

GENETIC ANALYSIS OF *XIMENIA AMERICANA* UNDER NATURAL  
CONDITIONS

A THESIS SUBMITTED IN PARTIAL FULFILMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE  
OF  
THE UNIVERSITY OF NAMIBIA  
BY

ANNASTASIA EKANDJO  
(200638599)

January 2015

Main Supervisor: Dr. Eroid Naomab

Co- Supervisor: Dr. Gladys Kahaka

**Abstract**

Cross-species hybridization approaches have been used for genetic analysis of species that do not have complete genome arrays available. The study used the cross-species DNA-DNA hybridization approach to characterize *Ximenia americana* biological processes. Preliminary micro-array studies done in the Department of Chemistry and Biochemistry at the University of Namibia has created *X. americana* DNA-DNA hybridization to a well-known *Arabidopsis thaliana* genechip (ATH1). *A. thaliana* probe-pairs that hybridised to the *X. americana* genomic DNA on the basis of the perfect-match (PM) probe signal were selected and analysed using a cel file parser script to generate a new high density probe mask files. These files effectively represent the first ever *X. americana* DNA-DNA hybridisation data. This study conducted gene ontology analysis of *X. americana*/*A. thaliana* hybridisation data. Such gene ontology analysis demonstrated that abiotic stress response genes are over-represented in relative comparison to model species *A. thaliana* under natural conditions. This observation was independently confirmed with PCR amplification of the following orthologous genes using *X. americana* genomic DNA: AT4G15910.1, SAD2, HXK1, ACC and ERF/AP2. Given the lack of genomic sequence information in *X. americana* background, primers for genomic amplification was design using *A. thaliana* genomic sequence information. Primers where designed to yield 100 bp genomic PCR product. Each of the selected genes was successfully amplified hence giving evidence of homology within primer binding sites. Nonetheless, the genomic amplification of these key abiotic factors in *X. americana* confirms the type of responses that supports the adaptation of *X. americana*

under natural conditions that are stress related (heat and drought stress). Although transcript levels of this important abiotic response factors could not be measured in absolute or relative terms, the study demonstrated inherent presence of such genes at the genomic level of *X. americana*. Further studies will be required to demonstrate that some sort of modification of any of these or other abiotic response factors within the genome of *X. americana* provides the key basis for its adaptation to the relatively dry and hot climatic conditions.

**Key words:** cross-species hybridization, gene ontology, abiotic response factors, micro-array and ATH1

## Table of Contents

Abstract .....	ii
List of Figures .....	vii
List of Tables .....	viii
Acronyms .....	ix
Acknowledgement.....	x
Dedication .....	xii
Declaration .....	xiii
CHAPTER 1: INTRODUCTION .....	1
1.1.    General introduction .....	1
1.2.    Statement of the problem.....	2
1.3.    Objectives of the research.....	3
1.4.    Significance of the study.....	3
CHAPTER 2: LITERATURE REVIEW .....	5
2.1.    Genetic Analysis .....	5
2.1.1.    Genetic information .....	5
2.1.2    Abiotic Stress Response in Olacaceae .....	5
2.2.    Tools and Application for Genetic Analysis.....	9
2.2.1.    Microarray .....	9

2.2.2	DNA-DNA hybridization cross species Genetic Analysis.....	13
2.2.3	Gene Ontology (GO).....	13
2.2.4	Polymerase Chain Reaction (PCR).....	16
CHAPTER 3: METHODOLOGY .....		18
3.	Materials and Methods.....	18
3.1	Sample collection.....	18
3.2	DNA extraction.....	18
3.3	DNA quantification: analysis of genomic DNA using ND-Nanodrop 2000c .....	19
3.4	Hybridization .....	19
3.5	Primer design .....	21
3.6	Gene ontology classification.....	21
3.6.1	Gene Ontology Analysis of <i>A.thaliana</i> Genome .....	22
3.6.2	Gene Ontology Analysis of <i>A.thaliana</i> / <i>X. Americana</i> Hybridization Data .....	22
3.6.3	Gene Ontology Analysis of <i>X.americana</i> PCR confirmed Loci.....	22
3.7	Independent confirmation using Standard PCR.....	23
3.7.1	Standard PCR amplification reactions .....	23
3.7.2	PCR products gel electrophoresis.....	25
CHAPTER 4: RESULTS AND DISCUSSION.....		27
4.2	Gene Ontology Analysis of <i>A.thaliana</i> genome Annotations .....	30

4.3	Gene Ontology Analysis of <i>A.thaliana</i> / <i>X. Americana</i> Hybridization Data.....	31
4.4	Comparative Gene Ontology Analyses.....	32
4.5	Independent confirmation.....	33
4.5.1	DNA Quality analysis and quantification.....	33
4.5.2	Independent confirmation of <i>X.americana</i> Biological process by Standard PCR.....	36
CHAPTER 5: CONCLUSION.....		42
5	Conclusion.....	42
REFERENCES.....		44
APPENDIX.....		54

**List of Figures**

Figure 1: <i>X.americana</i> plant, a) its fruits and b) flowers (retrieved from:.....	2
Figure 2: the Affymetrix expression array .....	12
Figure 3: Polymerase Chain Reaction parameters. ....	17
Figure 4: the graph of <i>X.americana</i> DNA concentration (139ng/ $\mu$ l).....	28
Figure 5: Comparison of Gene Ontology Classified Biological Processes between <i>A. thaliana</i> and <i>X. americana</i> .....	32
Figure 6: the Amplifications of <i>X.americana</i> DNA with different primers on the 1% Agarose gel.....	35
Figure 7:the genomic PCR based independent confirmation of <i>X.americana</i> biological process.....	37

**List of Tables**

Table 1: Reaction Master Mix for 25 $\mu$ l .....	23
Table 2: the primers IDs and the gene IDs for each primer .....	24
Table 3: Standard PCR reaction parameters .....	25
Table 4: the concentration of <i>X.americana</i> DNA samples .....	27
Table 5: the genomic DNA hybridization intensity values.....	29
Table 6: the GO biological process in <i>A.thaliana</i> genome .....	30
Table 7 : the DNA-DNA Hybridization GO Annotations .....	31
Table 8: the DNA concentration using ND- NanoDrop 2000c .....	34
Table 9: the ten targeted orthologous genes for independent confirmation.....	54

**Acronyms**

<b>DNA</b>	Deoxyribose Nucleic Acid
<b>PCR</b>	Polymerase Chain Reaction
<b>GO</b>	Gene Ontology
<b>dNTPs</b>	Deoxynucleotide triphosphates
<b>ACC</b>	1-aminocyclopropane-1-carboxylate
<b>HXK1</b>	Hexokinase 1
<b>ERF</b>	Ethylene Response Factor
<b>ATH1</b>	Thaliana Homeobox 1
<b>ABA</b>	Abscisic acid
<b>PM</b>	Perfect Match
<b>SAD2</b>	Sensitive to ABA and Drought2
<b>AP2</b>	APETALA2
<b>UV</b>	Ultra Violet
<b>SAM</b>	S-Adenosyl methionine
<b>LEA</b>	Late Embryogenesis Abundant
<b>cDNA</b>	complementary DNA

<b>DDH</b>	DNA-DNA Hybridization
<b>TAIR</b>	The Arabidopsis Information Resources
<b>T<sub>m</sub></b>	Melting Temperature
<b>NASC</b>	National Animal Supplement Council
<b>TBE</b>	Tris/Borate/ Ethylenediamine Tetraacetic Acid
<b>HsP90</b>	heat shock protein 90

**Acknowledgement**

Firstly, I would like to thank the man above, without him I would not accomplished all this and secondly I would like to thank my project supervisor, Dr. E Naomab for his support and effective style of supervision throughout the year. Your motivation and encouragement added to my strength to complete this project success fully, I forever remain grateful to you.

I would also like to extend my sincere gratitude to my co-supervisor, Dr. G Kahaka, for her support to keep me going with the project, you are such a blessing. Let me also not forget Dr E. Fabiano for his valuable support and the department of Chemistry and Biochemistry for allowing me to do my project. Friends and family, thank you.

## **Dedication**

I would like to dedicate this work to my late Dad Peter Ekandjo. My mom Lina Shindingeni thank you for all the love, support and guidance that helped moulds me into who I am today. My siblings, Miriam and Elizabeth, this is for you!!

**Declaration**

I, Annastasia Ekandjo, hereby declare that this study is a true reflection of my own research, and that this work, or part thereof has not been submitted for a degree in any other institution of higher education.

No part of this thesis may be reproduced, stored in any retrieval system, or transmitted in any form, or by means (e.g. electronic, mechanical, photocopying, recording or otherwise) without the prior permission of the author, or The University of Namibia in that behalf.

I, Annastasia Ekandjo, grant The University of Namibia the right to reproduce this thesis in whole or in part, in any manner or format, which The University of Namibia may deem fit, for any person or institution requiring it for study or research; providing that The University of Namibia shall waive this right if the whole thesis has been or is being published in a manner satisfactory to the University.

.....

Date .....

## CHAPTER 1: INTRODUCTION

### 1.1. General introduction

*Ximenia* although primarily indigenous to Africa is also found in other parts of the world, such as South America. This plant belongs to *Olacaceae* family and it is widely distributed in Africa. The most common species are *X. americana* and *X. caffra* which are found in the northern part of Namibia. This study focused on *X. Americana*. The *X. americana* grows up to 6 meters tall and has rough dark-grey bark. It has hairless green leaves and thorny branches. Regardless of climate change, *X. americana* bears fruits and flowers throughout the year (Orwa et al., 2009). *X. americana* is a drought-resistant plant, it mostly grows at low altitudes in a wide range of environment including, savannahs, dry woodlands, dry forests, and along coastal areas or on river banks (Orwa et al., 2009).

This specific plant is of economic importance because of its wood yield. *X. americana* fruit which is often used for jam, jellies and also kernel is reported to contain high oil quantity (Maikai, Maikai, & Kobo, 2009). In addition to its economic value, *X. americana*'s fruits are used for medicinal purposes which includes treating headaches, skin problems, sexually transmitted diseases, heal wounds as topical ointment, and relieve cough. *X. americana*'s used for medicinal purposes is reinforced by its free radical scavenging and antioxidant activity (Maikai et al., 2009; Chhabra & Viso, 2009; James, Owolabi, Ibiyeye, Magaji, & Ikugiyi, 2008).



**Figure 1: *X.americana* plant, a) its fruits and b) flowers (retrieved from: RonYeo@tidechaser.blogspot.com).**

### **1.2. Statement of the problem**

The socio-economic and ecological value of *X. americana* remains understated and under-valued. This is because the genome of this species has not yet been sequenced, which is a necessary precondition for bio-mining and bio-processing. Moreover, sequencing an organism's genome help with the understanding on how that organism interacts with the environment where it is found. For example the influence of heat and water stress on gene expression. In the absence of *X. americana* sequence information,

this study focused to evaluate the distribution of its biological processes using *A. thaliana* as hybridization platform. This approach has been applied to study biological processes of many un-sequenced plant and animal species. (Graham et al, 2010;Cardwell, Cardwell, Johnson, & Stevens, 2013; Gao, Yang, & Quiros, 2003).

### **1.3. Objectives of the research**

The specific objectives of this study are;

- ❖ To analyze and compare *X. americana* and *A. thaliana* DNA-DNA hybridization data using gene ontology
- ❖ To independently confirm ontology result with PCR amplification of orthologous genes within a given biological process. This requires designing primers in *A.thaliana* background and test the presence of genomic fragments in *X. americana* genomic DNA.

### **1.4. Significance of the study**

Preliminary studies done in the department of Chemistry and Biochemistry has created *X. americana* DNA-DNA hybridization to a well-known *A. thaliana* genechip (ATH1). This study analyzed these results and has demonstrated the distribution of *X. americana* biological processes under natural conditions on the basis of cross-species hybridization. Moreover, this study provided first insights into how *X. americana* is adapted to the harsh environmental conditions (i.e. drought and heat stress) in comparison to *A. thaliana*. However, further studies will be required to understand specific regulation of

*X. americana* genes under natural conditions as a way to appreciate why this particular species are considered to be drought tolerant.

## **CHAPTER 2: LITERATURE REVIEW**

### **2.1. Genetic Analysis**

#### **2.1.1. Genetic information**

Nucleic acids are universal components of all living matter and are needed for storage, transmission, and transfer of genetic information (Alberts et al, 2008). A nucleic acid is either a single- or double-stranded polynucleotide chain. Furthermore, it can take the form of either a deoxyribonucleic acid (DNA) or a ribonucleic acid (RNA) (Alberts et al, 2008; Strachan & Read, 2004).

The most important concept in molecular biology is the way information stored in DNA is passed on and expressed. It is sometimes referred to as the flow of genetic information. The transfer of information from a gene to a protein is a two-step process namely the transcription and translation (Taft, Pang, Mercer, Dinger, & Mattick , 2010; Taiz & Zeiger, 2010).

#### **2.1.2 Abiotic Stress Response in Olacaceae**

Important to the survival of every cell is to monitor the environment and to respond to external stimuli. For most cells this include appropriate communication with neighboring cells known as cell signaling (Krauss, 2008)

Drought and heat resistant plants such as *X. americana* have evolved mechanisms to adapt to the physiological stressful effects caused by water deficit. Two main responses include the mechanisms aimed at maintenance of water balance thus avoiding dehydration and through the alteration of gene expression (Abdallah, Moses, & Prakash,

2014; Savitri, Basuki, Aini, & Arumingtyas, 2013). Water maintenance seemed to be achieved mainly through the synthesis of sugar alcohols, amino acid proline and glycine betaine, which do not interfere with enzymatic reactions (Abdallah et al., 2014). A second form of response is through the alteration of gene expression either through suppression or enhancement of drought tolerance (Savitri et al., 2013). An example of a protein resultant from drought inducing genes is the late embryogenesis abundant which is part of the dehydrin family (*LEA-D11*) (Hanin et al., 2011). LEA tend to accumulate during the late embryogenesis (embryo maturation and desiccation) or as a result of stress stimuli such as drought, heat, which may cause cell dehydration (Hanin et al., 2011). Thus, LEA has a protective and stabilizing function towards cell membrane structures and to macromolecules (Savitri et al., 2013). Another example of stress induced protein is the heat shock protein (e.g. Hsp90, Hsp70) which are expressed after a heat threshold has been reached (Y. Liang et al., 2011b; Usman et al., 2014). High temperatures have both physiological and morphological effects (e.g. plant metabolism, structure) as well as molecular (e.g. membranes and proteins) (Y. Liang et al., 2011). To minimize these effects Hsp introduce a number of modifications at the cellular levels that include transient changes in gene expression and structural and function level such as close of stomata (Usman et al., 2014).

Additionally, drought stress considered to cause a wide range of physiological and biochemical responses in plants. These responses include reduced stomatal conductance and photosynthesis, and accumulation of osmolytes and proteins in cells (Wang et al.,

2011). A number of genes have been identified to be involved in drought response and tolerance, such as SAD2, HXK1, ACC and ERF/AP2.

SAD2 (super sensitive to ABA and drought 2) encodes an importin beta-domain family protein likely to be involved in nuclear transport in ABA signaling. Loss of function mutations in SAD2 exhibit increased tolerance for UV stress, increased production of UV protective secondary metabolites and suppression of nuclear localization of MYB4 (a repressor of UV stress response genes) (Verslues et al., 2006; Zhao et al., 2007). While a hexokinase (HXK1) in the plant glucose-signaling network functions as a glucose sensor to interrelate nutrient, light, and hormone signaling networks for controlling growth and development in response to the changing environment (Cho, Yoo, & Sheen, 2006; Karve, Xia, & Moore, 2012; Moore et al., 2003). Whereas 1-aminocyclopropane-1-carboxylate (ACC) synthase, catalyzes the conversion of 5-adenosylmethionine (SAM) to ACC, but ACC oxidase catalyzes the oxidation of ACC to ethylene. This plant hormone ethylene is produced in response to various kinds of environmental stress, such as wounding, physical load, disease, exposure to low temperature and chemicals and water stress (Karve et al., 2012; Kato et al., 2000; Li et al., 2012; X. Liang et al., 1992). However, ERF (ethylene response factor)/AP2 (The APETALA2) these are proteins that have important functions in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various responses to environmental Stimuli (Jung et al., 2010; Licausi et al., 2013; Nakano et al., 2006).

Moreover, abscisic acid (ABA) is a hormone that involved in many developmental plant processes, including leaf abscission and responding to environmental stress. This hormone is produced in the roots of the plant as well as the terminal buds at the top of the plant (Hartung, Sauter, & Hose, 2002). ABA is produced in the roots in response to decreased soil water potential and other conditions in which the plant may be under stress (Lenka, Lohia, Kumar, Chinnusamy, & Bansal, 2009). This ABA then translocates to the leaves where it quickly changes the osmotic potential of stomatal guard cells, triggering the leaves to shrink and the stomata to close. The closure of stomata reduces transpiration, hence there is no water loss from the leaves during water shortage (Hartung et al., 2002; Tuteja, 2007; Zalejski et al., 2006).

Furthermore, ABA regulates the adaptation process into two interacting steps. Firstly, ABA acts thru differential signal transduction pathways on cells which are the least and most affected by the imposed stress. Secondly, ABA may regulate through some genes/gene products, which control the expression of stress or adaptative-specific genes. Some genes are up-regulated and others are down-regulated, resulting in overall synthesis of genomic products which may play a role in plant survival under different environmental conditions (Iuchi et al., 2001; Lenka et al., 2009; Söderman, Brocard, Lynch, & Finkelstein, 2000; Water, Water, Kong, & Kong, 2000).

## **2.2. Tools for Genetic Analysis**

### **2.2.1. Microarray**

A microarray is a glass slide, onto which single-stranded DNA molecules are attached at fixed locations (spots) (Brazma & Vilo, 2000). In the microarray, the samples are arranged orderly where matching of known and unknown DNA samples is done based on base pairing rules. The experiment of microarray makes use of common assay systems such as microplates or standard blotting membranes. The sample spot sizes are typically less than 200 microns in diameter usually contain thousands of spots(Brazma & Vilo, 2000; Eisen & Brown, n.d.; Baldi & Hatfield, 2002).

Thousands of spotted samples known as probes (with known identity) are immobilized on a solid support (a microscope glass slides or silicon chips or nylon membrane). The spots can be DNA, cDNA, or oligonucleotides. These are used to determine complementary binding of the unknown sequences thus allowing parallel analysis for gene expression and gene discovery (Warner & Dieckgraefe, 2002).An experiment with a single DNA chip can provide information on thousands of genes simultaneously. An orderly arrangement of the probes on the support is important as the location of each spot on the array is used for the identification of a gene (Warner & Dieckgraefe, 2002; Baldi & Hatfield, 2002).

A number of microarray types are currently used including the glass DNA microarray which involves the micro spotting of pre-fabricated cDNA fragments on a glass slide

and the high-density oligonucleotide microarrays often referred to as a "chip" which involves *in situ* oligonucleotide synthesis (Warner & Dieckgraefe, 2002). *In situ* (on chip) oligonucleotide array format is a sophisticated platform of microarray technology which is manufactured by using the technology of *in situ* chemical synthesis that was first developed by Fodor et al., (1991). Nonetheless, the industry leader in the field of *in situ* oligonucleotide microarrays (Affymetrix) has further founded this type of technology in production of the so-called GeneChips which refers to its high density oligonucleotide based DNA arrays (Lockhart & Winzeler, 2000).

The basic principle of manufacturing Affymetrix's GeneChips is the use of photolithography and combinatorial chemistry to manufacture short single strands of DNA onto 5-inch square quartz wafers. Dissimilar to glass cDNA, the genes on the chip are designed based on sequence information alone, and then using an industry chip synthesizer, sequences are directly synthesized onto the surface of the 5-inch square quartz wafer at a pre-selected positions (Lockhart & Winzeler, 2000; Celis et al., 2000). GeneChips main advantages include speed, specificity and reproducibility. Speed, in terms of generating the array is prime advantage because, spotting the DNA onto the chip requires only that the DNA sequence of interest be known, therefore no time is spent in the handling of cDNA resources such as the preparation and accurate determination of handling bacterial clones, PCR products, or cDNAs, thus reducing the likelihood of contamination and mix up (Warner & Dieckgraefe, 2002).

Additional, advantages are a high specificity and reproducibility. Both of these attributes are due to the way oligonucleotide sequences printed on the chip are designed and the use of multiple, short sequence(s) representing the unique sequence of genes (Warner & Dieckgraefe, 2002; Celis et al., 2000). For example, when designing oligonucleotide sequences for a gene, each sequence is designed to be perfectly complementary to a target gene sequence, at the same time an additional partner sequence is designed that is identical except for a single base mismatch in its centre. This sequence mismatch strategy, along with the use of multiple sequence(s) for each gene increases specificity and helps to identify and minimize the effects of non-specific hybridization and background signal. This strategy also allows for the direct subtraction of cross-hybridization signals and discrimination between real and non-specific signals (Warner & Dieckgraefe, 2002; Celis et al., 2000). For these reasons and despite some disadvantages, GeneChips is the method of choice in this study.

Disadvantages of the *in situ* oligonucleotide array format include practical limitations in terms of affordability and flexibility. Firstly, *in situ* oligonucleotide array formats tend to have expensive specialized equipments e.g. to carry out the hybridization, staining of label, washing, and quantitation process. Secondly, ready made *in situ* oligonucleotide array format (GeneChips) are still expensive, although there has been reductions in cost as the market of microarrays has expanded. Thirdly, although short-sequences used on the array confer high specificity, they may have decreased sensitivity/binding compared with glass cDNA microarrays. Such low sensitivity however is compensated for by

using multiple probes (Freeman, Robertson & Vrana, 2000; Majtan & Timko, 2004, Cancer, 2002;Christensen, Angen, Mutters, Olsen & Bisgaard, 2000).

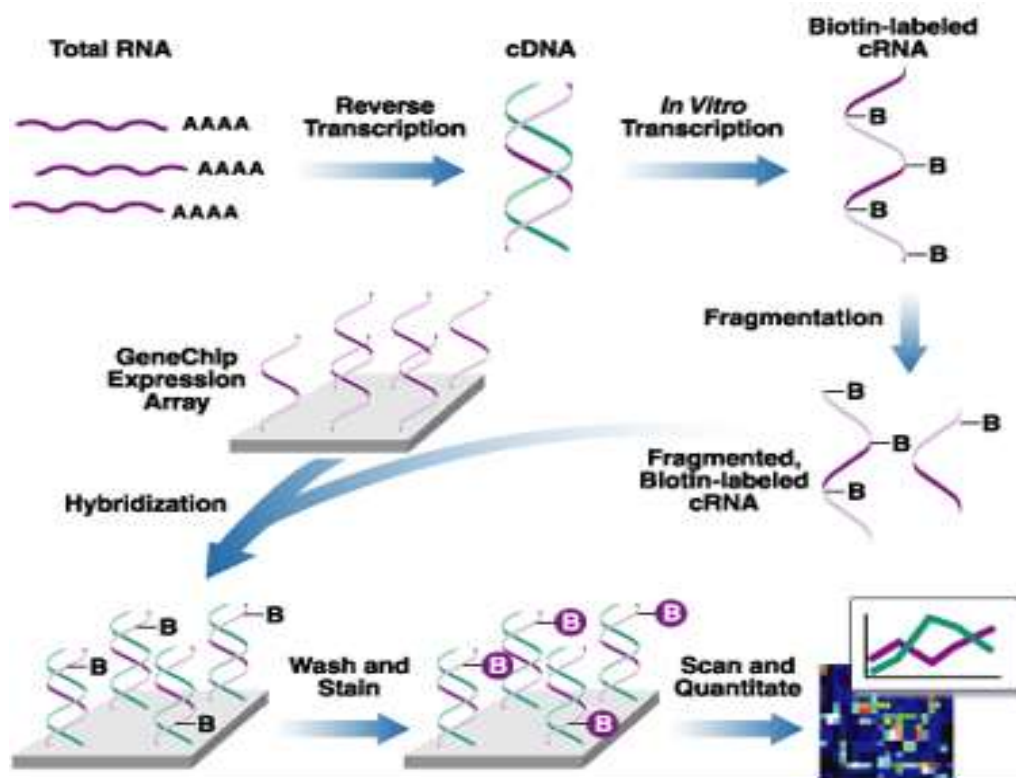


Figure 2: shows the Affymetrix expression array (Retrieved from:

[http://www.affymetrix.com/technology/ge\\_analysis/index.affx](http://www.affymetrix.com/technology/ge_analysis/index.affx)

Altogether, these microarray types (technology) have been widely applied. Application include gene discovery (i.e. identification of new genes, investigating gene functioning and expression levels under different conditions (van Hal et al., 2000)), disease diagnosis (i.e. assists with investigating different diseases such as heart diseases, mental illness, infectious disease and especially the study of cancer (Majtan & Timko, 2004)),

drug discovery (i.e. pharmacogenomics, that is the study of correlations between therapeutic responses to drugs and the genetic profiles of the patients (Majtan & Timko, 2004; van Hal et al, 2000)), and toxicological and toxicogenomics research (i.e. provides a robust platform for the research of the impact of toxins on the cells and their passing on to the progeny and an assessment of the correlation between responses to toxicants and the changes in the genetic profiles of the cells exposed to such toxicants, respectively; (Majtan & Timko, 2004 ; Warner & Dieckgraefe, 2002).

### **2.2.2 DNA-DNA hybridization cross species Genetic Analysis**

DNA-DNA hybridization (DDH) is one of the molecular cytogenetic tools used to gain a further understanding of the evolution, genetics and karyotypic alternations in various organisms (Silva & Souza, 2013). The technique can be used for studying genome wide comparisons among organisms conditioned on high quality DNA (Goris et al., 2007). The technique has four main steps that include the isolation of the DNA of the organism of interest as well as that of the reference organism, the heating of the mixture of DNA fragments to dissociate the DNA double-strands followed by the formation of cross species DNA-DNA hybrids. The latter, reannealing of fragments, is resultant from a decrease in temperature (Auch, von Jan, Klenk, & Göker, 2010; Silva & Souza, 2013). The degree of similarity or dissimilarity between strands (e.g. homologous ) determines their binding strength (Auch et al., 2010). As a consequence, the melting temperature has been used for infer degree of (dis)similarity among organisms. A 70% DDH value has been recommended as the baseline to delineate species, that is, the unit of study is

considered to be a different species from the reference one, if the DDH value is equal or more than 70% (Wayne et al., 1987).

The technique has been shown to be effective to differentiate genomes with homology of up to 95% (Parokonny et al., 1997). However, DDH has a few disadvantages that include the basic requirement for high quality DNA (Goris et al., 2007) which may prove to be difficult at times, it can be time consuming and labour intensive, different methods can yield different results and that libraries cannot be built given that the method relies on hybrid (Goris et al., 2007).

### **2.2.3 Gene Ontology (GO)**

Gene ontology (GO) is a utility which assigns functional roles to genes using a standardized control vocabulary (Berardini et al., 2004). Ontologies or key words are used to put genes in their respective biological functions. This tool (GO) is important for the representation and processing of related information on gene products across all species. When the gene is identified, it is important to know where the gene is allocated in the cell, what functions does the gene have on the molecular level and to what biological process do these functions contribute. Hence Gene Ontology is made of three smaller ontologies or aspects: Molecular Function, Biological Process, and Cellular Component (Ashburner et al., 2000; Bentink, 2003; Rhee, Wood, Dolinski, & Draghici, 2008; Smith & Kumar, 2004).

Smith & Kumar, (2004) defined Molecular function Ontology as the biochemical activity which including specific binding to ligands or structures of a gene product at a

molecular level. This aspect describes only what is done without specifying where or when the event actually occurs. Examples of broad molecular functional terms are catalytic activity, transporter activity or binding while Examples of narrower molecular functional terms are adenylate cyclase activity or Toll receptor binding (Ashburner et al., 2000).

Biological process refers to a biological objective to which the gene or gene product contributes. A process is of a series of events accomplished by one or more molecular functions (Smith & Kumar, 2004). Molecular function and biological processes terms are more closely related but the general rule is that the process must have more than one distinct step (Bentink, 2003). Examples of broad biological process terms are cellular physiological process or signal transduction. Examples of lower biological process terms are pyrimidine metabolism or alpha-glucoside transport(Ashburner et al., 2000). For the purpose of this study a biological process is defined according TAIR consortium as “Any process specifically pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms. A process is a collection of molecular events with a defined beginning and end following biological categories”.

Ashburner et al. (2000) defined Cellular component as the place in the cell where a gene product is active. Examples of Cellular component are an anatomical structure such as rough endoplasmic reticulum or nucleus and a gene product group such as ribosome, proteasome or a protein dimer(Ashburner et al., 2000).

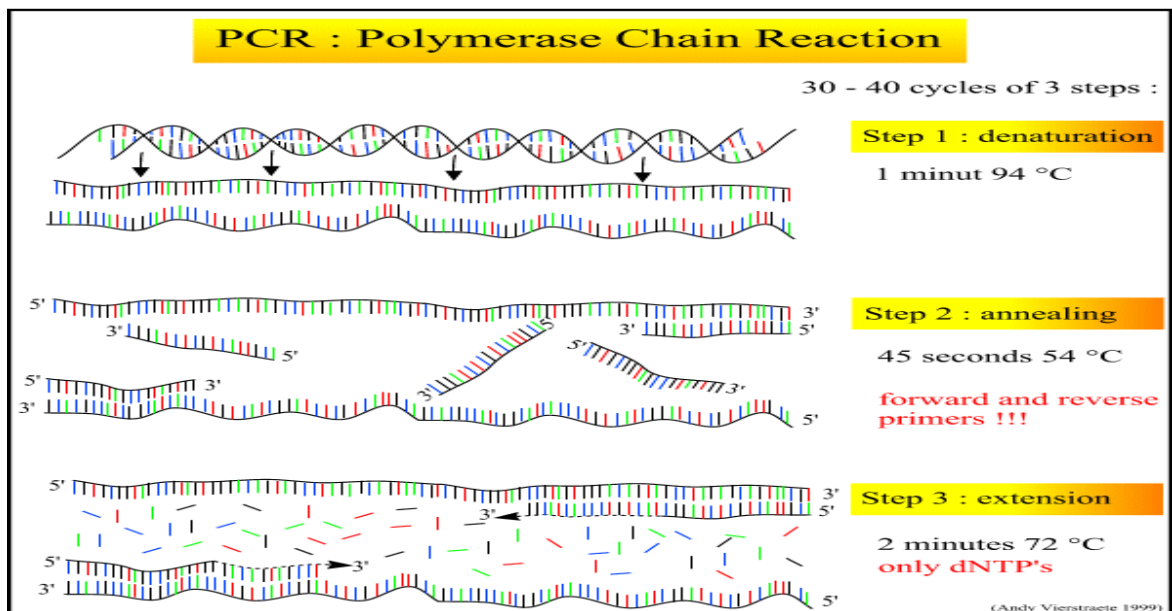
#### 2.2.4 Polymerase Chain Reaction (PCR)

There are two types of PCR namely standard PCR and real-time PCR. Real-Time PCR allow for the detection of PCR amplification during the early phases of the reaction. Measuring the kinetics of the reaction in the early phases of PCR provides a distinct advantage over traditional PCR detection. Traditional methods use Agarose gels for detection of PCR amplification at the final phase or end-point of the PCR reaction (Mullis, 1993). Although these two types of PCR differ fundamentally on how the actual amplification process occurs, their setup is similar. This setup involves several reagents: DNA template that contains the DNA region (target) to be amplified;

Two primers that are complementary to the 3' (three prime) ends of each of the sense and anti-sense strand of the DNA target; Taq polymerase or another DNA polymerase with a temperature optimum at around 70 °C; Deoxynucleotide triphosphates (dNTPs), the building-blocks from which the DNA polymerase synthesizes a new DNA strand; Buffer solution, which provides a suitable chemical environment for optimum activity and stability of the DNA polymerase; Divalent cations, magnesium or manganese ions; generally  $Mg^{2+}$  is used, but  $Mn^{2+}$  can be utilized for PCR-mediated DNA mutagenesis, as higher  $Mn^{2+}$  concentration increases the error rate during DNA synthesis Monovalent cation potassium ions (<http://www.eusem.com>).

PCR is a three step cycling process. The first step is to denature the target DNA so as to make it single-stranded and open up the complementary sequences of the primers. This is routinely done at 94 °C or 95 °C for up to one minute with 30 seconds being the norm.

The second step is to choose a primer annealing temperature ( $T_m$ ). The melting temperature of the primers determines this temperature. The usual place to set the  $T_a$  is about  $2^\circ\text{C}$  lower than the lowest  $T_m$  of the primers. Thus, if the primer melting temperatures are  $58.5^\circ\text{C}$  and  $59.2^\circ\text{C}$ , the  $T_m$  should be set at  $56.5^\circ\text{C}$  as a starting point. This temperature can, and should, be changed up or down depending upon the results. With regard to duration, the norm is around 30 seconds. This leaves the final step in the cycle, the polymerase extension step. The convention is to set this temperature to  $72^\circ\text{C}$ , the optimal temperature for Taq polymerase (Kainz, 2000; Mcpherson & Moller, 2000). A generic PCR cycling profile is presented in figure 3.



**Figure 3: Polymerase Chain Reaction parameters (retrieved at Andy Vierstraete ,1999).**

This study use standard PCR to confirm and characterized the previously created *X.americana* DNA-DNA hybridization.

## CHAPTER 3: METHODOLOGY

### 3. Materials and Methods

#### 3.1 Sample collection

The samples (*X. americana* leaves) were collected at Waterberg, Otjiwarongo in Otjozondjupa region during early summer (September) when average day time temperature reaches 30 °C with no recorded rainfall during that month. The samples were kept on ice from the field to the laboratory and subsequently stored in the freezer at -20 °C until the time of analysis.

#### 3.2 DNA extraction

DNA was extracted from leaves of *X. Americana* using a DNeasy Plant mini kit (Qiagen) according to manufactures' instructions with few modifications.

*X. americana* leaf (20mg) was ground with 800 µl API buffer using a mortar and pestle to get a homogeneous solution. The *X. americana* - buffer mixture was transferred to two 1.5 ml micro centrifuge tubes (400 µl) and 4 µl RNase was added to each tube. The mixture was incubated at 65°C for 10 minutes and it was mixed 3 times during incubation by inverting the tubes. To the lysate, 130 µl was added, mixed and incubated on ice for 5 minutes. The lysate was centrifuge for 5 minutes at 14, 000 rpm and then, it was transferred to QIAshredder Mini spin column (lilac) placed in a 2ml collection tube and centrifuged for 2 minutes at 14,000 rpm. Without disturbing the cell-debris pellet, the flow-through fraction from previous step was transferred into a new tube. The

volume of a flow-through was determined and 0.1 volume of sodium acetate (3M, pH5.2) was added. To the mixture, 0.7 volume of 50% isopropanol was added at room temperature. This mixture (650µl) including precipitate was transferred to the DNeasy Mini spin column placed in a 2 ml collection tube and it was centrifuged for 1 minute at 13,4 000 rpm and the previous step was repeated with the remaining sample. The DNeasy Mini spin column was placed into a new 2ml collection tube and 500 µl Buffer AW was added and it was centrifuged at 13,4 000 for 1 minute. The flow-through was discarded and was centrifuge at 14, 000 rpm for 2 minutes. The DNeasy Mini spin column was transferred to a 1.5 ml microcentrifuge tube and 50 µl Buffer AE (warmed at 65°C for 2 minutes) was directly added onto the DNeasy membrane. It was incubated at room temperature for 5 minutes, and then centrifuge for 1 minute at 14, 000 rpm to elute. The DNA was stored at -20 °C.

### **3.3 DNA quantification: analysis of genomic DNA using ND-Nanodrop 2000c**

The quantity of nucleic acid (DNA) from *X. americana* leave samples (1.5 µl ) were analyzed on a ND-NanoDrop 2000<sub>c</sub> spectrophotometer and results were obtained and tabulated.

### **3.4 Hybridization**

The Nottingham NASC arrays in the United Kingdom run a commercial service for DNA-DNA hybridization. DNA samples (50 µl) were run by NASC arrays with the following pre-defined experimental conditions. The DNA was labeled using Invitrogen BIoPrime kit according to manufacture instruction. About 100ng DNA was dissolved in

20  $\mu$ l of dilute buffer in a microcentrifuge tube. On ice, 20  $\mu$ l of 2.5X random primers solution were added. It was denatured by heat for 5 minutes in a boiling water bath; then immediately it was cooled on ice. To the solution on ice, 5  $\mu$ l of 10X dNTP mixture and distilled water to a volume of 49  $\mu$ l were added. It was mixed briefly and 1  $\mu$ l of Klenow fragment was added. It was mixed gently but thoroughly and it was centrifuge for 30 seconds. Thereafter, it was incubated at 37°C for 60 minutes and 5  $\mu$ l of stop buffer was added. While incubating, array was loaded with 200  $\mu$ l 1X hybe buffer and prehybe array in the oven for 10 minutes at 45°, 60 rpm. The probe was incubated for 5 minutes at 45°C and later the tubes were in for 5 minutes. The prehybe buffer was removed from the array and 200  $\mu$ l probes were loaded, the debris at the bottom of the tube was avoided. The septa were covered with tough spots to prevent leakage, and the arrays were placed in the oven at 45°C, 60rpm for 20 hours. The arrays were washed using the GeneChip operating software and after wash, CEL files data were exported and scanned.

Probe-pairs from each probe-set on the *A. thaliana* GeneChip array based on the hybridization efficiency of genomic DNA from *X. americana* was selected to homologous *A. thaliana* GeneChip array PM probes (Hammond et al, 2005). Genomic DNA from *X. americana* was biotin-labeled and hybridized to the *A. thaliana* ATH1 GeneChip array. Probe-sets were selected to analyses if the probe-set was represented by Perfect Match probes with gDNA hybridization intensities above a set threshold. Selection was performed using a .cel file parser script written in the Perl programming language (Xspecies Version 1.1, available at <http://affymetrix.arabidopsis.info/xspecies/>).

The regions that overlapped (define by probe sets) during cross-hybridization were selected. The primers were designed from those regions; probe sets (Hammond, 2005).

### **3.5 Primer design**

Primers were designed using the ProbeFinder Assay Design software ([www.lifescience.roche.com](http://www.lifescience.roche.com)). Genomic DNA sequences (see appendix 3) of *A. thaliana* genes (putative *X. americana* orthologous) were retrieved from Gene Bank data base (from the The *Arabidopsis* Information Resources (TAIR) website) to serve as templates for primer design. Primers were designed with a length from 15-30 bases and designed to line the region of interest without any other non-specific binding possibility. When primers were designed, necessary measures were followed according to Promega kit as per manufacture instruction to contain 40-60% (G+C); the 3'-ends of the primers were not complementary to avoid the production of primer-dimers. None of the primers contained three G or C nucleotides in a row near the 3'-end to avoid nonspecific primer annealing. Primers were design to yield similar melting temperatures. Primers were designed to yield a PCR product of between 70 and 100 bp. Fragments of this size have the advantage to be analyzed using multiple platforms such as Real-Time PCR or standard PCR.

### **3.6 Gene ontology classification**

Gene ontology analysis was done to evaluate if particular biological functions are over or under-represented during cross-hybridization. To classify the genes according to their putative function, the Gene Ontology (Gene Ontology Consortium, 2001) terms were

used to perform a classification analysis. The GO analysis was done on the TAIR Web site ([www.arabidopsis.org](http://www.arabidopsis.org)).

### **3.6.1 Gene Ontology Analysis of *A. thaliana* Genome**

*A. thaliana* genomic annotations were sub-classified using the bulk GO annotation retrieval tool available on the TAIR website. This tool was used to obtain a list of all GO annotations for a given set of genes using the locus identifiers. The number of genes (gene count) per category was noted and tabulated.

### **3.6.2 Gene Ontology Analysis of *A.thaliana/X. Americana* Hybridization Data**

Locus identifiers of more than 11000 probes (positive homology between *A. thaliana* and *X. americana*) were uploaded using the same bulk GO annotation retrieval tool available on the TAIR website. A list of all GO annotations for this set of identifiers was retrieved and number of genes (gene count) per category was noted and tabulated.

### **3.6.3 Gene Ontology Analysis of *X. americana* PCR confirmed Loci**

Locus identifiers of the 5 PCR positive *X. americana* loci (putative orthologous) were uploaded using the same bulk GO annotation retrieval tool available on the TAIR website. A list of all GO annotations for this set of identifiers was retrieved and number of genes (gene count) per category was noted and tabulated.

### 3.7 Independent confirmation using Standard PCR

Genes that were found during cross-hybridization were selected and analyzed to confirm if they are present in *X. americana* with standard PCR (table 1). First, DNA was isolated and quantified in *X. americana* leaves using the same protocols as in 3.2 and 3.3.

#### 3.7.1 Standard PCR amplification reactions

After *X. americana* DNA isolation and quantification, it was used in the standard PCR to amplify the targeted amplicon. The PCR tubes were placed on ice and the reagents (table 1) were added for each 25  $\mu$ l reaction. List of primers for PCR amplification is shown in table 2.

**Table 1: Reaction Master Mix for 25  $\mu$ l (Fermentas life science protocol, kit no: #K0171)**

Reagents/Components	Final concentration	1 $\times$ reaction
<b>PCR Master Mix 2<math>\times</math></b>	1 $\times$	12.5 $\mu$ l
<b>Primer forward</b>	0.3 $\mu$ M	1.2 $\mu$ l
<b>Primer Reverse</b>	0.3 $\mu$ M	1.2 $\mu$ l
<b>Template DNA</b>	500ng	2.5 $\mu$ l
<b>Nuclease free water</b>	Make up 25 $\mu$ l	7.6 $\mu$ l
Total	25 $\mu$ l	25 $\mu$ l

**Table 2: illustrate the primers IDs and the gene IDs for each primer**

#	Primer s ID	Forward primer	Reverse primer	Gene ID	Expected size (bp)
1	252921	cttccatgggaactcacgtt	ccccatacattgcacatcct	AT4G39030. 1	100
2	256075	caaattcagagcgtgaacca	tttctctcaaagtcaaattcaa gc	AT1G18150. 2	100
3	260133	aaccgctcttgcaccaagt	ccagttggcagcagacaaaaa	AT1G66340. 1	100
4	245523	aggttagatcaagaggcaga gtc	gcacgattggaaggtctgta g	AT4G15910. 1	100
5	263448	aactgcagaagttagctgcaca	tcagaatcgtcgtcatcatca	AT2G31660. 1	100
6	253705	ggtattctgaagaaactgggaa ga	ccatggctataaccgatttct g	AT4G29130. 1	100
7	259439	ttcaagctccgacaatttca	tgatgggttggtcaaaatca	AT1G01480. 1	100
8	264415	acatggacgatgggtcataag	gctgaaccggtaatcaatgg	AT1G43160. 1	100
9	249036	agctgcgccaacgaattat	cgattttgtcctccttgacc	AT5G44200. 1	100

10	255220	gaagttcaatgtttcgtttcatgt	ggattatacaaggcccaaaa a	AT4G05320. 2	100
----	--------	--------------------------	---------------------------	-----------------	-----

PCR reaction mixtures (master mix) were mixed thoroughly and PCR was carried out using thermal cycling conditions outlined in table 3. The PCR was left running over night. The PCR conditions were set as per Fermentas life science protocol, kit no: #K0171 (table 3).

**Table 3: Standard PCR reaction parameters**

Segment	Number of cycles	Target Temperature	Hold Time
<b>Pre-Denaturation</b>	1	94 °C	2 minutes
<b>Denaturation</b>	45	94 °C	1 minute
<b>Annealing</b>		60 °C	1 minute
<b>Extension</b>		72 °C	2 minutes
<b>Final extension</b>	1	72 °C	5 minutes
Hold	0	4 °C	0 minutes

### 3.7.2 PCR products gel electrophoresis

Exactly 2.2001g of agarose powder was weighed and transferred to a 500 ml screw cap reagent bottle. Then, 200 ml of 1×TBE buffer was added. The mixture was dissolved in

the microwave until no bubbles appeared in the solution. The solution was cooled under running tap water until the bottle was able to be held for at least 6 seconds. Thereafter, 5  $\mu$ l of ethidium bromide was added into the solution under the fume hood and it was poured into the gel tray. The 1% agarose gel was allowed to stand for 30 minutes and gently the comb was removed from the gel. The gel was transferred into the electrophoretic tank and filled with enough 1 $\times$ TBE buffer until the gel was completely covered with buffer. The PCR products (10  $\mu$ l) and loading dye (3  $\mu$ l) premixed were loaded on the gel. Ready to use 100bp ladder (8 $\mu$ l) was loaded in the first well. The gel was run for 60 minutes at 120V and the DNA bands were visualized on the UV transilluminator.

## CHAPTER 4: RESULTS AND DISCUSSION

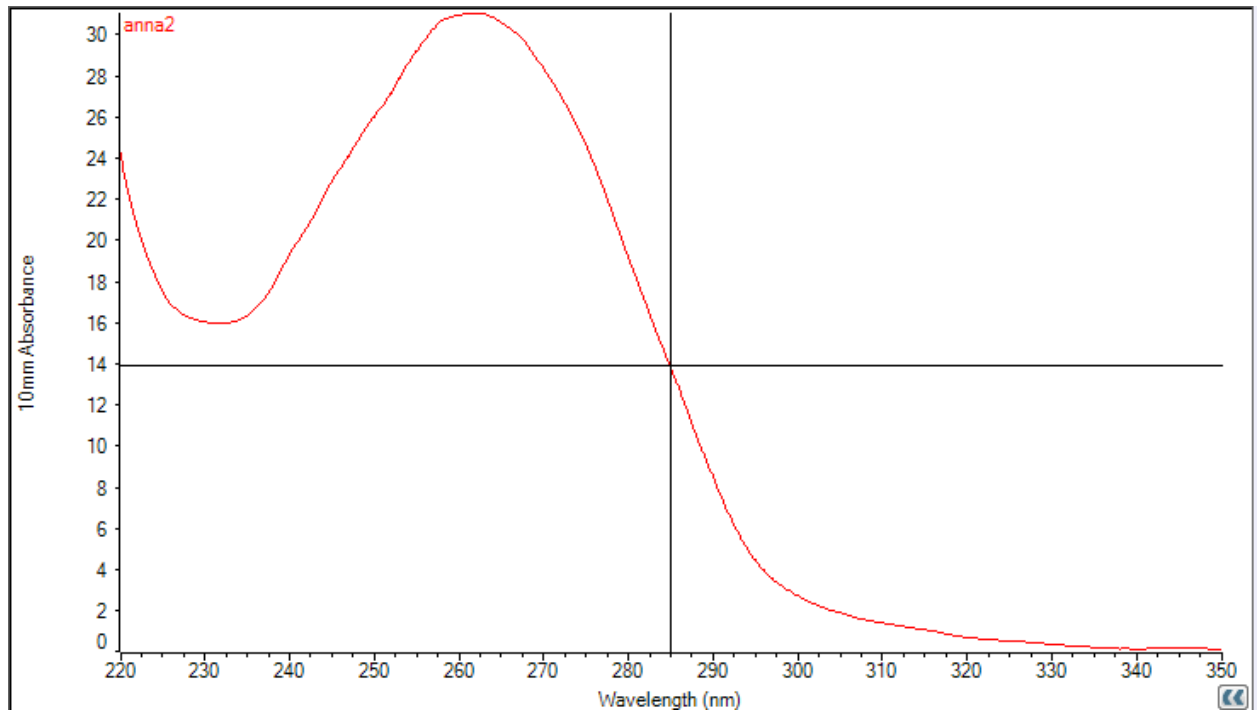
### 4.1 DNA quantification and Quality analysis using ND-Nanodrop 2000c

After DNA extraction, the nanoDrop 2000c was used to measure the concentration of DNA samples and quality of DNA i.e, the ratio at 260/280nm was to assess the purity of DNA. The ratio of ~1.8 is generally accepted as “pure” for DNA. If the ratio is appreciably lower, in may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280nm ([www.info@nanodrop.com](http://www.info@nanodrop.com)).

From the graph of DNA samples, the ratio is 2.05 indicating pure DNA with the concentration of 139ng/μl at 260/280 - ratio of absorbance at 260 nm and 280 nm (Table 4; Figure 4). This is a secondary measure of nucleic acid purity. The 260/280 values for a “pure” nucleic acid are commonly in the range of 1.8-2.2. If the ratio is appreciably lower, this may indicate the presence of co-purified contaminants.

**Table 4: shows concentration of *X. americana* DNA samples measured using nanodrop.**

Sample ID	Nucleic acid [ ] ng/μl	A <sub>260</sub>	A <sub>280</sub>	A <sub>260/280</sub>	A <sub>260/230</sub>	Sample type
XimeniaDNA	139	2.779	1.356	2.05	1.75	DNA
XimeniaDNA	62.5	1.251	0.611	2.05	1.44	DNA
XimeniaDNA	703	14.061	8.718	1.61	1.87	DNA



**Figure 4: the graph of *X. americana* DNA concentration (139ng/ $\mu$ l) measured using nanodrop.**

#### **4.2 *A. thaliana*/*X. americana* DNA-DNA Hybridisation Data**

*A. thaliana* ATH1 GeneChip arrays were used to study the putative orthologous genes of *X. americana*. Similar approach was applied to study ovine and *Brassica oleracea* (Graham et al., 2011; Hammond et al., 2005, 2006). The gDNA from *X. americana* was labelled and hybridized to the ATH1 GeneChip arrays. Hybridization signals were detected for all probes on the GeneChip for *X. americana* gDNA samples.

**Table 5: shows the genomic DNA hybridization intensity values**

Probe-set ID	Corresponding <i>A. thaliana</i>	<i>X. americana</i> DNA Leaves intensities
252921_at	AT4G39030	8.72618
256075_at	AT1G18150	7.74376
260133_at	AT1G66340	5.6179
245523_at	AT4G15910	6.54669
263448_at	AT2G31660	9.27173
253705_at	AT4G29130	5.6248
259439_at	AT1G01480	5.19156
264415_at	AT1G43160	7.21048
249036_at	AT5G44200	5.54982
255220_at	AT4G05320	5.10657

As shown in table 5 the average gDNA hybridization signal intensities for 10 selected putative orthologous genes on the ATH1 GeneChip array was 6.6589 for *X. americana*. Since gDNA from *X. americana* generated hybridization signals on the ATH1 GeneChip array, then this information can be used to identify probes and probe sets that have good homology between *A. thaliana* and *X. americana* for genetic analysis. Only one replicate was done in this study as less replicates are needed for experiments incorporating confirmatory testing such as PCR.

### 4.3 Gene Ontology Analysis of *A. thaliana* genome Annotations

The table below (table 6) shows number genes related to a particular GO biological process in *A. thaliana* genome.

**Table 6: shows GO biological process in *A. thaliana* genome**

GO Biological Process	Gene Count	%
response to abiotic or biotic stimulus	47	6
response to stress	46	6
unknown biological processes	66	6
protein metabolism	50	5
Transport	34	5
signal transduction	11	3
developmental processes	13	6
Transcription	9	3

The number of genes that carries annotations related to response to abiotic or biotic stimulus and response to stress are 47 and 46 respectively. Moreover, table 6 shows that about 66 genes in the genome of *A. thaliana* are not assigned to a biological function. In addition, both biological function genes that responsible for protein metabolism and transport are 5 % in the genome of *A. thaliana*. Furthermore, signal transduction and transcription have the least number of genes (3%).

#### 4.4 Gene Ontology Analysis of *A. thaliana*/*X. Americana* Hybridization Data

The table below (table 7) show the number of genes related to biological processes retrieve from the locus identifiers of hybridized probe sets (putative orthologous).

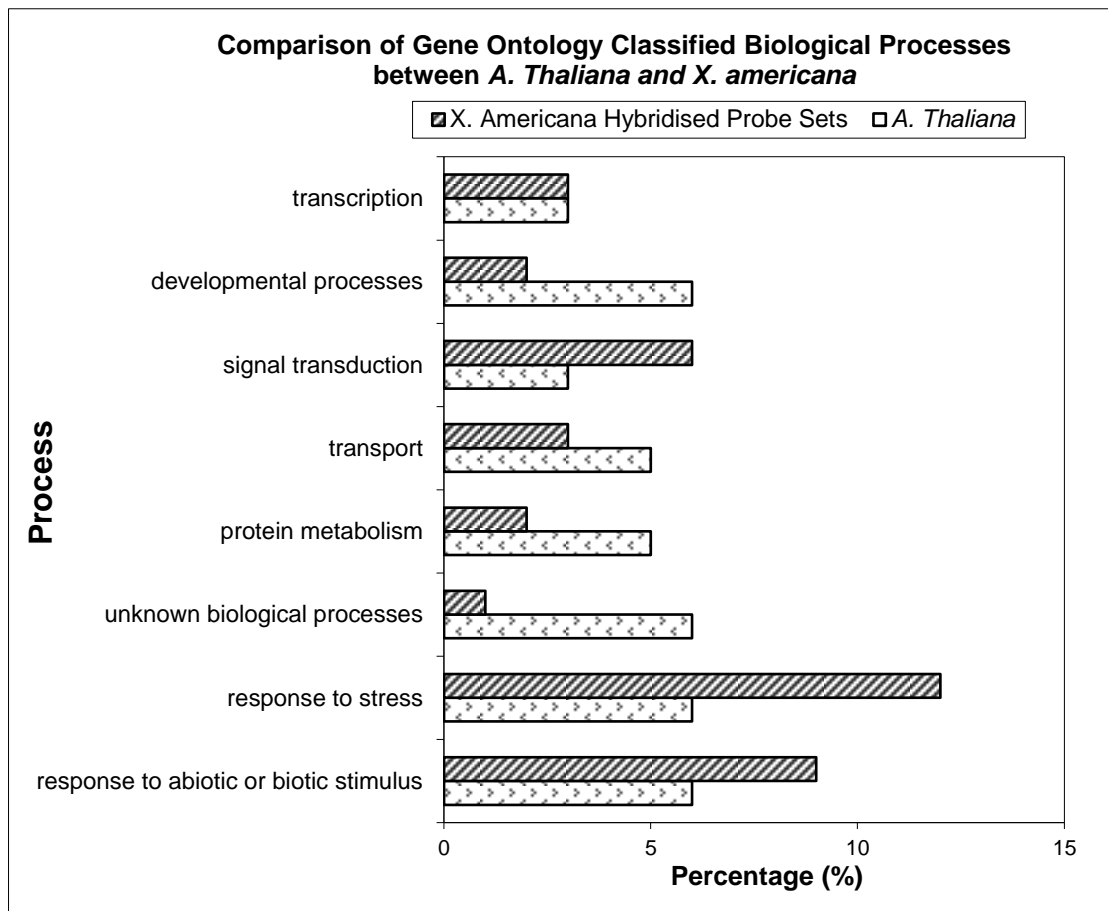
**Table 7: DNA-DNA Hybridization GO Annotations**

<b>GO Biological Process</b>	<b>Gene Count</b>	<b>%</b>
response to abiotic or biotic stimulus	1333	9
response to stress	1036	12
unknown biological processes	10226	1
protein metabolism	3961	2
transport	1699	3
signal transduction	906	6
developmental processes	1488	2
transcription	1701	3

The data shows (table 6) that number of genes related to response to abiotic or biotic stimulus and response to stress are 1333 and 1036 respectively. Moreover, the data shows that 10226 probe set representing putative orthologous genes of *X. americana* that are not assigned to a biological function. The number of this probe set is high compared to the *A. thaliana* because the genetic makeup of *X. americana* is not known yet. Putative orthologous genes related to Signal transduction has a least of 906 probe set.

#### 4.5 Comparative Gene Ontology Analyses

The Gene ontology classifications of biological processes were compared between *A. thaliana* and *X. americana*. Figure 5 compares gene-ontology results of biological process present in *Arabidopsis thaliana* with biological processes represented by hybridized regions between (probe set) *A. thaliana* and *X. americana* (<http://www.geneontology.org/>; <http://www.arabidopsis.org> ).



**Figure 5: Comparison of Gene Ontology Classified Biological Processes between *A. thaliana* and *X. americana***

The results illustrate that only 6% of biological processes in *A. thaliana* is related to stress response. The GO analysis further revealed that 6% of *A. thaliana* genome is not assigned to any specific biological process. Data suggest that 12% of biological process is assigned to stress response in the cross-hybridized region. The data suggest that representation of stress responses increased by 6% in *X. americana*. The data suggest over-representation of stress response in *X. americana* under natural conditions. This observation is confirmed by increase response to abiotic or biotic stimulus from 6% to 9%. Given the significance of the observation, an independent confirmation was needed.

#### **4.6 Independent confirmation**

The DNA-DNA hybridization identified more than 20 000 probes (genes) in this study. The same techniques were used in the study of *B. oleracea* and 40 099 genes were identified (Hammond et al., 2005). Putative orthologous genes of *X. americana* representing abiotic and stress responses were selected for independent confirmation in this study.

##### **4.6.1 DNA Quality analysis and quantification**

The result shows that the quality of DNA extracted was of reasonable standard for PCR amplification although is not pure (ratio of 260/280 which was 1.61) and the concentration of *X. americana* DNA samples were 1546.2 ng/μl and 700 ng/μl respectively on the two extracts (Table 8).

**Table 8: shows the DNA concentration using ND- NanoDrop 2000c**

Sample ID	Nucleic acid [ng/ $\mu$ l]	A <sub>260</sub>	A <sub>280</sub>	A <sub>260/280</sub>	A <sub>260/230</sub>	Sample type
XimeniaDNA	1546.2	30.925	19.226	1.61	1.97	DNA
XimeniaDNA	700	14.061	8.718	1.61	1.87	DNA

After quantification of *X. americana* DNA, using ND-NanoDrop 2000c (table 4), DNA amplification was done using designed primers to independently confirm if the ten genes that were found on the micro-array results are present in *X. americana*. The results in Figure 6 show that five genes out of ten targeted orthologous genes (appendix1) were positive (AT4G15910.1; AT2G31660.1; AT4G29130.1; AT1G01480.1; AT1G43160.1) as compared to the positive control AT4G05320.2 a housekeeping gene. The sizes of the genes were around 100 bp. The gel was run for 60 minutes.

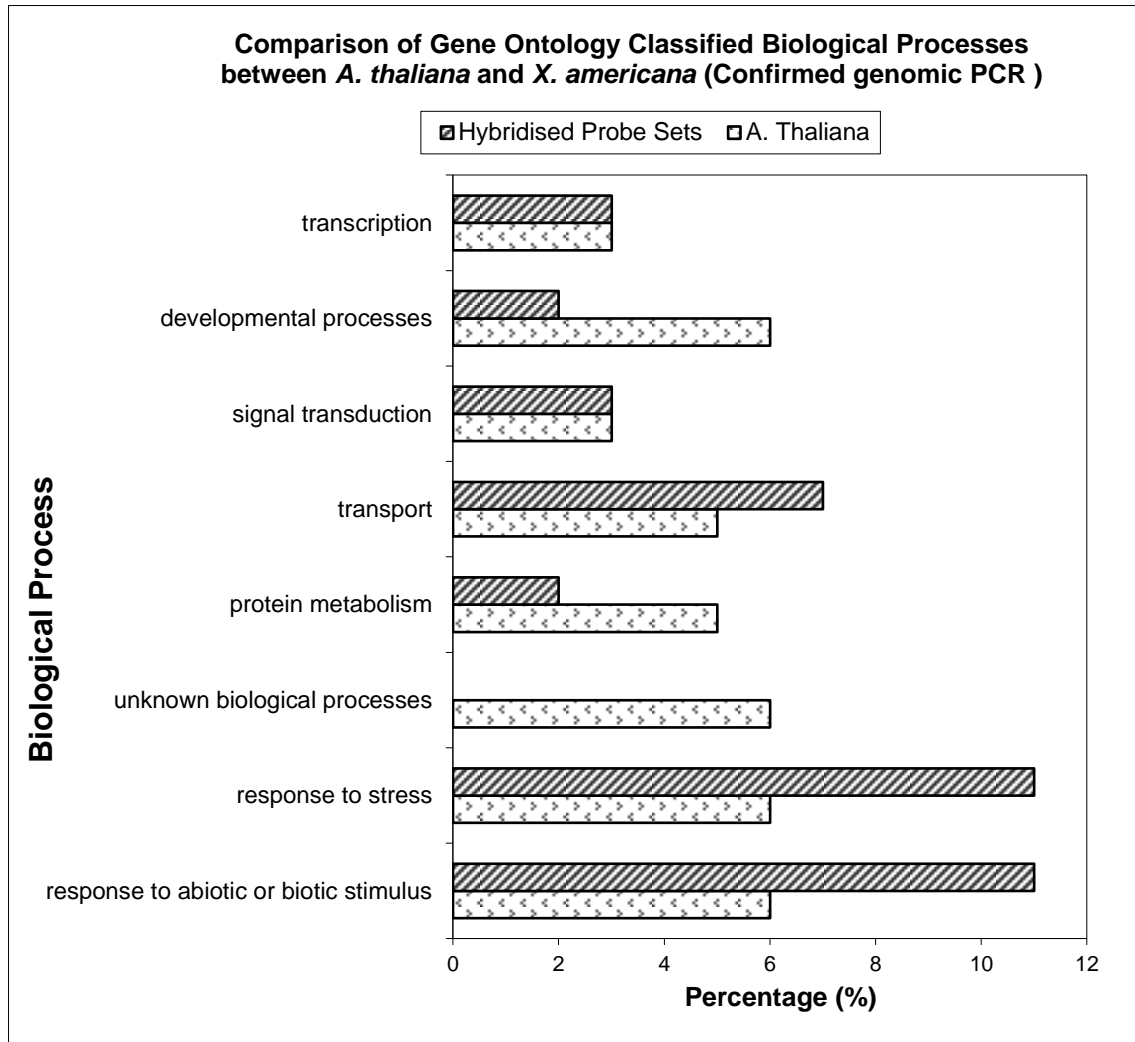


**Figure 6: shows the Amplifications of *X. americana* DNA with different primers on the 1% Agarose gel. A) DNA ladder (100bp); B) AT4G39031.1 (not presented); C) AT1G18150.2 (not present); D) AT1G66340.1 (not present); E) AT4G15910.1 (present); F) AT2G31660.1 (present); G) AT4G29130.1 (present); H) AT1G01480.1 (present); I) AT1G43160.1 (present); J) AT5G44200.1 (not present) and K) AT4G05320.2 (present, positive control)**

From PCR results, it was confirmed that five genes from the targeted ten genes are present in *X. americana*.

#### **4.6.2 Independent confirmation of *X. americana* Biological process by Standard PCR**

Gene ontology analysis of *X. americana* Genechip demonstrated that Abiotic stress response genes are overrepresented (12%) compared to Arabidopsis with 6%. Observation was independently confirmed with PCR amplification of the following orthologous genes: AT4G15910.1; SAD2; HXK1; ACC and ERF/AP2. Moreover it was confirmed that biological processes that response to abiotic and biotic stimulus, other metabolic processes, other cellular processes and transport are over-represented. Other biological processes that are found to be present and active include signal transduction, protein metabolism, transcription, and developmental processes (Figure 7).



**Figure 7: shows genomic PCR based independent confirmation of *X.americana* biological process.**

It is possible that the data reflects the *X. americana* environmental interaction. All five genes that were confirmed dependently by PCR and GO that are found in *X. americana* has to do with the environmental changes. These genes were over-represented in *X. americana* compared to *A. thaliana* because *X. americana* grows in dry, rocky places and harsh condition grows in a sandy, cold environment (Orwa et al., 2009). For

example AT4G15910.1 encodes a gene whose transcript level in root and leaves increases to progressive drought stress. This particular transcript level is also reported to be affected by changes of endogenous or exogenous abscisic acid level (Chen, Song, Zhang, & Zhang, 2013; Kawaguchi, Girke, Bray, & Bailey-Serres, 2004).

SAD2 (super sensitive to ABA and drought 2) encodes an importin beta-domain family protein likely to be involved in nuclear transport in ABA signaling. Loss of function mutations in SAD2 exhibit increased tolerance for UV stress, increased production of UV protective secondary metabolites and suppression of nuclear localization of MYB4 (a repressor of UV stress response genes) (Verslues, Guo, Dong, Ma, & Zhu, 2006; Zhao et al., 2007).

According to (Leung & Giraudat, 1998) ABA levels increase upon conditions of water stress during somatic growth, and ABA is an essential mediator in activating the plant responses to these adverse environmental stimuli. Leung & Giraudat,(1998) added that water loss can be limited by reducing stomata opening when Increased ABA levels. ABA is also involved in other aspects of stress adaptation. It was found that ABA-deficient mutants of *A. thaliana* are impaired in cold acclimation (Mantyla, Lang, & Palva, 1995) compared to *X. americana* that grows in dry, hot environment and in a root morphogenetic response to drought (Vartanian, 1994). The role of ABA in signaling stress conditions has also been extensively documented by molecular studies showing that ABA-deficient mutants are affected in the regulation of several genes by drought, salt, or cold (Leung & Giraudat, 1998).

From the ontology results (Figure 5) the data suggest that representation of stress responses increased by 6% in *X. Americana* and it shows also over-representation of stress response in *X. americana* under natural conditions, drought and heat which was dependently confirmed using PCR (Figure 6). The observation is confirmed by increase response to abiotic or biotic stimulus from 6% to 9%. According to (Ming-yi & Jian-hua, 2004) stated that one of the most essential regulator of plant responses to abiotic stresses is the phytohormone abscisic acid (ABA). Under drought, cold or salt stress condition, plants store higher amounts of ABA, with drought stress having the most prominent effect, thus why *X. americana* has more abiotic stress genes compared to *A. thaliana* due to the dry environment *X. americana* grows.

AT4G29130.1 which encodes a hexokinase (HXK1) in the plant glucose-signaling network functions as a glucose sensor to interrelate nutrient, light, and hormone signaling networks for controlling growth and development in response to the changing environment ( Cho, Yoo, & Sheen, 2006; Karve, Xia, & Moore, 2012; Moore et al., 2003) . Accordingly, metabolism, growth, development, and abiotic and biotic stress responses all are regulated, at least in part, by sugars (Harrington & Bush, 2003). In the study by Sarowar, Lee, Ahn, & Pai, (2008) they found that increasing hexokinase levels by HXK1/ HXK2 reduced or enhanced plant tolerance to methyl viologen. The Authors define Methyl Viologen as a compound that destroys the tissues of green plants by interfering with their photosynthesis. Its herbicidal property depends on the transformation of bipyridil residues to mono-cation radicals. The radical reacts with oxygen to render a superoxide anion, which can lead to the formation of H<sub>2</sub>O<sub>2</sub> and

hydroxyl radicals. Both products cause oxidative damage to cellular components, including protein inactivation, DNA damage, and lipid peroxidation (Timbrell, 2000).

AT1G01480.1 is a member of the 1-aminocyclopropane-1-carboxylate (ACC) synthase that catalyzes the conversion of 5-adenosylmethionine (SAM) to ACC, whereas ACC oxidase catalyzes the oxidation of ACC to ethylene. This plant hormone ethylene is produced in response to various kinds of environmental stress, such as wounding, physical load, disease, exposure to low temperature and chemicals and water stress (Karve et al., 2012; Kato, Hayakawa, Hyodo, Ikoma, & Yano, 2000; Li et al., 2012; X. Liang, Abel, Keller, Shen, & Theologis, 1992).

AT1G43160.1 encodes a member of the ERF (ethylene response factor)/AP2 (The APETALA2). These AP2/ERF proteins have important functions in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various responses to environmental Stimuli (Jung et al., 2010; Licausi, Ohme-Takagi, & Perata, 2013; Nakano, Suzuki, Fujimura, & Shinshi, 2006). Ethylene is observed as a stress hormone because its synthesis is induced by a variety of stress signals, such as mechanical wounding, chemicals and metals, drought, extreme temperatures and pathogen infection. 1-Aminocyclopropane-1-carboxylate (ACC) is an immediate precursor of ethylene in higher plants. Production of ethylene in plants is highly dependent on endogenous levels of ACC. Therefore, in the early stages of plant

response to stress, ACC accumulates along with a rapid burst in ethylene production (Bibi, Arshad, & Ahmed, 2006).

Although transcript levels of this important abiotic response factors could not be measured in absolute or relative terms, the study demonstrated inherent presence of such genes at the genomic level of *X. americana*. Further studies will be required to demonstrate that some sort of modification of any of these or other abiotic response factors within the genome of *X. americana* provides the key basis for its adaptation to the relatively dry and hot climatic conditions.

## CHAPTER 5: CONCLUSION

### 5 Conclusion

Orwa et al. (2009) describes *X. americana* as a drought-resistant plant, which mostly grows at low altitudes in a wide range of environment including, savannahs, dry woodlands, dry forests, and along coastal areas or on river banks. This study undertook a novel cross-species hybridization approach to study the biological processes that governs the adaptation of *X. americana* to its natural environment. Gene ontology analysis of perfectly match probe-pairs (during DNA-DNA cross hybridization with *A. thaliana*) showed increase in terms or ontologies that are associated with biological process and abiotic stress compared to *A. thaliana* ontology database. Such evidence not only confirms the conditions under which *X. americana* DNA samples were collected but suggest the molecular mechanisms that governs its adaptation to the prevailing natural conditions. *X. americana* grows under harsh conditions hence evolutionary adapted to survive under these conditions.

The *X. americana* DNA-DNA hybridization provides the first real data in understanding its molecular functions at the global genomic level. Such data alone however is not sufficient evidence to infer the molecular or biological functions of *X. americana*. Futures studies will involve creation of tentative *X. americana* DNA Chip from the perfect match hybridization probes. Such a chip can be used to study gene-expression profiles of *X. americana* under normal and drought conditions. Such an approach has been applied to study numerous un-sequences species. The study focused

on confirming the presence of 10 probe pairs out of the more than 20 000 perfect matched (PM) probe pairs. The confirmation of selected probe pairs did not only validate the hybridization data but confirmed the presence of stress related orthologous between *A. thaliana* and *X. americana* at the genomic level. Moreover, successful amplification of *X. americana* genomic fragments not only provided additional evidence of homology within primer binding sites. But illustrated some of the biological processes that integrate external stress signals such as drought stress and heat stress in *X. americana*.

## REFERENCES

- Abdallah, N. a, Moses, V., & Prakash, C. (2014). The impact of possible climate changes on developing countries. *GM Crops & Food*, 5(2), 77–80.  
doi:10.4161/gmcr.32208
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K & Walter P (2008) *Molecular Biology of the Cell*, 5th ed. New York: Garland Science
- Ashburner, M., Ball, C. a, Blake, J. a, Botstein, D., Butler, H., Cherry, J. M., ... Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1), 25–29. doi:10.1038/75556
- Auch, A. F., von Jan, M., Klenk, H.-P., & Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences*, 2(1), 117–134.  
doi:10.4056/sigs.531120
- Baldi, P., & Hatfield, W.G. (2002). *DNA microarrays and gene expression: From experiments to data analysis and modeling*. Cambridge, isbn: 0521800226.
- Bentink, S. (2003). *Gene Ontology as a tool for the systematic analysis*.
- Berardini, T.Z., Suparna Mundodi, S., Reiser, L., Eva Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L.A., Yoon, J., Doyle, A., Lander, G., Moseyko, N., Yoo, D., Xu, I., Zoeckler, B., Montoya, M., Miller, N., Weems, D., & Rhee, S. Y. (2004).

- Functional annotation of the Arabidopsis genome using controlled vocabularies. *Plant Physiology*, 135( 6), p. 745-755.
- Bibi, R., Arshad, M., & Ahmed, Z. (2006). Deaminase rhizobacteria extenuates acc-induced classical triple response in etiolated pea seedlings, 38(5), 1491–1499.
- Brazma, A., & Vilo, J. (2000). Gene expression data analysis. *FEBS Letters*, 480(1), 17–24. doi:10.1016/S0014-5793(00)01772-5
- Cancer, A. (2002). DNA Microarray Technology. What is it and how is it useful? Mandana Sassanfar and Graham Walker, 1–14.
- Cardwell, J. H., Cardwell, J., Johnson, P., & Stevens, J. (2013). A Review of Apomixis and Differential Expression Analyses Using Microarrays By.
- Chen, J., Song, Y., Zhang, H., & Zhang, D. (2013). Genome-Wide Analysis of Gene Expression in Response to Drought Stress in *Populus simonii*. *Plant Molecular Biology Reporter*, 31(4), 946–962. doi:10.1007/s11105-013-0563-6
- Cho, Y. H., Yoo, S. D., & Sheen, J. (2006). Regulatory Functions of Nuclear Hexokinase1 Complex in Glucose Signaling. *Cell*, 127(3), 579–589. doi:10.1016/j.cell.2006.09.028
- Eisen, M. B., & Brown, P. O. (n.d.). DNA Arrays for Analysis of Gene Expression.

- Goris, J., Konstantinidis, K. T., Klappenbach, J. a, Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, 57(Pt 1), 81–91. doi:10.1099/ijs.0.64483-0
- Graham, N. S., May, S. T., Daniel, Z. C. T. R., Emmerson, Z. F., Brameld, J. M., & Parr, T. (2011). Use of the Affymetrix Human GeneChip array and genomic DNA hybridisation probe selection to study ovine transcriptomes. *Animal : An International Journal of Animal Bioscience*, 5(6), 861–6. doi:10.1017/S1751731110002533
- Hammond, J. P., Bowen, H. C., White, P. J., Mills, V., Pyke, K. a., Baker, A. J. M., ... Broadley, M. R. (2006). A comparison of the *Thlaspi caerulescens* and *Thlaspi arvense* shoot transcriptomes. *New Phytologist*, 170(2), 239–260. doi:10.1111/j.1469-8137.2006.01662.x
- Hammond, J. P., Broadley, M. R., Craigon, D. J., Higgins, J., Emmerson, Z. F., Townsend, H. J., ... May, S. T. (2005). Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. *Plant Methods*, 1(1), 10. doi:10.1186/1746-4811-1-10
- Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., & Masmoudi, K. (2011). Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signaling & Behavior*, 6(10), 1503–1509. doi:10.4161/psb.6.10.17088

- Harrington, G. N., & Bush, D. R. (2003). The bifunctional role of hexokinase in metabolism and glucose signaling. *The Plant Cell*, *15*(11), 2493–2496.  
doi:10.1105/tpc.151130
- Hartung, W., Sauter, A., & Hose, E. (2002). Abscisic acid in the xylem: where does it come from, where does it go to? *Journal of Experimental Botany*, *53*(366), 27–32.  
doi:10.1093/jexbot/53.366.27
- Iuchi, S., Kobayashi, M., Taji, T., Naramoto, M., Seki, M., Kato, T., ... Shinozaki, K. (2001). Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in *Arabidopsis*. *Plant Journal*, *27*(4), 325–333. doi:10.1046/j.1365-313X.2001.01096.x
- Jung, K.-H., Seo, Y.-S., Walia, H., Cao, P., Fukao, T., Canlas, P. E., ... Ronald, P. C. (2010). The submergence tolerance regulator Sub1A mediates stress-responsive expression of AP2/ERF transcription factors. *Plant Physiology*, *152*(3), 1674–1692.  
doi:10.1104/pp.109.152157
- Kainz, P. (2000). The PCR plateau phase - Towards an understanding of its limitations. *Biochimica et Biophysica Acta - Gene Structure and Expression*, *1494*(1-2), 23–27.  
doi:10.1016/S0167-4781(00)00200-1

- Karve, a., Xia, X., & Moore, B. D. (2012). Arabidopsis Hexokinase-Like1 and Hexokinase1 Form a Critical Node in Mediating Plant Glucose and Ethylene Responses. *Plant Physiology*, *158*(4), 1965–1975. doi:10.1104/pp.112.195636
- Kato, M., Hayakawa, Y., Hyodo, H., Ikoma, Y., & Yano, M. (2000). Wound-induced ethylene synthesis and expression and formation of 1-aminocyclopropane-1-carboxylate (ACC) synthase, ACC oxidase, phenylalanine ammonia-lyase, and peroxidase in wounded mesocarp tissue of *Cucurbita maxima*. *Plant & Cell Physiology*, *41*(4), 440–447. doi:10.1093/pcp/41.4.440
- Kawaguchi, R., Girke, T., Bray, E. a., & Bailey-Serres, J. (2004). Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant Journal*, *38*(5), 823–839. doi:10.1111/j.1365-313X.2004.02090.x
- Lenka, S. K., Lohia, B., Kumar, A., Chinnusamy, V., & Bansal, K. C. (2009). Genome-wide targeted prediction of ABA responsive genes in rice based on over-represented cis-motif in co-expressed genes. *Plant Molecular Biology*, *69*(3), 261–271. doi:10.1007/s11103-008-9423-4
- Leung, J., & Giraudat, J. (1998). Abscisic Acid Signal Transduction. *Annual Review of Plant Physiology and Plant Molecular Biology*, *49*, 199–222. doi:10.1146/annurev.arplant.49.1.199

- Li, G., Meng, X., Wang, R., Mao, G., Han, L., Liu, Y., & Zhang, S. (2012). Dual-level regulation of ACC synthase activity by MPK3/MPK6 cascade and its downstream WRKY transcription factor during ethylene induction in arabidopsis. *PLoS Genetics*, 8(6). doi:10.1371/journal.pgen.1002767
- Liang, X., Abel, S., Keller, J. a, Shen, N. F., & Theologis, a. (1992). The 1-aminocyclopropane-1-carboxylate synthase gene family of Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22), 11046–11050. doi:10.1073/pnas.89.22.11046
- Liang, Y., Zhang, F., Wang, J., Joshi, T., Wang, Y., & Xu, D. (2011). Prediction of drought-resistant genes in Arabidopsis thaliana using SVM-RFE. *PloS One*, 6(7), e21750. doi:10.1371/journal.pone.0021750
- Licausi, F., Ohme-Takagi, M., & Perata, P. (2013). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. *New Phytologist*, 639–649. doi:10.1111/nph.12291
- Mantyla, E., Lang, V., & Palva, E. T. (1995). Role of Abscisic Acid in Drought-Induced Freezing Tolerance, Cold Acclimation, and Accumulation of LT178 and RAB18 Proteins in Arabidopsis thaliana. *Plant Physiology*, 107(1), 141–148. doi:10.1104/pp.107.1.141

Ming-yi, J., & Jian-hua, Z. (2004). Abscisic Acid and Antioxidant Defense in Plant Cells  
2 Expression of Genes Encoding Antioxidant 3 Effect of ABA on Enzymatic and  
Non-enzy-.

Moore, B., Zhou, L., Rolland, F., Hall, Q., Cheng, W.-H., Liu, Y.-X., ... Sheen, J.  
(2003). Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and  
hormonal signaling. *Science (New York, N.Y.)*, 300(5617), 332–336.  
doi:10.1126/science.1080585

Nakano, T., Suzuki, K., Fujimura, T., & Shinshi, H. (2006). Genome-Wide Analysis of  
the ERF Gene Family. *Plant Physiology*, 140(February), 411–432.  
doi:10.1104/pp.105.073783.currently

Parokony, a. S., Marshall, J. a., Bennett, M. D., Cocking, E. C., Davey, M. R., & Brian  
Power, J. (1997). Homoeologous pairing and recombination in backcross  
derivatives of tomato somatic hybrids [*Lycopersicon esculentum* (+) L.  
peruvianum]. *Theoretical and Applied Genetics*, 94(6-7), 713–723.  
doi:10.1007/s001220050470

Rhee, S. Y., Wood, V., Dolinski, K., & Draghici, S. (2008). Use and misuse of the gene  
ontology annotations. *Nature Reviews. Genetics*, 9(7), 509–515.  
doi:10.1038/nrg2363

- Sarowar, S., Lee, J. Y., Ahn, E. R., & Pai, H. S. (2008). A role of hexokinases in plant resistance to oxidative stress and pathogen infection. *Journal of Plant Biology*, 51(5), 341–346. doi:10.1007/BF03036136
- Savitri, E. S., Basuki, N., Aini, N., & Arumingtyas, E. L. (2013). Identification and characterization drought tolerance of gene LEA-D 11 soybean ( *glycine max L . Merr* ) based on, 2013(January), 32–37.
- Silva, G. S., & Souza, M. M. (2013). Genomic in situ hybridization in plants. *Genetics and Molecular Research*, 12(3), 2953–2965. doi:10.4238/2013.August.12.11
- Smith, B., & Kumar, A. (2004). Controlled vocabularies in bioinformatics: A case study in the gene ontology. *Drug Discovery Today: BIOSILICO*, 2(6), 246–252. doi:10.1016/S1741-8364(04)02424-2
- Söderman, E. M., Brocard, I. M., Lynch, T. J., & Finkelstein, R. R. (2000). Regulation and function of the Arabidopsis ABA-insensitive4 gene in seed and abscisic acid response signaling networks. *Plant Physiology*, 124(4), 1752–1765. doi:10.1104/pp.124.4.1752
- Orwa, C., Mutua, A., Kindt, R., Jamnadass, R., & Simons, A. (2009). Agroforestry Database: a tree reference and selection guide. Version 4.0. <http://www.worldagroforestry.org>

Timbrell, J. (2000). Principles of biochemical toxicology 3rd ed. *London: Taylor and Francis.*

Tuteja, N. (2007). Abscisic Acid and abiotic stress signaling. *Plant Signaling & Behavior*, 2(3), 135–138. doi:10.4161/psb.2.3.4156

Usman, M. G., Rafii, M. Y., Ismail, M. R., Malek, M. A., Latif, M. A., & Oladosu, Y. (2014). Heat Shock Proteins : Functions And Response Against Heat Stress In *Plants*, 3(11), 204–218.

Verslues, P. E., Guo, Y., Dong, C. H., Ma, W., & Zhu, J. K. (2006). Mutation of SAD2, an importin  $\beta$ -domain protein in Arabidopsis, alters abscisic acid sensitivity. *Plant Journal*, 47(5), 776–787. doi:10.1111/j.1365-313X.2006.02833.x

Wang, D., Pan, Y., Zhao, X., Zhu, L., Fu, B., & Li, Z. (2011). Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. *BMC Genomics*, 12(1), 149. doi:10.1186/1471-2164-12-149

Warner, E. E., & Dieckgraefe, B. K. (2002). Basic Science on the Cutting Edge Application of Genome-Wide Gene Expression Profiling by High-Density DNA Arrays to the Treatment and Study of Inflammatory Bowel Disease, 8(2), 140–157.

Water, C., Water, C., Kong, H., & Kong, H. (2000). Pathways to abscisic acid-regulated gene expression. *Review Literature And Arts Of The Americas*, (120), 357–396.

- Wayne, L. G., Brenner, D. J., Colwell, R. R., Grimont, P. a. D., Kandler, O., Krichevsky, M. I., ... Truper, H. G. (1987). Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *International Journal of Systematic Bacteriology*, 37(4), 463–464. doi:10.1099/00207713-37-4-463
- Zalejski, C., Paradis, S., Maldiney, R., Habricot, Y., Miginiac, E., Rona, J.-P., & Jeannette, E. (2006). Induction of abscisic acid-regulated gene expression by diacylglycerol pyrophosphate involves Ca<sup>2+</sup> and anion currents in Arabidopsis suspension cells. *Plant Physiology*, 141(4), 1555–1562. doi:10.1104/pp.106.080218
- Zhao, J., Zhang, W., Zhao, Y., Gong, X., Guo, L., Zhu, G., ... Guo, Y. (2007). SAD2, an importin -like protein, is required for UV-B response in Arabidopsis by mediating MYB4 nuclear trafficking. *The Plant Cell*, 19(11), 3805–3818. doi:10.1105/tpc.106.048900
- <http://www.eusem.com>
- <http://www.arabidopsis.org>
- <http://www.lifescience.roche.com>
- <http://www.info@nanodrop.com>
- <http://www.geneontology.org/>
- <http://affymetrix.arabidopsis.info/xspecies/>

## APPENDIX

### 1. PCR of Targeted orthologous genes

Table 9; shows the ten targeted orthologous genes for independent confirmation.

<b>Orthologue Gene</b>	Description of Orthologue Target	Left Primer	Right Primer	Expected Target Size (bp)
<a href="#"><u>AT4G3903</u></a> <a href="#"><u>0.1</u></a>	Encodes an orphan multidrug and toxin extrusion transporter. Essential component of salicylic acid-dependent signaling for disease resistance. Member of the MATE-transporter family. Expression induced by salicylic acid. Mutants are salicylic acid-deficient.	cttccatg ggaactca cggt	ccccatac attgcacat cct	100
<a href="#"><u>AT1G1815</u></a> <a href="#"><u>0.2</u></a>	Encodes mitogen-activated protein kinase 8 (MPK8). MPK8 connects protein phosphorylation, Ca <sup>2+</sup> , and ROS in the wound-signaling pathway.	caaattca gagcgtg aacca	tttctctcaa agtcaaatt caagc	100
<a href="#"><u>AT1G6634</u></a> <a href="#"><u>0.1</u></a>	Similar to prokaryote sensory transduction proteins. Contains a histidine kinase and a response	aaccgctc ttgtcacca agt	ccagttgg cacgacaa aaa	100

	<p>regulator domain. Homodimer. Membrane component. Binds ethylene. Mutations affect ethylene binding and metabolism of other plant hormones such as auxin, cytokinins, ABA and gibberellic acid. Ethylene receptor. Has histidine kinase activity. Is regulated by RTE1. Mutations in ETR1 block ethylene stimulation of flavonol synthesis.</p>			
<p><a href="#">AT4G1591</a> <a href="#">0.1</a></p>	<p>Encodes a gene whose transcript level in root and leaves increases to progressive drought stress. The transcript level is also affected by changes of endogenous or exogenous abscisic acid level.</p>	<p>aggttag atcaagag gcagagtc</p>	<p>gcacgatt ggaaggtc tgtag</p>	<p>100</p>
<p><a href="#">AT2G3166</a> <a href="#">0.1</a></p>	<p>SAD2 (super sensitive to ABA and drought 2) encodes an importin beta-domain family protein likely to be involved in nuclear transport in ABA signaling. Subcellular localization of</p>	<p>aactgcag aagttagc tgcaca</p>	<p>tcagaatc gtcgtcatc atca</p>	<p>100</p>

	<p>GFP-tagged SAD2 showed a predominantly nuclear localization, consistent with a role for SAD2 in nuclear transport. Mutation of SAD2 in Arabidopsis alters abscisic acid sensitivity. SAD2 was ubiquitously expressed at low levels in all tissues except flowers. SAD2 expression was not induced by ABA or stress. Loss of function mutations in SAD2 exhibit increased tolerance for UV stress, increased production of UV protective secondary metabolites and suppression of nuclear localization of MYB4 (a repressor of UV stress response genes). Regulates microRNA activity.</p>			
<p><a href="#">AT4G2913</a> <a href="#">0.1</a></p>	<p>Encodes a hexokinase (HXK1) in the plant glucose-signaling network. Functions as a glucose sensor to interrelate nutrient, light, and hormone signaling networks for controlling</p>	<p>ggtattctg aagaaact gggaaga</p>	<p>ccatggct ataaccga tttctg</p>	<p>100</p>

	growth and development in response to the changing environment.			
<a href="#"><u>AT1G0148</u></a> <a href="#"><u>0.1</u></a>	a member of the 1-aminocyclopropane-1-carboxylate (ACC) synthase (S-adenosyl-L-methionine methylthioadenosine-lyase, EC 4.4.1.14) gene family, isolated from a flower-specific cDNA library.	ttcaagctc cgacaatt ca	tgatgggtt ggtaaaaa tca	100
<a href="#"><u>AT1G4316</u></a> <a href="#"><u>0.1</u></a>	encodes a member of the ERF (ethylene response factor) subfamily B-4 of ERF/AP2 transcription factor family (RAP2.6). The protein contains one AP2 domain. There are 7 members in this subfamily.	acatggac gatgggtc ataag	gctgaacc ggtaatca atgg	100
<a href="#"><u>AT5G4420</u></a> <a href="#"><u>0.1</u></a>	Encodes a nuclear cap-binding protein that forms a heterodimeric complex with ABH1 (ATCBP80) and is likely to participate in RNA metabolism. Its mRNA is ubiquitously expressed. Loss of function mutations suggest a role in	agctgcgc caacgaat tat	cgattttgt cctcctga cc	100

	processing of pri-miRNA and mRNA splicing.			
<a href="#">AT4G0532</a> <a href="#">0.2</a>	One of five polyubiquitin genes in A. thaliana. These genes encode the highly conserved 76-amino acid protein ubiquitin that is covalently attached to substrate proteins targeting most for degradation. Polyubiquitin genes are characterized by the presence of tandem repeats of the 228 bp that encode a ubiquitin monomer. Induced by salicylic acid. Independent of NPR1 for their induction by salicylic acid	gaagttca atgttcggt tcatgt	ggattatac aaggcccc aaaa	100

## 2. Primer design

AT4G39030.1

ATGCTAATCAAATCCCAGATTGACTCTTTTCTCTCCTCTCTGTCCAAAACAAG  
ACGGATCCCGGTTAACTCCCACCAAACACTGGTTCGCAGAATCGGTGATAACT  
CGGAGAACAACCTCGGTGCGATCACCGCAACTCCGAGTTTTTCATAAAAATCCGG  
TGGTGATTTCGACGGAGGATCAAATTAGAGAGAGTTACGAGGAACTGCGTCA  
GAATTGATCGGGAAATCGACGAAGAAGAAGAAGAGGAGAAGGAGAGA  
GGGGATCTGGTGAAACAGAGCATATGGGAACAGATGAAAGAGATAGTGAA  
GTTACAGGTCCGGCGATGGGGATGTGGATTTGTGGACCGTTGATGAGTCTC  
ATCGACACGGTCGTCATCGGCCAAGGAAGCTCCATCGAACTCGCTGCTCTTG  
gtacgctttcaatcatactgtctttccggttagtagccaatfttgaaccggtggactggtcccctacccaatagctaagctgtaa  
gattcatacactgftttggtcctttgtagtctttcgaatatattacacagatttgaaccaataatagtttaaaacaggtaac  
aaattaggcaagtgtccagctatcatcttatttttaatcattttccaagccaatcattagattgtgtcggtttaccaactcgaaa  
attagtagtgtgtaacgftttcacgattcttctagacagaaaaaataataatcgtaacataaftaaagaa  
caaatacatatgftaaagftaaatggtggtctgaagaaaaggcaaatggttttagtctcgttggtcaaagftaagaccgftttca  
agcatttgattatacaaattgactagcaacgcaaggtaactaacataatcaatgftaaaaagaaaaaattgfttgattttgg  
tgagaatacggtaataatgtagtatatttatatgatggctgagaaaattaccgfttaatagGACCGGGAACAGT  
ACTATGCGACCACATGAGTTATGTCCTTCATGTTCCCTCTCCGTGGCTACATCCA

ATATGGTTGCTACTTCTCTTGCCAAACAGgttattacttttactctttcaacttggttttcttctctga  
tgcgaaacatctcaaatgttattatattcttgagGACAAGAAAGAAGCGCAACATCAAATCTCT  
GTTTTACTCTTTATTGGATTGGTTTGTGGACTAATGATGCTTCTGCTCACAAG  
ATTATTCGGTCCTTGGGCTGTTACTGgtaaacatttttagtacaaaagtagttccgtttaagaaatatt  
gagtgaatgtgattfacactctgtcatggtttttgagCTTTTACAAGGGGGAAGAACATTGAGA  
TTGTCCCTGCAGCTAATAAGTATATTCAGgttcttgctggtgatcattctctgttcaaatctctgatg  
aatggtgtctctttaactggaatctctgtctcgagATTCGAGGCTTAGCGTGGCCGTTTATCCT  
TGTTGGATTGGTTGCTCAAAGTGCAAGgtttctfacactccctcttattaatcagcttaacgttcgatg  
tctgtgcgatttccgctctttacttaattgatgtgagaatctgtctcagTCTTGGAATGAAGAACTCATG  
GGGACCTCTTAAAGCATTAGCTGCAGCAACGATCATTAAACGGTCTTGGCGAT  
ACAATCCTATGCTTGTCTTGGACAAGGTATCGCTGGAGCTGCTTGGGCAA  
CAACTGCTTCACAGgtaatattaattatagatcatgaagccaacactgaaatgaaatgcagaagattcattacg  
gatttatgtgttggttatgtagATTGTTTCGGCTTATATGATGATGGATTCTCTGAACA  
AAGAAGGCTACAATGCTTACTCTTTTGCATCCCTTACCAGCAAGAACTATG  
GAAGATCTCTGCACTCGCTGCACCTGTTTTTATCTCCATTTTCTCCAAGATTG  
CTTCTACTCTTTCATAATTTACTGTGCTACTTCCATGGGAACCTCACGTTTAA  
GCCGCTCATCAGgtacattacaaaacagagatctttatgccagtcaggttggtgtgatattataattgtggtttatta  
cagGTGATGGCTCAGACGTATAGGATGTGCAATGTATGGGGTGAGCCACTCTC  
TCAAACGGCTCAGTCATTTATGCCAGAGATGTTATATGGTGCGAACCGTAAT  
CTTCCAAGGCTAGAACGTTGCTTAAAGTCCCTGATGATTATTGGAGCCACGC  
TTGGATTAGTCTTGGGAGTTATCGGAACAGCGGTTCCAGGCCTGTTTCCTGG  
TGTCTACACACATGATAAAGTCATCATATCCGAGgtctggtttgttttcaatcaacattgttc

ctgtgttttgattctaactggctaatctgattcttgatatgttgccagATGCATAGACTGCTTATAACCTTTC  
 TTCATGGCGTTGTCTGCATTGCCATGACAGTATCCCTCGAGGGTACATTGCT  
 Ggtaaattacaaactctgttcgattaggtggttcttgagaagcaagcagctgaataagtctctatattttcaatcttgaacaag  
 atagatgcttaaaccaaggtgttctgttcaccagGCGGGACGTGATCTCAAATTCGTCAGCTC  
 AGTGATGAGTTCAAGCTTCATAATTGGTTGTCTCACACTAATGgtactttatagagtctc  
 tcataaaccttaactacgtttattaatctcaaagacactttcaatttgagcaaatctaaaatactttttgttggataatgacagTT  
 CGTGACAAGAAGTGGCTATGGTTTACTCGGCTGCTGGTTCGTTCTCGTCGGA  
 TTCAATGGgtaaagacagacatacacagcaacatttgttatcagagattttctttatccaaatctgtctttaaccaaat  
 atggttaacattgcagGGCCGGTTTGGTCTATATCTACGACGGCTTCTTTTCGCCTGGAG  
 GCATACTTAATTCAGATGGGCCGAGTCCATACACGGTGGAGAAGATTAAT  
 CCATT**TA**

AT1G18150.2

**TTA**AGAATTGTGAAGAGAAGCAACTTTATCAGTAAGCTCTGCAACTGTTTCA  
 TTATCTTCTCCTCAGCTATGGTATCctgcaaaacaagagaaaagaacaacagtaagcaaaactcatc  
 gtcattgtcctttcaagatcttctgttttcttcttcttcttactaactaacCTCTTTGTCAGTTTTAGATTGGA  
 CACCAATGCATTTAGAACCGCTGATGCTTGCACTCTTCATGAGGTTACGAGC  
 ACTGTAACCTCCTCCATTGTCTGCTTCTCGGATTCCAAGGTTGAAGCCACA  
 GCAGCTGCTGCTTCTCTCAACATCACTGCTTCTTCTGTCAGTTTCACCATT  
 TGGAGCAGGAAGTCTCTCTctgcaacatcatcatagagaagcacaatgctctcagagggagattgagaaga  
 gaagagaagagatagagagattctgtgaaagacattacCTTGGCAAGGAAGCATGATGTCTGTGAA

GTGCAGTACTTCTTCCTCCTCCTGCTGCTCCTGGTTTACCTTGATTTTCTTCAA  
 GATGAGCAAATTGCCTTTTAAATCGGTCTACCCCActgcgtaaacacaaaaagattatc  
 gagtcataactttctgtttcaagacttatagttctgtaactgtaattagctaacCTTGGGTACATAAAGCTAA  
 GCTGATCACCACCACGAAGGTATTCTCCAACATCTGAGGATGATATTCCAA  
 TATctgatattfcaaaaatcgaagaaaagtgattgtaggatcaatgaataacattgaaacaaattcagattaagaaaaggat  
 atatgtagaaggctttgtgtttgtttgttacCTCTCGGTAAATTAATTCTCTGACATCATCTTTAA  
 CTAACTTC**TTTCTCTCAAAGTCAAATTC**AGC**TTTGAAATTGGCTGTGTTGT****T**  
**GGTTCACGCTCTGAATTTG**ACAAACCACTGAAATATGGATCAGCTAGTGCctgt  
 aatatgaaatcaaaaacgacgaatataagagtttctgtataaaaacaaaaggaaaagtaaagcaaatttacttctcacttac  
 ATCTTCAGCTGATGCACGATCTTTCGGATCAAAGGCAAGCAGGCGTTCCAGA  
 AGGCGGAGAGCCAAAGGATCAGCTTTAGGGAATTTATGAGAGAAAGGGACT  
 GGTTGTTTTTTCCTCATGCTGCTAAGATATCTCCTCGCCTTTTCATTTCTAATct  
 gtagaagtccaagattgcaaagctcattacaaggattaaattacttcaatgatgatggaacattaaattgctacagaaaatag  
 aacaagtgtttgagtactaacCCTTGATATTGACTCAGGAGGTGGAGTACCAAGAAAGT  
 CAGTCATTAGATCCAATTGGTGCACAACGTTTTTCCCGGGAAACAAAGGCTT  
 ACCCAAAAGCATTTCGCAAAAATGCAACCGACTCCAGATATCAATCGC  
 AGGGGTATActgaaatataacaagagtagacttagcttgatgctgctgtcttctcaacaactatagaaccacaaa  
 caatattctacaaaatagaaaactagaagcattttctgtagggagaaatagcaagagtagtgggtttgtgaaaaagaatctt  
 acTTTGGAGAAAAACGATCCACAGAGTTCAGGGGCACGATACCACCGAGTAG  
 CTACATAATCctacaaaatccatataaacgaaatgagctcacaagcatttgaatagttaagaggaagcctaataagaa  
 cagttttgctttcacattgagttagatgactcacAGTCCAAAATATAGCGGTTGGGGCATCGTTAA  
 AAGAAACGCGAGCCAGCCCAAAAATCACAGATCTTCAATTTGCAATCAGCAT

TAGCTAGAATGTTCTTCGGTTTCAAATCCCGATGAAACACATTAGctgaaaacaga  
gaaccctaagtgttcaatattatctataacaaaaatcatgaggatcgaaagagaccaatgtgtaaaccagaaactaacCTG  
CGTGAACATATTTTCAGACCACGAAGAAGCTGGTACAAGAAAACTGATAAT  
GCTCAGGAGTCAAATCATCATTCGCCTTGATCACTTGATGAAGATCAGATTC  
CATCAACTCAAAGACAACATAAATATCGCGGAATTCTCTGCGAGAAGGAGG  
TAACATGATGTGTTTAATCTCCACAACATCAGGATGACGAAGCAACCGAAG  
CAATTTGATCTCTCTGAGAATCCTAGTAGCATCGGAGACATGCTCAAACACA  
TCGTTGATCTTCTTGATCGCAACTCTCTCTCCAGTGTGTGAGTCAACAGCAGA  
AGCCACAACACCATAACTTCCTTTGCCAACAACCTCTTGAATCTGATACCTG  
TTAGCTTCTCCATACTCCGTGAAGAATTCAGTTTCACCctgaagaattttattcatgattcata  
cttaaaattgacaaaatgatggatcaagacaataagatcaatcaaacgaaacacaaaactgaaactttatcctttgtggaagca  
aatgcaataacacagaacattaagagaacaaaacaagtctctcaatctcaatcaaaagacaacatctaagagcaaatagcata  
aaagggtagagaaggagaagaagaaggtacCTTCTTGTGAGGATCCATAGGTAAATGATTA  
CGTTTAGGAACGTTAATAAGCGTTAGACCAGAGAAATCGAGATCTTCTTCAA  
TAATAAGCTCTCCAGAGTTAGCGGAATGATTGGGGTTAGAGGAAGAAGAGA  
AAGAAGAAGAATTGAAAATGGGTTGTTCGTGATTATTATTATTATTGTTGGA  
AGAAGAAGAAGAAGAAGAAGAAGGTCGCTGAAAAAGCCATCGACGAA  
CACCGTCGACGAGATTCCCACCACCACC**CAT**

AT1G66340.1

ATGGAAGTCTGCAATTGTATTGAACCGCAATGGCCAGCGGATGAATTGTTAA  
TGAAATACCAATACATCTCCGATTTCTTCATTGCGATTGCGTATTTTTCGATT  
CCTCTTGAGTTGATTTACTTTGTGAAGAAATCAGCCGTGTTTCCGTATAGATG  
GGTACTTGTTCAGTTTGGTGCTTTTATCGTTCTTTGTGGAGCAACTCATCTTA  
TTAACTTATGGACTTTCACTACGCATTCGAGAACCGTGGCGCTTGTGATGAC  
TACCGCGAAGGTGTTAACCGCTGTTGTCTCGTGTGCTACTGCGTTGATGCTTG  
TTCATATTATTCCTGATCTTTTGAGTGTTAAGACTCGGGAGCTTTTCTTGAAA  
AATAAAGCTGCTGAGCTCGATAGAGAAATGGGATTGATTCGAACTCAGGAA  
GAAACCGGAAGGCATGTGAGAATGTTGACTCATGAGATTAGAAGCACTTTA  
GATAGACATACTATTTTAAAGACTACACTTGTTGAGCTTGGTAGGACATTAG  
CTTTGGAGGAGTGTGCATTGTGGATGCCTACTAGAACTGGGTTAGAGCTACA  
GCTTTCTTATACACTTCGTCATCAACATCCCGTGGAGTATACGGTTCCTATTC  
AATTACCGGTGATTAACCAAGTGTTTGGTACTAGTAGGGCTGTAAAAATATC  
TCCTAATTCTCCTGTGGCTAGGTTGAGACCTGTTTCTGGGAAATATATGCTAG  
GGGAGGTGGTCGCTGTGAGGGTTCGCTTCTCCACCTTCTAATTTTCAGATT  
AATGACTGGCCTGAGCTTTCAACAAAGAGATATGCTTTGATGGTTTTGATGC  
TTCCTTCAGATAGTGCAAGGCAATGGCATGTCCATGAGTTGGAACCTCGTTGA  
AGTCGTCGCTGATCAGgttttacattgctgagaattctctctttgctatgttcgatcttctataactttctctctt  
attatagGTGGCTGTAGCTCTCTCACATGCTGCGATCCTAGAAGAGTCGATGCGA  
GCTAGGGACCTTCTCATGGAGCAGAATGTTGCTCTTGATCTAGCTAGACGAG  
AAGCAGAAACAGCAATCCGTGCCCGCAATGATTTCTAGCGGTTATGAACC  
ATGAAATGCGAACACCGATGCATGCGATTATTGCACTCTCTTCCTTACTCCA

AGAAACGGA ACTA ACCCCTGAACAAAGACTGATGGTGGAAACAATACTTAA  
 AAGTAGTAACCTTTTGGCAACTTTGATGAATGATGTCTTAGATCTTTCAAGG  
 TTAGAAGATGGAAGTCTTCAACTTGA ACTTGGGACATTCAATCTTCATACAT  
 TATTTAGAGAGgtaactttgaacagctctatgttcataagttataactattgtgtacttgattgtcatattgaatcttggtg  
 cagGTCCTCAATCTGATAAAGCCTATAGCGGTTGTTAAGAAATTACCCATCAC  
 ACTAAATCTTGCACCAGATTTGCCAGAATTTGTTGTTGGGGATGAGAAACGG  
 CTAATGCAGATAATATTAATATAGTTGGTAATGCTGTGAAATTCTCCAAAC  
 AAGGTAGTATCTCCGTAAACCGCTCTTGTCAACCAAGTCAGACACACGAGCTGC  
 TGACTTTTTTGTCTGCTGCAACTGGGAGTCATTTCTACTTGAGAGTGAAGgttatta  
 tcttgatcttgggatcttataccatagctgaaagtatttcttaggtcttaattttgatgattattcaaatatagGTAAAAGACT  
 CTGGAGCAGGAATAAATCCTCAAGACATTCCAAAGATTTTCACTAAATTTGC  
 TCAAACACAATCTTTAGCGACGAGAAGCTCGGGTGGTAGTGGGCTTGGCCTC  
 GCCATCTCCAAGAGgtttgagccttataaaagacgtttttccaacttttctgtctctgtgtgtaaaagtactca  
 taagcgttaatatgacaagGTTTGTGAATCTGATGGAGGGTAACATTTGGATTGAGAGC  
 GATGGTCTTGGAAAAGGATGCACGGCTATCTTTGATGTTAAACTTGGGATCT  
 CAGAACGTTCAAACGAATCTAAACAGTCGGGCATACCGAAAGTTCCAGCCA  
 TTCCCCGACATTCAAATTTCACTGGACTTAAGGTTCTTGTCATGGATGAGAA  
 CGGgtagtataagcttctcaccttctcttgcaaaatctctcgccttacttctgcaaatgcagatattggcgtttagaaaaaacg  
 caaatthaatcttatgaaaaccgatgattatttgggtgcagGGTAAGTAGAATGGTGACGAAGGGAC  
 TTCTTGACACCTTGGGTGCGAAGTGACCACGGTGAGTTCAAACGAGGAGTG  
 TCTCCGAGTTGTGTCCCATGAGCACAAAGTGGTCTTCATGGACGTGTGCATG  
 CCCGGGGTCGAAA ACTACCAAATCGCTCTCCGTATTCACGAGAAATTCACAA

AACAACGCCACCAACGGCCACTACTTGTGGCACTCAGTGGTAACACTGACA  
 AATCCACAAAAGAGAAATGCATGAGCTTTGGTCTAGACGGTGTGTTGCTCAA  
 ACCCGTATCACTAGACAACATAAGAGATGTTCTGTCTGATCTTCTCGAGCCC  
 CGGGTACTGTACGAGGGCATG**TA**

AT4G15910.1

**ATG**GCCGCTCGTTCACTCTCCGGTGCCGTTAAATCTCTTTGCTCCGCCGCATC  
 CGGAAGTCTGTCTTGTTCATTGTCTTAAGgtacttatgttgaactgatccatttttcaggtacaat  
 atataaacattacacactctttcattgcttagatcgataaagaattataattctaattgaataatagttgatggttgaattgagtg  
 aaggctaaagattgatctatcattaataaacgtaattaattacttagatctattaagataattatgcatgatattatcacgtattttct  
 acgtttcgttatgttatttagtcattgctgtttatataatcttacctcaaataaattattggaatgtag**GAGGAGTTACGT**  
 TGCTACATCGCAGAATGTAACAGCAGCAGGATTGAGTAAGGGAGGTTCCAC  
 CAGAGTTATGGTGGGGAAGATGGAACAGAG**AGGTTTAGATCAAGAGCCAGA**  
**GTC**TGCTTGGGGACCAGATCCAGTTACTGGATA**CTACAGACCTTCCAATCGT**  
**CC**GGCTGAGATTGATCCAGCTGAGCTCAGAGAGTTGCTTTTGAAAAACAAA  
 GCAAAGTCTTTC**TG**

AT2G31660.1

**ATG**GATCTGCATAGCCTCGCTTTGATCCTCCGAACCGCCGCTCTTAGCCCTAT  
 TCCCGATGAGCGTAAAGTTTCTGAGCAGCAACTCAATCAGgtttctcagttccacttcaat  
 ttctcttcgatgtatttacagggttgcttggtcgcgagtttgattgttcttattctgctagtagtatcttaaagtttgagctgagttcatgt

atagagtttcctataagttcactttctggtttactgaagatagttcaaagattcccactcttttgaacaaatcgttattgctttaatc  
actgtagccagatttcagtggtgttactatttctaataatgatcaattaaagttatgcagtttctgatgcctgagtttctcgttattctttgtc  
ttgcagTTGGAGCACACGCCCCAGCATTGGTGAGGTTATTGCAGATAGCCGTG  
GATGGAAATTGTGACATGGCAGTGCCTCAAATTGCTAGTATTCAGTTCAAAA  
ATTTAATTGCTAAGAACTGGTCACCTGAGGATTGTGgtaataatttctatttcaattatcttctc  
cggcttagcgttgaagatcttagattaatttgcattttgatgggtgcagGTCCTGCAGTAAGGCAGCAGC  
AGATATTTGAGAGTGACAAAGAATTGGTGAGGGACAATATTCTCGTCTATGT  
CACCCAAGTTCCAACCTTACTCAGgtaaattgccagatagctttggaattatcattcaagtcatttggttg  
ctgaactctgtctgacgtggtctgatttaagcgattttatgaaaatattgtaacctgatggctaatacgcctacaaattgctc  
atcttctgatagattgcttcaagctgatagctctcttttatatacatcttgaatgtctctacattatccttaggttctagata  
ttaaaaaacggcagatagaagatgatttttttaagttagttttttctcattaccttggccaatagcatcttataagcctcaac  
aacaaccataattttttccctaatttcttaccagATCACAACTTGGAGAGTCTCTCAAGACAAT  
TATTTATGCTGACTACCCAGAACAATGGCCACGTCTTCTGGATTGGGTGAAG  
TACAACTTGCAAAATCAACAAATTTATGGGGCTTTGTTTGTGTTGCGGATAC  
TCTCTAGAAAATATGAgtaagttcaccactttctgtccgatccctatgtctgtgtgatgattaaagttttagt  
gaaattttgatatgcgttattgagacatctcaagattatcttgggtgcaatagecgaatttcgagcttgcattatcacact  
gagtagtgactaataaaattaggtgtaaccgataatattaagagaagggaagaagtgttaggcaggagaaatgagagtgtt  
aaaacgtattgtagaacacaatggtttcgtgtctctgcaattcatgatgtcctgtcaagtatgtgacgacttccagctagaatg  
caagttgctctttatgatgacacctaagtagtatatccggtgatgatgtttcatgattttatttctagGTTCAAGTCAG  
ATGAGGAGAGGACACCGGTTTCCCGCATTGTGGAGGAGACATTCCTCAACT  
TTTGACTATTTTTAATGGGCTTATCCAGATACCAAATCCGTCTCTGGAGATAG  
CAGAACTTATGAAGTTGATATGCAAAATATTTTGGTCATCCATATATgtaagttgc

tagagagatcagttacatatgcctctgcggaatttcacattttatctttcttaacataatgcatatccttaaacacaattcaaacca  
gttaagctgttaatgtctttttactgttctctgtgggcagCTGGAGCTTCCAAGGCAATTATTTGATCT  
AAACGTCTTTAATGCCTGGATGGTTCTGTTCTTGAGTGTTTCGGAACGTCCTG  
TTCCAGTTGAAGGTCAGCCTATGGATCCGGAACCTTAGAAAATCTTGGGGCTG  
GTGGAAGGTCAAGAAGTGGACAGTGCACATTTTAAACAGGCTCTATAGTCGg  
tatgtcttgacattaatatttttaacaagaaagtagcgttgaaaggttatagaatattctttaacagGTTTGGAGATCC  
AAAACCTCAAAGTCCAGAAAACAAACCTTTCGCCCAAATGTTTCAAAGAA  
TTATGCAGGCAGAATTTTGGGAAGGCCACCTAAATTTCTTGAACACAATTCGT  
GTTGGAGGCTATCTTCCTGACAGAGTTATCAATCTTCTCCTTCAGTACTTAAG  
CAACAGgtttgtgtacaatcactgttagacattagttcatgtacatacatgtacacattcagatccccactggctaactaa  
glatcatctttctctgacattataaaaatttttttggtcactgaatttaagtcttgaattaattagaaaactcttagaggtttagatgt  
tttgtaaaaaaaattcgttgaaactaattcgtgtttctgcagCATTTCGAAGAATAGCATGTACAAGTT  
GCTTCTGCCTCGGCTGGATGTTCTGCTTTTTGAGATTGTTTTCCCTCTAATGT  
GCTTTAATGATAATGATCAAAGCTTTGGGAGGAAGACCCTCATGAATATGT  
GAGGAAAGGTTACAgttaagtactacataataatgactttcttaaaaagcttaattcttaaaatcatcatttgaatgtta  
tttgtgtgctgtagATATCATTGAAGACTTGTACAGTCCGCGAACAGCGTCCATGGAT  
TTTGTAATGAGTTGGTTCGAAAACGTGGGAAAGAAAACCTTCCGAAGTTTG  
TTAAGTTTGTCGTTGAGATCTTCTTGAGgtattacattgagctcttctgcttgaactttcatatcagtt  
gtgtttattgacaactgtttgtccactgtttgtagTTATGAAAAAGCTACTGTAGAGGAGAAGCCT  
TATCGCCAAAAGATGGTGCGATGCTGGCTGTTGGAGCTCTTTGTGACAAAT  
TGAAGCAAACGGATCCCTATAAATCTCAACTAGAGCTAATGTTGGTGCAACA  
TATTTTTCCTGATTTCAATAGTCCAGTTGGACATCTTAGAGCAAAGgtatttttgctt

taatctgtgtgtggatatctcagtaatccattgtfactagcttaccggaatgaggatthtthtgcattccacgaagtctgtgtactt  
 caatttattctagaacgcttagcattgttgcttagttggctactgcattgctgctgatagtttagtatagacttcccgttgta  
 cccagtatttgatgcatattctttgtgatctcctttgatctgttgacgggtgcattttttggatgtagGCTGCATGGGTA  
 GCTGGGCAATATGCACACATCAATTTCTCAGATCAGAACAACCTTCGTAAAG  
 CATTGCACAGTGTGTTTTCGGGACTGCGCGATCCTGATCTCCCTGTTTCGTGTT  
 GATTCAGTTTTTGCATTGCGTTCATTTGTTGAGGCGTGCAAGGgtatgtgattccgcct  
 acgcattttatatgctgaatgtgttgcactgtttacctgtgtgtggctttttgattgtgtgctgggtaatgtgtggctaataaa  
 actttctggcaaatgcagATTTAAATGAGATACGTCCAATCCTCCCTCAGTTACTTGAT  
 Ggtgggttactctctttatgaatactagatggcagattgttgaagtaaactaacgctcaaacttattgtgatcctgcgtccatca  
 gAATTTTTCAAACATGAATGAGGTAGAGAATGAGGACCTTGTTTTTACTTT  
 GGAGACCATCGTAGATAAGTTTGGCGAAGAGATGGCTCCTTTTGCTTTTGA  
 TTATGTCAAATCTGgtaagaaaaattggttctctgagttgatattcttgaatactcaaacttgtaataaatatgat  
 ttctgcctttgctttgttcagGCGGCTGCGTTTTGGAGATGTCTAAACACGTCAGAAGCCA  
 ATGACGACTCTGATGACATGGGAGCTTTAGCTGCCGTTGGTTGTTTGCCTGC  
 CATAAGTACAATCCTTGAATCTGTTAGCAGTCTCCCGCAGCTGTTTGTGAG  
 ATAGAACCAACTATACTTCCAATAATGCAGAAAATGTTGACCACTGATGGCC  
 AAGgtatactcatatagttttggatagttttagactccgagataaattaagttatagttgtgtccagaggatgtacattgga  
 ctgattggtagccaactagccatttgggtccattacattagtatggtttttttcagaattgtataatccattttttgtctgtcagcct  
 tgtgtgtgtgtgtagcatcacgatcttaagtttattaattccttttctagctatttctatttaataattgacactgaagcatctatatttcat  
 cctatttgcattggcagAGGTATTTGAAGAAGTTTTGGAGATTGCATCATAACATGACCT  
 TTTATTCACCTAGCATTTCCTTGGATATATGGAGTCTCTGGCCTTTAATGGTG  
 GAAGCATTGGTTGATTGGGGAATCGATTTCTTTCCAAgtatggttctctcacatacctctaaa

ttttctagactttggattttatagtgactccatcacggtagcactccacgtcattcattttttgcatacgttcactcaggtfactcttt  
gaaatttagATATTTTGGTTCCAATGGACAACTTTATATCAAGGGGAACGGCTCA  
TTTTCTCACTTGCAAAGAGCCCGACTATCAGCAAAGTCTATATAACGTTCTTT  
CAACTgtgagatatactctgatagtgctagtgagtagatgtctgtgcttgattggtgtaaggaactgaaggtataaaatt  
ctgcaaatacatgtagcatagaatcttctttatcgtttaaaaaattcgagttcattttacatgcctattgtggtaaaaagttgtg  
gaactctgtttattgtgaaaacactaggaatactatatgttacattcagactacatgccattttttgctaataatataataagcatctctt  
atgtttttccgcacaaggatagcttatatatgtctcacgtttccactctacagCTTATGACCGACAGAAACATA  
GAAGACAGTGAAATTGAGTCTGCTCCAAAGCTTATTGAAGTTGTTTTCCAGA  
ATTGCAAAGGGCAGGTGGATCAGTGGGTTGAACCATATCTGCGTCTCACTGT  
TGACCGGTTACAACGGGCTGAGACTTCGTATGTTAAATCTCTTCTTATACAA  
GTGgtaagaactatgtggaggattgtgacctgctactggtgatttctactaccggttatgctcttaccattagaggagatag  
tatgtcaatgctactgtatgaatcaagtgggtcctattgaactgattctgttgatgaccacaatgatgggtgaaattccaagt  
tttttaatagcaacagcaccagcgttatataccatgaagttctcaacataatacttgattctgttcgcagGT  
GGCAAATATGCTTTACTACAATCCAGGTTTGACGCTTGGTGTATTGCATAAC  
ACAGGACTTGCTTCCAAAGTCTTTGACCTTTGGTTCCAGATGTTGCAGCAA  
AGAGAAAAAGTGGCCTGCCAGCAAACCTTCAAAAAGgtactgtgatgtttttgtatgtcacatct  
tgttcatgtaactataagcgaatggttgaaattctattgttcgctttgatagctgataaaggggtgttactctctgttaccagG  
GAACATGATAAAAAGGTTTGCTGCTTGGGTTTGACTTCATTACTTGCTCTCCC  
TGGGGGTCAATTCCCCGATGAAGCGCTGCAGCGTGTTTTTCAGGGCAACACTT  
GATCTTCTAGTTGCATATAAGAATCAGCTAGCTGgtatgtcctgctcatgagtcgatccatttt  
ctctcacttgcatgtaatctttacatttattccctttccaatcttctcgcagAAGCAGCTAAGGAGACAGAA  
GTAGATTATGAAGAAGAAATGAATGGACTCCAGAGTAGTGATGATGATTAT

GATGATGATGGGTCTGATGGGGAGATGGATGATACAGAGGAAGGAGATGAA  
GCACAAAGTGTTAAACTGCAGAAGTTAGCTGCACA Ggtttgtgaagtgtttgatcgacttgc  
tctgcaagcatggttttttctgcttggactgtttccaattcgttatcctttcatagGCAAAGGCCTTCCACT  
ATGATGATGATGATGACGACGATTCTGATGATGACTTTAGTGATGAGGACGA  
GTTTCAGTCACCCATCGATGAGGTGGATGCCTTTGTATTCTTTGTTCGATGCAA  
TCAGAGgtaccattaataaaaaatcctatggttaccagtttgaccacataaaatactatatggagcttcatgttctctggcgta  
tcttgcttagtaatcacgcttctgtttattatacagTTATGCAGGCATCAGATGCG  
CAGAGGTTTCAGAACCTGAACCAGTCACTGGACTTCACTTACCAGGCGATTG  
CAAACGGAATAGCACAGCATGCAGAGCTGAGAAGAGTGGAGATCGAGAAG  
GAGAAGCAAAGAAGTTAGCAGCTGCTTCTACTCCTGTTACTGCTCTATGA

AT4G29130.1

TTAAGAGTCTTCAAGGTAGAGAGAGTGAGAAGCAGCAAGAAGCGCAGCTCC  
AATGCCTGATCCATCATTGGAGTGAGTGACTTCAACGCTTCCCTGAAGCTTCA  
TCTCCAAGCAACTCTTTTAGTGAGCTCTCCATACACTCACTAAACTGAGTGT  
AATGCTCAAACAATCCACCATCCATGGCTATAACCGATTCTGCACCTCCTC  
GTCTTTAGTAGTATCTCTTCCAGTTTCTTCAGAATACCATAGATTCCAGCAG  
CAGAGAGACGAGCTCCTCGGGTTGCTATGATGTTGCAGAGACTGATCACAA  
CTTTTCTCATTTCAGAGAAGTTGTAGGGACCTCCAATATATCCTTAATCTTG  
CTCCCAACAATCTTCAAGTCTGGAGAAGTGTCGTTGTGCATAGCCGACATGT  
GAGGAGTCctgcaaatgattgtgaactttagactaaagccctcctcaaaggatgaaacttagtaaagagtacCTAA

TGATGAATGGTATTCTCAGCTTAGATGGGACTGTATCGCCAAAGAAAGCAGC  
ATCTTCAGCCATCTTTAGAAGAACTCTTCGCAAAATCTCTCCCAAGTACATA  
CCGGAAATGATTTTCTCAAGAATctgcaaaatgaaattccacatagaaactcattctcgagacacctgta  
ggtattggcattttgtaacaacaaccggggtttaaaaactgtacCTGTTTCGCCTGGATTTCAGACTCTCG  
AAATCCAGCGTGTGATCAAACCTCGGTTAATGGAAGATGTGATGACCTGAAG  
TTTCCCATTCATGTTTATAACCATTTCTCCTGATTTTGGGAAGCAGACCATG  
CCATTTAGGGATCGCGGTTGCACGCTCAACATAGGCTGCGTTTGTCCCAGTG  
CCTAAAATAACAGCAGCAACAACATCCGGGTTATAGTATCTACCACCGGCTA  
GTGTTCCAACGGTATCATTAAACctgtaaaacgtcaaaaacaatcaacaacacatgacaaaagggatc  
actagagagcataaggcgaaagagaactcacAAGTGCTGCGATTTCGCATGTCAAGACCAACTC  
TTTCCAGAGCCTTATTAAGTGCTCCAACAACATCTTGTCCAACctatgcaacaggttaa  
acaagattaaggcaaaggcgaaaactaacattttaagtgaatagtgaacgaaaatcgagcttaacTGCTTCTTCG  
ATGGAAAAGCCTTTTGTCCATTTGATGAGACTACCAGAGGACAGAGAAGTCT  
GCTTAACAGGAAACGAGAAAGTGAAACCTAATTCCCTCTGTCTACCTTCTGG  
AAGATGAAAGTCTTCGCATTCTGTAGCGACAAACTTCGCAAGAGCTTCAGCT  
ATAAAATTGAACAACCTCctgaatccatcaaaagaacgaaaagaatcttggtacacaaagatcatctaagacta  
aacatgacaaattgaagaacagaaaactgacATCTGAACCACCAGTCATCAAATGAGGAGGAA  
TCGAAACTTCTTCGAATTCTTGTTTAACAACACGCTCTTGCTTCCCGCCAAGA  
AGCACACGCATGACACGGAAGTTTGTCCCCCTAGGTCCAATGCATAAAAG  
AGACCCTTTTCATCCctgttatattaatggtagtaaatgtaagacagacctacaacatgaaactagacgcattgccg  
cattggtaacaatattaaagcgaaaatgaaaccgacccaaaatgttctacgtaataataacataatcaatgccatcatgccatcta  
atcacgaagtccataagacaaaagtgatgtctgtggacagacttttcagacaaatcatattcacatcactgtcattgtctgcctta

gtaactatgtattaaaccttaatacaaaagggcaataactctcaaatccatgaaaacaaatctacaaaatcaagaagtatacaaca  
 aaacctataagaaaatgagacaatgattcagtgaggacatgatctggattccaaaagattcactgattctgaacagtagatcaata  
 atcctatctaategatcaatgaccattaagagacaatgaattacatctccaaatcaggacgaaactcattcataacgtaatcc  
 atattacacaatcaacaacattgaatggatgacatattaggagatagaatattctatcttctaataatgacaagattgaaaattca  
 atgaaaagcaacagaattgaaatcaattagaaaacattac**CCGGAAGGAAGATTATCAACGTAGCT**  
**GATAAGCATCTTGAGTTTGCTACCACCGTCGGATGCAAGACCAGCATGCATC**  
**TCAACGGTCATAGCATCAGCCACTTGTCTCAGTTTCGAGATCGGAGTCGCAC**  
**AATCCTCTTCAAAGGCCTTGAGGATAGCCAAAACACGTCCCCACTTCCCTGA**  
**GCTCTGCATCCGTCGTCGAACAACCAAAACAGCCACCGCACAAACCGCCGC**  
**CGTGCAAACAACAGTCGCTCCAACAGCTACTTTACC****CA**

AT1G01480.1

**ATG**GGTCTTCCGGGAAAAAATAAAGGTGCAGTTTTGTCGAAGATAGCGACT  
 AACAATCAACACGGAGAGAACTCAGAGTACTTTGATGGATGGAAAGCTTAC  
 GACAAAGATCCTTTTCATCTTTCCCGTAACCCCATGGGATCATCCAAATGG  
 GTCTTGCAAGAGAATCAGgtacagattatatataatccaataaatcatgttatatgtgtgtcgttgatgaacttc  
 catctattagctattatatatgaacacgtatacacatcaagctaataacctttttctctttcaagtcaagtaataacttaataacaca  
 ttttctaaactcttacagCTTTGCTTAGATTTGATCAAAGATTGGGTCAAAGAGAACCC  
 AGAAGCTTCTATTTGCACCCTTGAAGGTATTCATCAGTTTAGCGACATCGCT  
 AATTTCCAAGACTACCATGGTCTTAAGAAGTTTAGACAGgtactataaatcattcattattca  
 gatattctgtaatcagctacggacatattagaaaacaattttfacatggaaagtaataacaccttaacaacatcagttgatgat



GTTGCACGGAACCTGGATGGTTTAGGATTTGCTTTGCCAACATGGACGATGA  
 TACTCTCCATGTGGCGCTTGGACGGATCCAAGATTTTCGTGTCTAAGAACAAG  
 AACAAGATCGTCGAGAAAGCATCTGAAAATGATCAGGTAATCCAGAACAAG  
 AGTGCTAAAAAGCTGAAATGGACGCAGACCAATCTTCGACTAAGTTTCCGA  
 CGACTTTACGAGGATGGTCTCTCGTCTCCAGGGATAATGTCACCACACTCAC  
 CTCTTCTCCGAGCA**TGA**

AT1G43160.1

**ATG**GTGTCTATGCTGACTAATGTTGTCTCTGGTGAGACCGAACCCCTCGGCAT  
 CTGCG**ACATGGACGATGGGTCATAAG**AGAGAAAGAGAAGAGTTTTCTTTGC  
 CTCCTCAA**CCATTGATTACCGGTTTCAGC**TGgtatgcatgttttctaaatgaattttttgtgagaa  
 cactatactaaccaataaaacattctttaacaaaaacaaaatcaataaaacagTGACTAAAGAATGTGAAA  
 GCTCAATGTCCTTGGAGAGGCCAAAAAATATAGAGGAGTAAGGCAACGAC  
 CATGGGGAAAATGGGCGGCGGAGATTCGAGACCCACACAAGGCGACACGTG  
 TATGGCTTGGGACATTCGAGACAGCCGAGGCCGCCGCAAGAGCCTATGATG  
 CGGCAGCACTTCGCTTTAGAGGAAGCAAAGCAAAGCTTAATTTCCCGAAA  
 ATGTTGGAACCTCAGACGATTCAACGAAATTCTCATTCTTGCAAACTCTAT  
 GCAACCTTCTCTGACATACATCGATCAATGTCCAACCTCTATTATCTTACTCTC  
 GATGTATGGAGCAACAACAACCATTAGTAGGCATGTTGCAGCCAACAGAAG  
 AGGAAAATCACTTTTTTCGAAAAACCATGGACCGAATATGATCAATACAATTA  
 CTCCTCTTTTGGT**TAA**

AT5G44200.1

TTAAGATCTTCTCTTCCGATCATCTTCACCATCGTCATCGGAGTCACCGCTCT  
CACGGAATCGTGGGTCTTCTCctgttacaatcacacaaaattccatagcagacagtgtttagcatcagata  
ctaaatatactcgaaacgtttagtaaggaaaaagaaaaggccttgatttacCGGTCTCATGTCACTATCAG  
TGTCTCTCCTAGACTCATGATCTGTGTTTCTTCTTGAGTTATCGCGGCCATAG  
CGATCATCGTCTCTTTGTCGTTTCCGATGGTAGTctgcaaaaacattaaaaacagtgattgtgc  
tattgtggctgaggattacactattactacgcattatgtcattgggataaaaatacCTCCTCTTCCATGGCGATT  
TTGTCCTCCTTGACCATAGTTTCCACCACCACGCCTTCCATTCCATctgtgattaac  
atatgcaaaaagcatttaggaatctagtgtttgcaggctctattfacggatttcacataattgactgaataagataactaaca  
ctaggaccatattgtttcacatttcactatctattcgaagggaactgggaaaacaataagcagatgctttgcaagataaagag  
tctfattctgagcactacaattctaagtaagcaacttaagactcgcattcatggtacttttagftcaccattgtccacagtatagtc  
gaaatcatgaaaccaggaataagatcaagtaaatggtagaaagtgaaaaaacatacAATTCGTTGGCGCAG  
CTTGAGGATAAGCTCCCAATGAGCCAGTACCGTAATCCACGAGCTGCCTTTG  
TGCTTCGAGTTCCTTCTGAACTAATTTCCCATAACCACCTCTAGgtaatcaatgagta  
aggaaaaaacatggattggcgggaaattcataaccgagaatcaatcttctaccacaaaacaggataactaacaagaac  
aacagagtgcataactatcggactaaaaggatatCAGGATCGTAATCTGTACGGTATTCATCAC  
GAACctgtatgaaattagaaaaccatgatgacctttaattgtaaagaacggtcagggaattacagagaggatgatgcatct  
gatgaatacCTGGCCACCGCTTCTACCACGGCCCCATTGTCTTCCTTCTTGGAATC

CCCAATCAAAATCCACACGAATAGGCCGGTCATCGAGAATAGTCCCACTTAT  
 AACTTGACTGCATCTTCAGTATCCTCTCTAGAGTAGAACctgataaaaacaagacaca  
 aacaaaaactcagcaaccaccataatgaaacttacatataaaaactgtgaagatgaaagcttacAGGACGAAACAA  
 AAGCCACAAGGTGTTTTAGTGTTCTTATCCAAACCCATGATGATCTTTTTGAT  
 TTCTCCAGCTCTAGAGAAAAGCTCGTAAAGCTGTTCTTCAGTAGTGTA  
 GACACATTGCCAATGTAAACCGTTGTAGATGCTCTTAGCGCTTCATCAA  
 CCTCTTGTTCTGAAAATCTTCTATCACGATAAGCTGATAGCTTTGCTTGC  
 TCctgacaatcaaaaaccccaaacatgaaatcgtgagaccaaacctagacaacaactcaaaaatcaattagaatcgttca  
 gaccatcagattccagacaacccggatactaaatcactaaaaacattaacgaatctcgattttctatgtagaaaaagcaaac  
 ccgaaatggttccgataaaaatctaaactttatgtacCTTGAACAAAGAAGCCAT

AT4G05320.2

ATGCAGATCTTTGTTAAGACTCTCACCGGAAAGACAATCACCTCGAGGTGG  
 AAAGCTCCGACACCATCGACAACGTTAAGGCCAAGATCCAGGATAAGGAGG  
 GCATTCCTCCGGATCAGCAGAGGCTTATTTTCGCCGGCAAGCAGCTAGAGGA  
 TGGCCGTACGTTGGCTGATTACAATATCCAGAAGGAATCCACCCTCCACTTG  
 GTCCTCAGGCTCCGTGGTGGTATGCAGATTTTCGTTAAAACCCTAACGGGAA  
 AGACGATTACTCTTGAGGTGGAGAGTTCTGACACCATCGACAACGTCAAGG  
 CCAAGATCCAAGACAAAGAGGGTATTCTCCGGACCAGCAGAGGCTGATCT  
 TCGCCGGAAAGCAGTTGGAGGATGGCAGAACTCTTGCTGACTACAATATCC  
 AGAAGGAGTCCACCCTTCATCTTGTTCTCAGGCTCCGTGGTGGTATGCAGAT

TTTCGTTAAGACGTTGACTGGGAAAACACTATCACTTTGGAGGTGGAGAGTTCT  
 GACACCATTGATAACGTGAAAGCCAAGATCCAAGACAAAGAGGGTATTCT  
 CCGGACCAGCAGAGATTGATCTTCGCCGGAAAACAACCTTGAAGATGGCAGA  
 ACTTTGGCCGACTACAACATTCAGAAGGAGTCCACACTCCACTTGGTCTTGC  
 GTCTGCGTGGAGGTATGCAGATCTTCGTGAAGACTCTCACCGGAAAGACCAT  
 CACTTTGGAGGTGGAGAGTTCTGACACCATTGATAACGTGAAAGCCAAGAT  
 CCAGGACAAAGAGGGTATCCCACCGGACCAGCAGAGATTGATCTTCGCCGG  
 AAAGCAACTTGAAGATGGAAGAACTTTGGCTGACTACAACATTCAGAAGGA  
 GTCCACACTTCACTTGGTCTTTCGTCTGCGTGGAGGTATGCAGATCTTCGTG  
 AAGACTCTCACCGGAAAGACTATCACTTTGGAGGTAGAGAGCTCTGACACC  
 ATTGACAACGTGAAAGCCAAGATCCAGGATAAGGAAGGAATCCCTCCGGAC  
 CAGCAGAGGTTGATCTTTGCCGGAAAACAATTGGAGGATGGTTCGTACTTTGG  
 CGGATTACAACATCCAGAAGGAGTCGACCCTTCACTTGGTGTTGCGTCTGCG  
 TGGAGGTATGCAGATCTTCGTCAAGACTTTGACCGGAAAGACCATCACCTT  
 GAAGTGGAAAGCTCCGACACCATTGACAACGTCAAGGCCAAGATCCAGGAC  
 AAGGAAGGTATTCCTCCGGACCAGCAGCGTCTCATCTTCGCTGGAAAGCAGC  
 TTGAGGATGGACGTACTTTGGCCGACTACAACATCCAGAAGGAGTCTACTCT  
 TCACTTGGTCCTGCGTCTTCGTGGTGGTTTC **TAA**atctcgtctctggtatgcttaa **gaagttcaatg**  
**ttcgtttcaatg**taaaactttggtggtttgt **ttttggggcctt**glataatccctgatgaataagtgttctactatgttccgttctgttat  
 ctctttcttctaatgacaagtcgaactcttctttatcatcgcttcggttttattatctgtgcttctttgttaatacgcctgcaaagtgact  
 cgactctgtttagtgagttctgcgaaacttgtaaatagtccaattgttggcctctagtaatagatgtagcgaaagtggtgagctgtt  
 gggttctaaggatggcttgaacatgtaatcttttggttctgagatgatgaacattcgttgtt **gtaagaatgcctgtaatg**ccca

caaatgtagaaaatggttcgtaccttgcctcaagcattgatatgtctgatgagaggaaactgcaagatactgagcttggttaacg  
aaggagaggcagttctcctccaagcattcattgacaatgccttgatcatcttaagtagagttctgttggaagttgaaa  
cttgaagaaacgactctcaagtaaattgatgatcacaagtgaagtgtatgtt**acataagtgatattcaccctttccatcaatc**  
**aaaacatcatatagtaatccattggttatacaaacatcaaaatacttacctctgaaatgaggaaaaaatgcaaagagattttg**  
**aaaattccaacaatggagagaaaaaaaaaaaaaacaagaattaaatcaatcagtacattataaaggttatgagacctatgat**  
**acTTA**