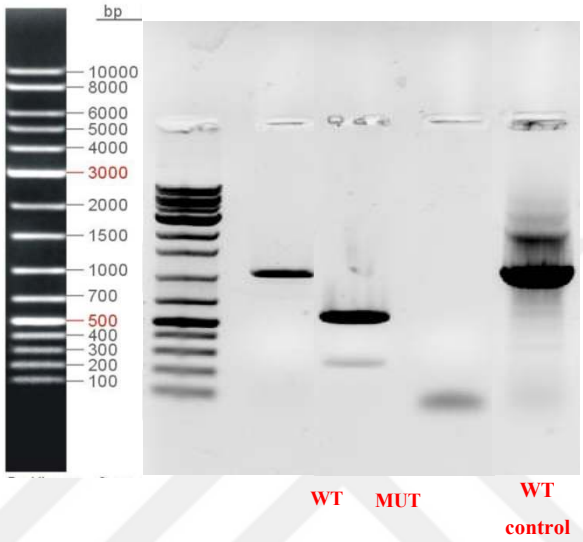

Figure 2.17: LYM1-10 wild-type band, mutant band from two different samples, and positive wild-type control and ladder by Biomatik 1kb plus ladder.

Similarly, the same applies for LYM2 genotyping. However, no mutant was observed in this screening. Instead, the inverted orientation of FST into LYM2 was taken into consideration and the following schematic representation was created.

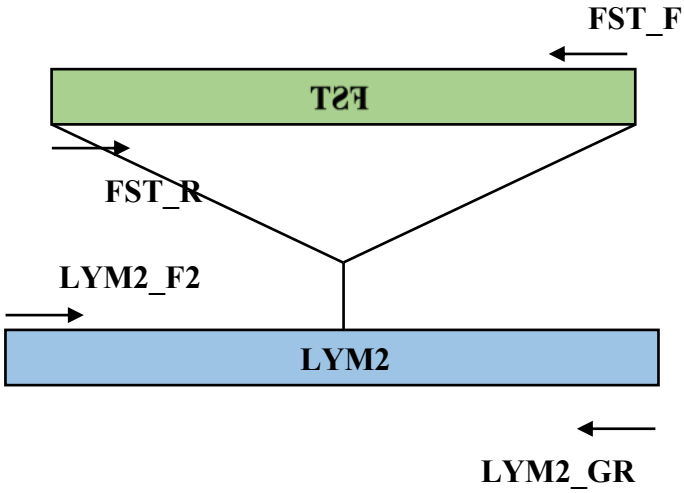

Figure 2.18: A representation of FST insertion into LYM2 in an inverted manner together with corresponding primers.

The band which would be in length of size amplified by LYM2_F2 and LYM2_GR primers corresponds to the WT genotype again whereas the region amplified by FST_R & LYM2_GR or LYM2_F2 & FST_F would be the indicator of mutant genotype in this case. The primers that were used include LYM2_F2, LYM2_GR, and FST_F. The following gel was obtained as a result of PCR reaction. The following gel was obtained as a result of PCR reaction.

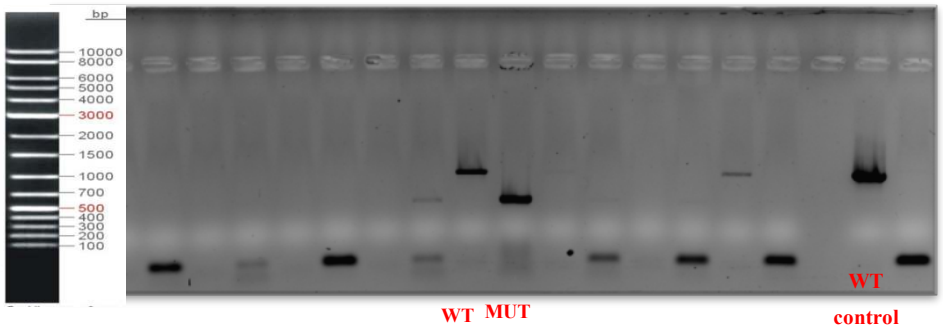

Figure 2.19: LYM2-8 wild-type band, mutant band from two different samples, and positive wild-type control and ladder by Biomatik 1kb plus ladder.

Overall, FST was found to be positioned in two different orientation in LYM1 and LYM2 seeds and mutant seeds were detected using the primers mentioned above to amplify the regions observed on the agarose gel.

2.4 Conclusion

LYM1 and LYM2 CDS were successfully cloned into pGWB12 plant expression vectors for downstream applications such as co-immunoprecipitation and further protein-protein interaction studies. It is aimed to identify interacting proteins of LYM proteins and their specific interaction with DNF2, which is thought to process LYM proteins. Moreover, potential homozygous mutants of LYM1 and LYM2 seeds were detected by performing genotyping utilizing PCR.

VITA

Duğçar Ebrar Erdoğan was born in Istanbul on 25.11.1997. He received his Bachelor's degrees from Chemistry and Molecular Biology & Genetics Departments at Istanbul Technical University in 2020, and 2021, respectively. In 2020, he started his master's studies in Molecular Biology and Genetics program at Koç University under the supervision of Asst. Prof. Onur Öztaş. He studied genome-wide mapping of nucleotide excision repair in Arabidopsis root and shoot by using computational methods in his first year, and then worked on LYM1/2 proteins during nodulation in *Medicago truncatula* by using experimental methods in his last year. He is currently planning to go Germany for his PhD studies.

BIBLIOGRAPHY

Adebali, O., Chiou, Y. Y., Hu, J., Sancar, A., & Selby, C. P. (2017). Genome-wide transcription-coupled repair in Escherichia coli is mediated by the Mfd translocase. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(11), E2116–E2125. https://doi.org/10.1073/pnas.1700230114

Bono, J. J., Fliegmann, J., Gough, C., & Cullimore, J. (2019). Expression and function of the Medicago truncatula lysin motif receptor-like kinase (Lysm-RLK) gene family in the legume–rhizobia symbiosis. *The Model Legume Medicago Truncatula*, 439–447. https://doi.org/10.1002/9781119409144.ch55

Canturk, F., Karaman, M., Selby, C. P., Kemp, M. G., Kulaksiz-Erkmen, G., Hu, J., Li, W., Lindsey-Boltz, L. A., & Sancar, A. (2016). Nucleotide excision repair by dual incisions in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4706–4710. https://doi.org/10.1073/pnas.1604097113

Chakraborty, S., Nguyen, B., Wasti, S. D., & Xu, G. (2019). Plant Leucine-Rich Repeat Receptor Kinase (LRR-RK): Structure, Ligand Perception, and Activation Mechanism. *Molecules 2019, Vol. 24, Page 3081*, *24*(17), 3081. https://doi.org/10.3390/MOLECULES24173081

Chang, C., Damiani, I., Puppo, A., & Frendo, P. (2009). Redox changes during the legume-rhizobium symbiosis. *Molecular Plant*, *2*(3), 370–377. https://doi.org/10.1093/MP/SSN090

Cheng, Y., Zhou, W., el Sheery, N. I., Peters, C., Li, M., Wang, X., & Huang, J. (2011). Characterization of the Arabidopsis glycerophosphodiester phosphodiesterase (GDPD) family reveals a role of the plastid-localized AtGDPD1 in maintaining cellular phosphate homeostasis under phosphate starvation. *Plant Journal*, *66*(5), 781–795. https://doi.org/10.1111/j.1365-313X.2011.04538.x

Cleaver, J. E., Lam, E. T., & Revet, I. (2009). Disorders of nucleotide excision repair: The genetic and molecular basis of heterogeneity. *Nature Reviews Genetics*, *10*(11), 756–768. https://doi.org/10.1038/nrg2663

Cui, M., Gu, M., Lu, Y., Zhang, Y., Chen, C., Ling, H. Q., & Wu, H. (2020). Glutamate synthase 1 is involved in iron-deficiency response and long-distance transportation in Arabidopsis. *Journal of Integrative Plant Biology*, *62*(12), 1925–1941. https://doi.org/10.1111/jipb.12985

Dai, G.-Y., Yin, J., Li, K.-E., Chen, D.-K., Liu, Z., Bi, F.-C., Rong, C., & Yao, N. (2020). The Arabidopsis AtGCD3 protein is a glucosylceramidase that preferentially hydrolyzes long-acyl-chain glucosylceramides. *Journal of Biological Chemistry*, *295*(3), 717–728. https://doi.org/10.1016/s0021-9258(17)49930-3

de Laat, W. L., Jaspers, N. G. J., & Hoeijmakers, J. H. J. (1999). Molecular mechanism of nucleotide excision repair. *Genes & Development* , *13*(7), 768–785. http://genesdev.cshlp.org/content/13/7/768.short

Deger, N., Yang, Y., Lindsey-Boltz, L. A., Sancar, A., & Selby, C. P. (2019). Drosophila, which lacks canonical transcription-coupled repair proteins, performs transcription-coupled repair. *Journal of Biological Chemistry*, *294*(48), 18092–18098. https://doi.org/10.1074/jbc.AC119.011448

diCenzo, G. C., Tesi, M., Pfau, T., Mengoni, A., & Fondi, M. (2020). Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nature Communications 2020 11:1*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-16484-2

Farkas, A., Maróti, G., Dürgo, H., Györgypál, Z., Lima, R. M., Medzihradszky, K. F., Kereszt, A., Mergaert, P., & Kondorosi, É. (2014). Medicago truncatula symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(14), 5183–5188. https://doi.org/10.1073/PNAS.1404169111/-/DCSUPPLEMENTAL/SAPP.PDF

Fliegmann, J., Canova, S., Lachaud, C., Uhlenbroich, S., Gasciolli, V., Pichereaux, C., Rossignol, M., Rosenberg, C., Cumener, M., Pitorre, D., Lefebvre, B., Gough, C., Samain, E., Sébastien, #, Fort, S., Driguez, H., Vauzeilles, B., Beau, J.-M., Nurisso, A., … Bono, J.-J. (2013). *Lipo-chitooligosaccharidic Symbiotic Signals Are Recognized by LysM Receptor-Like Kinase LYR3 in the Legume Medicago truncatula*. https://doi.org/10.1021/cb400369u

Fliegmann, J., Uhlenbroich, S., Shinya, T., Martinez, Y., Lefebvre, B., Shibuya, N., & Bono, J. J. (2011). Biochemical and phylogenetic analysis of CEBiP-like LysM domain-containing extracellular proteins in higher plants. *Plant Physiology and Biochemistry*, *49*(7), 709–720. https://doi.org/10.1016/J.PLAPHY.2011.04.004

Fousteri, M., & Mullenders, L. H. F. (2008). Transcription-coupled nucleotide excision repair in mammalian cells: Molecular mechanisms and biological effects. *Cell Research*, *18*(1), 73–84. https://doi.org/10.1038/cr.2008.6

Friedberg, E. C. (2001). How nucleotide excision repair protects against cancer. *Nature Reviews Cancer*, *1*(1), 22–33. https://doi.org/10.1038/35094000

Gaul, L., & Svejstrup, J. Q. (2021). Transcription-coupled repair and the transcriptional response to UV-Irradiation. *DNA Repair*, *107*(August), 103208. https://doi.org/10.1016/j.dnarep.2021.103208

Gautrat, P., Mortier, V., Laffont, C., Keyser, A. de, Fromentin, J., & Goormachtig, S. (2019). *Unraveling new molecular players involved in the autoregulation of nodulation in Medicago truncatula*. *70*(4), 1407–1417. https://doi.org/10.1093/jxb/ery465

Hanawalt, P. C., & Spivak, G. (2008). Transcription-coupled DNA repair: Two decades of progress and surprises. *Nature Reviews Molecular Cell Biology*, *9*(12), 958–970. https://doi.org/10.1038/nrm2549

Horváth, B., Yeun, L. H., Domonkos, Á., Halász, G., Gobbato, E., Ayaydin, F., Miró, K., Hirsch, S., Sun, J., Tadege, M., Ratet, P., Mysore, K. S., Ané, J. M., Oldroyd, G. E. D., & Kaló, P. (2011). Medicago truncatula IPD3 is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses. *Molecular Plant-Microbe Interactions*, *24*(11), 1345–1358. https://doi.org/10.1094/MPMI-01-11-0015/SUPPL_FILE/MPMI-01-11-0015E4.PDF

Hu, J., Adar, S., Selby, C. P., Lieb, J. D., & Sancar, A. (2015). Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes and Development*, *29*(9), 948–960. https://doi.org/10.1101/gad.261271.115

Hu, J., Li, W., Adebali, O., Yang, Y., Oztas, O., Selby, C. P., & Sancar, A. (2019). Genome-wide mapping of nucleotide excision repair with XR-seq. *Nature Protocols*, *14*(1), 248–282. https://doi.org/10.1038/s41596-018-0093-7

Hu, Z., Cools, T., & de Veylder, L. (2016). Mechanisms Used by Plants to Cope with DNA Damage. *Annual Review of Plant Biology*, *67*, 439–462. https://doi.org/10.1146/annurev-arplant-043015-111902

Huang, F., Zhu, Q. H., Zhu, A., Wu, X., Xie, L., Wu, X., Helliwell, C., Chaudhury, A., Finnegan, E. J., & Luo, M. (2017). Mutants in the imprinted PICKLE RELATED 2 gene suppress seed abortion of fertilization independent seed class mutants and paternal excess interploidy crosses in Arabidopsis. *Plant Journal*, *90*(2), 383–395. https://doi.org/10.1111/tpj.13500

Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T., & Parcy, F. (2002). bZIP transcription factors in Arabidopsis. *Trends in Plant Science*, *7*(3), 106–111. https://doi.org/10.1016/S1360-1385(01)02223-3

Jang, J. Y., Choi, C. H., & Hwang, D. J. (2010). The WRKY superfamily of rice transcription factors. *Plant Pathology Journal*, *26*(2), 110–114. https://doi.org/10.5423/PPJ.2010.26.2.110

Johnson, X., Wostrikoff, K., Finazzi, G., Kuras, R., Schwarz, C., Bujaldon, S., Nickelsen, J., Stern, D. B., Wollman, F. A., & Vallon, O. (2010). MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization of rbcl mrna in chlamydomonas and arabidopsis. *Plant Cell*, *22*(1), 234–248. https://doi.org/10.1105/tpc.109.066266

Kaya, S., Adebali, O., Oztas, O., & Sancar, A. (2022). Genome-wide Excision Repair Map of Cyclobutane Pyrimidine Dimers in Arabidopsis and the Roles of CSA1 and CSA2 Proteins in Transcription-coupled Repair†. *Photochemistry and Photobiology*, *98*(3), 707–712. https://doi.org/10.1111/php.13519

Kelly, S., Radutoiu, S., Stougaard, J., Chiou, T.-J., Versaw, W. K., & Fujiwara, T. (2017). Legume LysM receptors mediate symbiotic and pathogenic signalling. *Current Opinion in Plant Biology*, *39*, 152–158. https://doi.org/10.1016/J.PBI.2017.06.013

Kitashova, A., Schneider, K., Fürtauer, L., Schröder, L., Scheibenbogen, T., Fürtauer, S., & Nägele, T. (2021). Impaired chloroplast positioning affects photosynthetic capacity and regulation of the central carbohydrate metabolism during cold acclimation. *Photosynthesis Research*, *147*(1), 49–60. https://doi.org/10.1007/s11120-020-00795-y

Kubeš, M., Yang, H., Richter, G. L., Cheng, Y., Młodzińska, E., Wang, X., Blakeslee, J. J., Carraro, N., Petrášek, J., Zažímalová, E., Hoyerová, K., Peer, W. A., & Murphy, A. S. (2012). The Arabidopsis concentration-dependent influx/efflux transporter ABCB4 regulates cellular auxin levels in the root epidermis. *Plant Journal*, *69*(4), 640–654. https://doi.org/10.1111/j.1365-313X.2011.04818.x

Kubo, M., Furuta, K., Demura, T., Fukuda, H., Liu, Y. G., Shibata, D., & Kakimoto, T. (2011). The CKH1/EER4 gene encoding a TAF12-like protein negatively regulates cytokinin sensitivity in arabidopsis thaliana. *Plant and Cell Physiology*, *52*(4), 629–637. https://doi.org/10.1093/pcp/pcr021

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larkin, R. M., Stefano, G., Ruckle, M. E., Stavoe, A. K., Sinkler, C. A., Brandizzi, F., Malmstrom, C. M., Osteryoung, K. W., & Chory, J. (2016). REDUCED CHLOROPLAST COVERAGE genes from Arabidopsis thaliana help to establish the size of the chloroplast compartment. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(8), E1116–E1125. https://doi.org/10.1073/pnas.1515741113

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, L., Habring, A., Wang, K., & Weigel, D. (2020). Atypical Resistance Protein RPW8/HR Triggers Oligomerization of the NLR Immune Receptor RPP7 and Autoimmunity. *Cell Host and Microbe*, *27*(3), 405-417.e6. https://doi.org/10.1016/j.chom.2020.01.012

Li, W., Adebali, O., Yang, Y., Selby, C. P., & Sancar, A. (2018). Single-nucleotide resolution dynamic repair maps of UV damage in Saccharomyces cerevisiae genome. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), E3408–E3415. https://doi.org/10.1073/pnas.1801687115

Lin, J. H., Yu, L. H., & Xiang, C. bin. (2020). ARABIDOPSIS NITRATE REGULATED 1 acts as a negative modulator of seed germination by activating ABI3 expression. *New Phytologist*, *225*(2), 835–847. https://doi.org/10.1111/nph.16172

Lopez, L., Perrella, G., Calderini, O., Porceddu, A., & Panara, F. (2022). Genome-Wide Identification of Histone Modification Gene Families in the Model Legume Medicago truncatula and Their Expression Analysis in Nodules. *Plants*, *11*(3). https://doi.org/10.3390/plants11030322

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Luu, T.-B., Ourth, A., Pouzet, C., Pauly, N., & Cullimore, J. (2021). A newly-evolved chimeric lysin motif receptor-like kinase in Medicago truncatula spp. tricycla R108 extends its Rhizobia symbiotic partnership. *BioRxiv*, 2021.11.23.469708.

https://www.biorxiv.org/content/10.1101/2021.11.23.469708v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.11.23.469708v1.abstract

Mao, P., Smerdon, M. J., Roberts, S. A., & Wyrick, J. J. (2016). Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(32), 9057–9062. https://doi.org/10.1073/pnas.1606667113

Marteijn, J. A., Lans, H., Vermeulen, W., & Hoeijmakers, J. H. J. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature Reviews Molecular Cell Biology*, *15*(7), 465–481. https://doi.org/10.1038/nrm3822

Mendoza-Suárez, M., Andersen, S. U., Poole, P. S., & Sánchez-Cañizares, C. (2021). Competition, Nodule Occupancy, and Persistence of Inoculant Strains: Key Factors in the Rhizobium-Legume Symbioses. *Frontiers in Plant Science*, *12*, 1684. https://doi.org/10.3389/FPLS.2021.690567/BIBTEX

Mus, F., Crook, M. B., Garcia, K., Costas, A. G., Geddes, B. A., Kouri, E. D., Paramasivan, P., Ryu, M. H., Oldroyd, G. E. D., Poole, P. S., Udvardi, M. K., Voigt, C. A., Ané, J. M., & Peters, J. W. (2016). Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Applied and Environmental Microbiology*, *82*(13), 3698–3710. https://doi.org/10.1128/AEM.01055-16

Oztas, O. (2017). *the Key Question in Symbiotic Nitrogen Fixation: How Does Host Maintain a Bacterial Symbiont? Doctoral Dissertations. 945.* https://scholarworks.umass.edu/dissertations_2/945

Oztas, O., Selby, C. P., Sancar, A., & Adebali, O. (2018). Genome-wide excision repair in Arabidopsis is coupled to transcription and reflects circadian gene expression patterns. *Nature Communications*, *9*(1), 1–8. https://doi.org/10.1038/s41467-018-03922-5

Pani, B., & Nudler, E. (2017). Mechanistic insights into transcription coupled DNA repair. *DNA Repair*, *56*(June), 42–50. https://doi.org/10.1016/j.dnarep.2017.06.006

Pankievicz, V. C. S., Irving, T. B., Maia, L. G. S., & Ané, J. M. (2019). Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops. *BMC Biology*, *17*(1), 1–17. https://doi.org/10.1186/s12915-019-0710-0

Park, E., & Nebenführ, A. (2013). Myosin XIK of Arabidopsis thaliana Accumulates at the Root Hair Tip and Is Required for Fast Root Hair Growth. *PLoS ONE*, *8*(10), 1–18. https://doi.org/10.1371/journal.pone.0076745

Petit, C., & Sancar, A. (1999). Nucleotide excision repair: From E. coli to man. *Biochimie*, *81*(1–2), 15–25. https://doi.org/10.1016/S0300-9084(99)80034-0

Quinlan, A. R. (2014). BEDTools: The Swiss-Army tool for genome feature analysis. In *Current Protocols in Bioinformatics* (Vol. 2014). https://doi.org/10.1002/0471250953.bi1112s47

Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., & Stougaard, J. (2003). *Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases*. www.nature.com/nature

Radutoiu, S., Madsen, L. H., Madsen, E. B., Jurkiewicz, A., Fukai, E., Quistgaard, E. M. H., Albrektsen, A. S., James, E. K., Thirup, S., & Stougaard, J. (2007). LysM domains mediate lipochitin–oligosaccharide recognition and Nfr genes extend the symbiotic host range. *The EMBO Journal*, *26*(17), 3923–3935. https://doi.org/10.1038/SJ.EMBOJ.7601826

Reardon, J. T., & Sancar, A. (2005). Nucleotide Excision Repair. *Progress in Nucleic Acid Research and Molecular Biology*, *79*(04), 183–235. https://doi.org/10.1016/S0079-6603(04)79004-2

Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., Frugoli, J., Dickstein, R., & Udvardi, M. K. (2020). Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *The Plant Cell*, *32*(1), 15–41. https://doi.org/10.1105/tpc.19.00279

Rutten, L., Miyata, K., Roswanjaya, Y. P., Huisman, R., Bu, F., Hartog, M., Linders, S., van Velzen, R., van Zeijl, A., Bisseling, T., Kohlen, W., & Geurts, R. (2020). Duplication of symbiotic lysin motif receptors predates the evolution of nitrogen-fixing nodule symbiosis. *Plant Physiology*, *184*(2), 1004–1023. https://doi.org/10.1104/pp.19.01420

Růžička, K., Strader, L. C., Bailly, A., Yang, H., Blakeslee, J., Łangowski, Ł., Nejedlá, E., Fujita, H., Itoh, H., Syono, K., Hejátko, J., Gray, W. M., Martinoia, E., Geisler, M., Bartel, B., Murphy, A. S., & Friml, J. (2010). Arabidopsis PIS1 encodes the ABCG37 transporter of auxinic compounds including the auxin precursor indole-3-butyric acid. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(23), 10749–10753. https://doi.org/10.1073/pnas.1005878107

Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A., & Lopez-Bigas, N. (2016). Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature*, *532*(7598), 264–267. https://doi.org/10.1038/nature17661

Sancar, A. (1993). Photobiology school nucleotide excision. *Society*, *57*(5), 905–921.

Sancar, A., Lindsey-Boltz, L. A., Kang, T. H., Reardon, J. T., Lee, J. H., & Ozturk, N. (2010). Circadian clock control of the cellular response to DNA damage. *FEBS Letters*, *584*(12), 2618–2625. https://doi.org/10.1016/j.febslet.2010.03.017

Sancar, A., Lindsey-Boltz, L. A., Ünsal-Kaçmaz, K., & Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual Review of Biochemistry*, *73*, 39–85. https://doi.org/10.1146/annurev.biochem.73.011303.073723

Sasaki, T., Fukuda, H., & Oda, Y. (2017). CORTICAL MICROTUBULE DISORDERING1 is required for secondary cell wall patterning in xylem vessels. *Plant Cell*, *29*(12), 3123–3139. https://doi.org/10.1105/tpc.17.00663

Schärer, O. D. (2013). Nucleotide excision repair in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *5*(10), 1–19. https://doi.org/10.1101/cshperspect.a012609

Staal, J., Kaliff, M., Dewaele, E., Persson, M., & Dixelius, C. (2008). RLM3, a TIR domain encoding gene involved in broad-range immunity of Arabidopsis to necrotrophic fungal pathogens. *Plant Journal*, *55*(2), 188–200. https://doi.org/10.1111/j.1365-313X.2008.03503.x

Suo, Y., Hu, F., Zhu, H., Li, D., Qi, R., Huang, J., & Wu, W. (2021). Big3 and big5 redundantly mediate vesicle trafficking in arabidopsis. *Biomolecules*, *11*(5), 1–12. https://doi.org/10.3390/biom11050732

Swiderski, M. R., Birker, D., & Jones, J. D. G. (2009). The TIR domain of TIR-NB-LRR resistance proteins is a signaling domain involved in cell death induction. *Molecular Plant-Microbe Interactions*, *22*(2), 157–165. https://doi.org/10.1094/MPMI-22-2-0157

Wozniak, K. J., & Simmons, L. A. (2022). Bacterial DNA excision repair pathways. *Nature Reviews Microbiology*, *0123456789*. https://doi.org/10.1038/s41579-022-00694-0

Wrzaczek, M., Brosché, M., Salojärvi, J., Kangasjärvi, S., Idänheimo, N., Mersmann, S., Robatzek, S., Karpiński, S., Karpińska, B., & Kangasjärvi, J. (2010). Transcriptional regulation of the CRK/DUF26 group of Receptor-like protein

kinases by ozone and plant hormones in Arabidopsis. *BMC Plant Biology*, *10*, 1–19. https://doi.org/10.1186/1471-2229-10-95

Wu, S., Huang, Y., Selby, C. P., Gao, M., Sancar, A., & Hu, J. (2022). A new technique for genome-wide mapping of nucleotide excision repair without immunopurification of damaged DNA. *Journal of Biological Chemistry*, *298*(5), 1–13. https://doi.org/10.1016/j.jbc.2022.101863

Zhao, Z., Yang, X., Lü, S., Fan, J., Opiyo, S., Yang, P., Mangold, J., Mackey, D., & Xia, Y. (2020). Deciphering the novel role of atmin7 in cuticle formation and defense against the bacterial pathogen infection. *International Journal of Molecular Sciences*, *21*(15), 1–21. https://doi.org/10.3390/ijms21155547

Zipfel, C., & Oldroyd, G. E. D. (2017). Plant signalling in symbiosis and immunity. *Nature*, *543*(7645), 328–336. https://doi.org/10.1038/nature22009

APPENDIX A

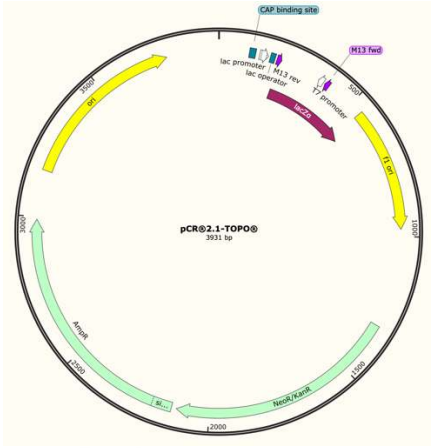

Figure A.1: Plasmid map of pCR2.1-TOPO®

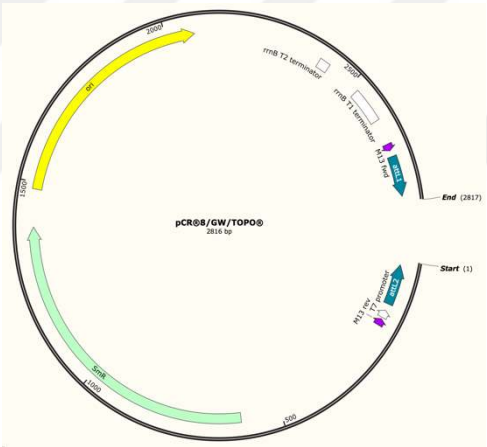

Figure A.2: Plasmid map of pCR8® GW TOPO vector

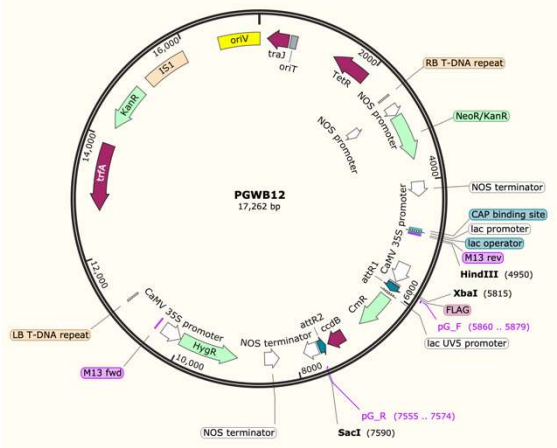

Figure A.3: Plasmid map of pGWB12 plant expression vector.