Potential of Allele Mining for Improving Drought Tolerance in Crops

Akash Sinha, Ankita Chauhan and Pushpa Lohani*

Department of Molecular Biology and Genetic Engineering, College of Basic Science and Humanities, GB Pant University of Agriculture & Technology, Pantnagar- India

*Corresponding author

A B S T R A C T

Keywords

Allele mining, Crop improvement, Abiotic stress, Drought tolerance, Germplasm collection

Article Info

Accepted:
10 April 2020
Available Online:
10 May 2020

Drought is the major abiotic stress that results in severe loss of yield to crops. It is estimated that there will be a steep rise in global water consumption in the coming years. On the other hand, it is also estimated that the sources of water will deplete due to rise in temperature and climate change. It is, therefore, critical to find out such genotypes of crops that have the ability to tolerate drought without much loss of yield. The genetic and molecular basis of drought tolerance has been investigated extensively and genes encoding drought-related transcription factors and functional proteins have been identified by allele mining. Allele mining is a promising way to isolate naturally occurring variation in alleles of individual genes with useful agronomic qualities. The superior alleles of such genes need to be fished out. Germplasm collections worldwide have immense unexploited allelic variations in genes. Deciphering untapped useful nucleotide diversity patterns for drought-related genes can be performed by allele mining. The recent advancements made in the field of next generation sequencing have made the approach of allele mining less cumbersome, practicable and cheaper. This review explores the concept, potential and applications of allele mining for drought tolerant genes and its importance in strengthening the goal of achieving climate resilient agriculture.

Introduction

Drought can be defined as deficiency or absence of precipitation for a long period of time eg a year or many years in a region compared to statistical multi-year average rainfall for that region. It results in shortage of water for numerous activities like agriculture and environment sector. Drought is the consequence of anticipated natural precipitation reduction over an extended period of time, usually a season or more in length.

There are many definitions proposed around the world to classify drought in terms of reduced rainfall over different time periods, its impact on water reservoir levels as well as reduction in agricultural productivity. FAO classifies drought according to

meteorological, agricultural, hydrological, and socio-economic criteria. However, an agricultural drought is said to occur when there is insufficient soil moisture to meet the needs of a particular crop at a particular time. Droughts are very devastating of all natural hazards as their occurrence and duration is uncertain. In addition, droughts can subsequently lead to other hazards, such as extreme heat and wildfires. Their impact on wildlife and farming areas is enormous, often killing crops, grazing lands, edible plants and even in severe cases, trees. A terrifying consequence of drought is wildfire as the dyeing and drying vegetation catches fire easily. Thus, high temperature combined with drought poses a very serious situation.

Droughts bring with them prolong periods of inadequate water supplies leading to a sharp decline in agriculture produce. The decreased agricultural productivity is reflected as incidences of malnutrition, famine etc. leading to ill health and death of many people. Droughts' duration and their intensity have generally increased over the years. Direct impacts of drought include reduced crop, rangeland, and forest productivity, reduced water levels, increased fire hazard, damage to wildlife and fish habitat, increased livestock and wildlife mortality rates, increase in rate of insect infestations, increase in reports of plant diseases etc. Indirect impacts include reduced income for farmers and agribusiness, risk of foreclosures on bank loans to farmers and businesses, increased prices for food and timber, increased unemployment, reduced tax revenues, increased crime and insecurity and migration.

The intergovernmental panel on climate change forecasts that the condition is going to exacerbate and the end of this century will witness widespread drought stress in agriculture as a result of drying subtropics as the greenhouse gas concentrations are likely to remain elevated (Solomon *et al.,* 2007). In warm regions, crop yields can drop ~3 – 5% with every 1°C increase in temperature. Agriculture activities alone consume about 75% of the global water. Since dryland populations are mainly concentrated in the developing countries where majority of the population is involved in agriculture or allied activities, planning of suitable mitigation strategies is imperative. Various approaches have been tried to address the problem of drought leading to failure of crops, most of which involve breeding for drought tolerance with marker assisted selection. But drought tolerance is a multigenic quantitative trait involving complex genetic control. It involves huge gene families and complex interactions between the transcription factors and cis-elements on the promoters of target genes (Wang *et al.,* 2009). Also it has low heritability and high G x E interactions. Hence, the approach of marker assisted selection for imparting drought tolerance has not been successful in contributing significantly to crop improvement (Fleury *et al.,* 2010). Another approach is performing manipulation at molecular scale. But this demands intense study about the pathways, gene networks and cross talk between them as they overlap each other in the case of abiotic stress responses. Shinozaki *et al.,* (2007) discovered that about 40% of genes induced by drought or high salinity are also induced by cold stress. Also a risk exists that enhancing tolerance to one stress may also lead to imparting sensitivity to another. For example enhancing production of the osmolyte proline to counter drought stress may prove to be an inappropriate effort in field conditions where multiple stresses co-occur since proline has toxic effect under heat stress (Rizhsky *et al.,* 2004)

So the problem of imparting and enhancing drought tolerance can be overcome by using allele mining techniques which involves the

identification and isolation of novel and superior alleles of agronomically important genes from crop gene pools to suitably deploy for the development of improved cultivars. The natural variations observed among different alleles of genes coding for important traits can be harnessed using allele mining tool and can be utilized in crop improvement programs (Kumar *et al.,* 2010). It is critical to have rich genetic diversity for any crop improvement program as it is a prerequisite in the development of superior recombinants. Accurate assessment of the level and pattern of genetic diversity is of great importance for crop breeding. Genetic diversity analysis is usefulfor estimating and establishing of genetic relationship in germplasm collection, identifying diverse parental combinations to create segregating progenies with maximum genetic variability for further selection and introgression of desirable genes from diverse germplasm into the available genetic base.

Molecular basis of drought tolerance and use in allele mining

Internal cell mechanisms induce certain pathways and gene expression patterns in response to moisture stress by altering the level of specific transcription factors. Microarray gene expression data provides a global view of transcriptional regulation. Identification of significantly regulated target genes which differ in their expression between drought tolerance and drought susceptible genotypes under drought stress might potentially serve as suitable candidate for allele mining.

Using GO analysis of expression profiling of Affimetrix Rice Genome array, Lenka *et al.,* (2011) suggested that drought tolerance of drought tolerant was found to be linked to enhanced enzymatic activity, whereas drought susceptibility of drought susceptible genotypes was governed by significant down

regulation of transcriptional regulatory protein encoding genes. Another method for identification of stress responsive genes in sequenced genotypes is using ESTs generated from drought stressed seedlings. A direct approach for discovering genes associated with stress response was provided by ESTs; Gorantla *et al.,* (2007) in order to identify genes associated with water stress response in rice, performed comparative analysis with public databases and expression profiles and identified 125 putative genes expressed under drought stress.

The stress-inducible genes can be classified into two groups. The first group includes proteins that most probably function in abiotic stress tolerance. The examples of the proteins are chaperones, late embryogenesis abundant (LEA) proteins, osmotin, antifreeze proteins, mRNA-binding proteins, key enzymes for osmolyte biosynthesis, water channel proteins, sugar and proline transporters, detoxification enzymes, and various proteases. The second group comprises of regulatory proteins. Regulatory proteins comprise of various protein kinases, different transcription factors, phosphate hydrolyzing proteins, enzymes catalyzing phospholipid metabolism and many other protein molecules involved in signal transduction pathways such as calmodulin-binding protein etc. Regulatory RNAs including siRNAs and miRNAs have also been discovered as important regulators in drought stress response and tolerance (Shinozaki and Yamaguchi-Shinozaki, 2007). The different categories of genes associated with drought tolerance are compiled in table 1.

The two most important groups of genes that have been widely used to counter drought stress are genes for transcription factors and of osmolyte biosynthesis. The single functional gene approach has seen little success in conferring drought stress tolerance

to plants due to the complexity of stress responses regulated by multi-genes (Mittler *et al.,* 2011 and Varshney *et al.,* 2011). This has lead to more attention on studies of regulatory genes and it was found that transcription factors play role of master regulators in multiple abiotic stress responses by regulating a big spectrum of downstream responsive genes (Wang *et al.,* 2009).Overall view of molecular response of transcription factor genes in drought tolerance is presented in Fig1. The DREB subfamily, the most extensively studied of all transcription factors, can regulate expression of multiple dehydration/cold regulated (RD/COR) genes by interaction with DRE/CRT cis elements (A/GCCGAC) present in the promoters of RD/COR genes which are responsive to dehydration and low temperature stress, such as RD 29A/COR 78 and COR 6.6 (Liu *et al.,* 1998; Lucas *et al.,* 2011). Another important family of transcription factors is the MYB which have been recently well summarized by Li *et al.,* (2015) and its members have been found to be active players in regulating drought related responses. For example AtMYB60 and AtMYB61 improved drought tolerance by regulation of stomatal movement (Liang *et al.,* 2005; Jung *et al.,* 2008) and AtMYB96 improved drought tolerance by activating cuticular wax deposition (Seo *et al.,* 2011).Transcription factors are master regulators of gene response. A transcription factor can control expression of diverse target genes involved in various physiological processes. A considerable fraction of genome of all eukaryotes is represented by genes encoding transcription factors (Riechmann *et al.,* 2000). For instance, out of the total annotated genes, 2.6% of rice genome is constituted of transcription factors (Guo *et al.,* 2008). Genome wide identification of drought responsive regulons in contrasting drought tolerant genotypes has helped in unraveling system level interplay between different genetic pathways that confer drought

tolerance; although the information about function and cross talk between them are still limited.

Recent researches have seen validation of studies about the active role of transcription factors by overexpression of their genes in transgenic plants. For example VrDREB1Afrom Vigna radiate when overexpressed in Arabidopsis showed enhanced tolerance to drought and salinity (Chen *et al.,* 2005), TaMYB3R1 from wheat when overexpressed in Arabidopsis showed enhanced tolerance to drought and salinity (Cai, 2015). BdWRKY36 from *Brachypodium distachyon* when overexpressed in tobacco enhanced tolerance to drought (Sun *et al.,* 2015), TaNAC29 from wheat when overexpressed in *Arabidopsis* showed enhanced tolerance to drought and salinity (Huang *et al.,* 2015) and TaZIP from wheat when overexpressed in Arabidopsis showed tolerance to drought, salt and freezing (Zhang *et al.,* 2015)The majority of plant transcription factors so far characterized that have a role in stomatal movements is from the model species *Arabidopsis thaliana.* The first transcription factors for which a role in stomatal opening/closure has been clearly demonstrated were the Arabidopsis AtMYB60 and AtMYB61 proteins.They are members of the R2R3MYB family, a 126 member subgroup within the MYB superfamily that, with 198 proteins in Arabidopsis, represents the largest transcription factor group in Arabidopsis(Chen *et al.,* 2005).

The expression of the AtMYB60 gene is specifically localized in guard cells. Its expression is up-regulated by signals that induce stomatal opening, such as white and blue light, and negatively down-regulated by darkness, desiccation and abscisic acid treatment, signals that promote stomatal closure.

Leaves from the atmyb60-1 knock-out mutant displayed a reduction in the light-induced aperture of stomatal pores of approximately 30% compared to wild-type leaves. These data indicate that this transcription factor represents a positive regulator of stomatal opening that is silenced in stress conditions (Comai *et al.,* 2004). Two other Arabidopsis R2R3MYB genes have been described for their involvement in guard cell movement: AtMYB44, and AtMYB15. AtMYB44 gene expression was induced by ABA and by different abiotic stresses. The gene was highly expressed in guard cells. Transgenic Arabidopsis plants overexpressing the gene are more tolerant to drought and high salinity than the wild-type (Ding *et al.,* 2012).Studies reveal that different genotypes undertake different regulatory pathways in response to water stress. Transcript profiles of drought tolerant wheat genotypes on comparison with susceptible genotypes showed that tolerant genotypes induced bZIP and HDZIP expression (transcription factors involved in ABA regulatory pathway) while sensitive genotypes induced genes encoding TFs that bind to ethylene response elements (Ergen *et al.,* 2009).

Another mechanism by which plants cope with moisture stress is by accumulation of high molecular weight, non-toxic metabolites that function as adaptive osmolytes. These metabolites increase water retention by osmotic adjustments. They include mannitol, proline, glycine, betaine, trehalose, fructan, inositol, and inorganic ions.These organic substances can regulate the plasma osmotic potential, and protect the enzymes and plasma membranes. In addition, changes in the ion and water channels control the export and import of ions and moisture for plant cells, which also contributes to osmotic adjustments. Another group of genes involved in drought tolerance are those involved in biosynthesis of enzymes involved in anti-oxidant defense systems. This includes genes encoding for enzymes viz. superoxide dismutase (SOD), catalase (CAT), ascorbate peroxidase (APX), glutathione peroxidase (GPX), glutathione reductase (GR), glutathione S-transferase (GST), dehydroascorbate reductase (DHAR), mono-dehydroascorbate reductase (MDAR), thioredoxin peroxidase (TPX), alternative oxidase (AOX), peroxiredoxin (PrxR/POD), etc (Apel and Hirt, 2004; Mittler *et al.,* 2011).

Allele mining

Huge genetic variation exists in crop gene pools for the drought tolerance genes. It is critical to make use of these genetic variations, to identify and isolate novel and superior alleles of genes having agronomic importance from available gene pools, and use them for developing improved cultivars. Allele mining is a practical way to make use of naturally occurring allelic variations of genes with desirable traits. Therefore allele mining is a promising approach which has potential applications in crop improvement programs. Potent drought resistant alleles as well as new haplotypes can be discovered using the technique of allele mining. It may also pave way for developing allele specific markers for improved marker assisted selection. The main objective of allele mining lies in identification and isolation of unknown and superior alleles from within genetic resource collections, present at a known locus that are candidates for conferring important traits. A large number of allele mining studies have been performed in recent years for dissection of useful alleles in imparting disease resistance (Wang *et al.,* 2009; Bhullar *et al.,* 2010). Intensive breeding efforts have concentrated the favorable alleles already selected during early domestication and thus contributed to further narrowing of the gene pool (Simmonds, 1976; Ladizinsky, 1985).

Gene banks preserve the genetic diversity which is otherwise lost in cultivated material. The available germplasm resources need to be screened to fish out potent alleles to enhance qualitative agronomic traits of crops (Qasim and Ashraf, 2006). Gene banks have rich diverse collection of germplasm which can be utilized to enhance the genetic potential of crops via genetic improvement programs. It is well known that phenotypic traits are controlled by genes and affected by environment, and a large numbers of accessions can adapt to environments. Germplasm collection can provide potent allele for novel traits and there will be no need to transform genes from different taxas.

Allele mining is a useful strategy for rapid characterization of diversity stored in gene bank accessions at a genetic locus of agronomic importance (Bhullar *et al.,* 2010). But handling the entire germplasm is a whooping task, whether for conventional plant breeding or for allele mining and hence must involve sampling strategies to narrow it down to a manageable size while maintaining the variability. Development of core and mini collections out of the entire collection is an effective strategy to simplify the conservation of germplasm resources and proper utilization of the existing variation in gene banks. A core collection is a subset of accessions from the entire collection which capture most of the available genetic diversity of the species. This representative subset is then subjected to screening for drought tolerance, followed by further analysis of the promising genotypes having drought tolerance.

These tolerant genotypes are often excellent genetic resources for stress tolerance but are poor yielders.One such example is the Indian landrace selection Nagina 22 (N 22), traditional rice genotype that is highly tolerant to drought. Several breeding programmes can be contemplated with such untapped germplasm accessions, most of which involves inbred or recurrent backcrossing or recurrent selection (Cortes *et al.,* 2012). Also such identified genotypes may serve in genetic engineering programs for gene transfer amongst distant species/genera.

Reasons for diversity in alleles

Wild relatives of cultivated plants didn't have to suffer from bottle necks or selective sweeps. But the cultivated plants had to undergo these processes during the course of domestication when suitable traits were selected for improvement. Thus it can be safely presumed that the wild gene pools are intact and conserve much of the variation present originally. This is supported by the fact that wild relatives are often better adapted to stressful conditions than their cultivated versions (Cortes *et al.,* 2012). Consequently it can be expected that the traits that were not subjected to diversifying selection or genes that are part of the domestication syndrome, the wild relatives have higher genetic diversity as compared to cultivated ones. This trend has been demonstrated in studies on crops like rice (Li *et al.,* 2011). Purifying selection and local adaptation are what most commonly observed in analysis of wild and cultivated varieties.

Mutations in coding regions have an excellent effect on the phenotype through changing the particular encoded protein structure as well as function. Singh *et al.,* (2015), in their study on natural allelic diversity in OsDREB1F gene in rice observed a transversion in the coding region which was responsible for non-synonymous substitution and caused an amino acid change of aspartate into glutamate which is precursor of proline in plants. This they predicted was probably responsible for drought tolerance in wild rice accessions carrying the alleles.

Mining for suitable stress inducible promoters

The adaptation of plants to environmental changes during the course of evolution has seen the participation of promoter region in a series of those changes. Polymorphisms occurring within such non-coding sequences have been found to have profound effect on phenotype by effecting alteration in the gene expression. Mutations arising within a *cis*response element can generate expression variance by changing the way transcription factors bind. Tighter or looser binding can lead to up or down regulation of transcription. EcoTILLING approach was used by Yu *et al.,* (2012) to determine the polymorphisms in 1kb promoter region of drought tolerant genes in natural varieties who observed them to be widespread. They sequenced promoters of 8 genes associated with drought resistance in 5 varieties and observed that the binding sites of the transcription factors were altered by insertions. Variations in the cis elements of the stress associated genes were found to enrich more stress related cis elements. They observed promoters were dehydration inducible, hormone responsive, and those involved in wound induced signaling.

Moreover, growth defects are often observed due to constitutive over expression of drought tolerant genes when a constitutive promoter is used (Martignago *et al.,* 2019). Therefore identification of stress inducible promoters which can have use in genetic engineering is important. Promoter mining is generally used for the expression study of the given gene and for prediction of genes. Table 2 provides various databases used for gene or promoter mining.

Strategies for allele mining

The various strategies used in allele mining programme have been described in detail.

Screening for drought tolerant genotypes

The accessions obtained from the germplasm collections need to be screened for drought tolerance. Screening for drought tolerant accessions involves not just the ability to survive but also the ability to produce a good harvestable yield under water limited condition. Intrinsic variation in drought tolerance of susceptible and tolerant genotypes can be investigated by scoring various indices of stress induced injury. This can be done by imparting moisture stress to the plants and evaluating them through various physiological and biochemical parameters imparted drought stress to two genotypes of rice, N22 which is drought tolerant traditional landrace and IR64 which is a susceptible cultivated variety (Lenka *et al.,* 2011). They compared Relative water content (RWC), total chlorophyll content and excised leaf water content in the two drought tolerant and drought susceptible genotypes and concluded that Drought tolerant showed better ability to conserve moisture in comparison to drought susceptible in response to dehydration. They also observed better drought tolerance and recovery ability than drought susceptible by visual comparison and wilting symptoms of the two cultivars.

In order to provide the greatest potential for identification of genetic variation, the genotypes must be selected from different geographical locations. When one of the objectives of allele mining is to develop a plant with good harvest index as a part of various yield components, then while phenotyping it is important to consider that both cultivated and wild gene pools are taken into account to exploit variation for drought tolerance. This is useful because several of the wild relatives would not be valuable for plant breeding given the adaption and photoperiod requirement of the equatorial versus sub-tropical zones.

Thus the accessions to be screened may include hybrids, restorer lines, CMS lines, local varieties, introgression lines, land races, wild relatives etc. A core collection that represents the entire diversity present in the germplasm needs to be prepared. However direct selection from the germplasm collections can also be done based on literature or based on available passport data that shows the genotypes to be drought tolerant (Cseri *et al.,* 2011) Screening can be also performed in vitro by evaluating the genotypes on polyethylene glycol (PEG) induced drought.

Drought may affect a plant at any stage of life, but certain stage such as germination and seedling are critical (Kingsbury *et al.,* 1984). Screening in seedling stage can be done for shoot growth, leaf rolling, canopy temperature, chlorophyll content (Chen *et al.,* 2005) Primary response to drought stress in general involves inhibition of shoot growth which allows for the diversion of cellular essential solutes from growth requirements to stress related functions. This decreases plant height and hence curbs the yield potential (Yang *et al.,* 2010). Genotypic variations revealed via osmolyte accumulation can be made to correlate their level with plant tolerance to drought. Various protocols have been described for the determination of level of osmolyte accumulation in plants. For instance, proline content determination is widely done by method described by (Bates *et al.,* 1973). The influence of seed traits on their tolerance to drought stress can be evaluated using parameters for seed quality detection and classification. Grain shape of plant seed, seed germination and seedling growth characters are important factors.

After all the accessions from the core collections are phenotyped for different parameters of moisture stress the subset of tolerant genotypes need to be identified.

Based on phenotypic responses, the genotypes are identified for allele mining. These could also serve as potential donor for drought stress tolerance in breeding programs.

There are two main methods available for the identification of sequence polymorphisms for a particular gene. They are (i) EcoTilling and (ii) sequencing based allele mining.

EcoTILLING

The term EcoTILLING was first used by Comai *et al.,* (2004) when they adapted the TILLING approach (Fig 2) to discover DNA polymorphisms occurring in natural populations of Arabidopsis thaliana. EcoTILLING has been used for rice, maize, barley, melon, wheat, wild peanut, invasive aquatic plant, black cotton wood, mung bean, potato, common bean, beet, musa, tomato, chickpea, cotton (Zhang *et al.,* 2011). To determine variation in individuals through artificially induced mutations it is a powerful reverse genetics tool for functional genomics where knockout methodologies cannot be applied (Comai *et al.,* 2004). Tilling allows the identification of allelic variation of trans-gene in a high-throughput manner. EcoTILLING involves identification of natural variance within populations or even natural mutations within germplasm without using mutagenesis. It can also be used for discovering single nucleotide polymorphism (SNPs) and small insertions and deletions (InDels) associated with the allele.

Moreover, Eco-TILLING has the potential to indicate precisely haplotypes at loci of interest as well as describe variations in microsatellite (SSR) repeat number. EcoTILLING most commonly involves discovery of polymorphisms by enzymatic mismatch cleavage followed by fluorescence detection by Li-Cor DNA analyzer (Till *et al.,* 2006).

In this approach the PCR products are amplified using infra-red dye labeled primers at the 5' end so that it can be detected in one of the two channels of the Li-Cor. After this PCR amplification and digestion using mismatch specific endonuclease is performed. The products after being purified are loaded on denaturing polyacrylamide gel and then the cleaved products are visualized in both channels of the Li-Cor. Polymorphism detected by EcoTILLING is important in order to pinpoint the mismatch. Cel-1 is the most commonly used enzyme used in EcoTILLING projects and it cleaves at 3' side of mismatches in heteroduplexes. It can be easily extracted in an inexpensive extraction method from celery stalks (Till *et al.,* 2006). Other endonucleases used are Brassica petiole extract, ENDO 1 from Arabidopsis which is believed to be more efficient that Cel-1 (Triques *et al.,* 2007).

Cseri *et al.,* (2011) used the EcoTILLING approach for allele mining in barley candidate genes for drought tolerance and observed that EcoTILLING has very high efficiency and shows little discrepancy in detecting natural polymorphisms by regenotyping the candidate gene. EcoTILLING approach was used to detect polymorphisms of transcription factor promoters (Yu *et al.,* 2012). PCR products after Cel-1 digestion between Nipponbare and testing materials were detected and they observed 69 genes with 2 alleles, 52 genes with 3 alleles, 46 gene markers with 4 alleles and 23 gene markers with 5 alleles.

The EcoTILLING approach has seen a number of useful modifications over the years. Ibiza *et al.,* 2010 were the first to use cDNA instead of genomic DNA in EcoTILLING and thus avoided DNA intron sequence problems and number of reactions was reduced. A protocol described by Torjek *et al.,* (2008) which involves use of fluorescently labeled NTPs into PCR products instead of labeled primers is used for EcoTILLING experiments in many studies now. Another variation to the traditional Eco-TILLING method has been shown by Raghvan et.al (2007), where they used a cost effective method of detecting mutations in alleles on agarose gels, which is rapid and cheap, but less sensitive. Another modification involves use of non-denaturing polyacrylamide gels stained with ethidium bromide to detect mutations (Uauy *et al.,* 2009).

The technique of EcoTILLING requires much sophistication and includes several steps, from making DNA pools of reference and test genotypes, specific conditions for efficient cleavage by nuclease, detection of mismatch in polyacrylamide gels using Li-Corgenotyper and ultimately confirmation through sequencing (Kumar *et al.,* 2010). Thus although cheaper as compared to sequence based approach, this method is cumbersome and requires more technical know-how.

Sequencing based allele mining

Another approach for allele mining is PCR-based amplification of alleles of a gene in diverse genotypes followed by DNA sequencing to recognize nucleotide variance in the alleles. By using this approach, different alleles among a variety of cultivars can be identified and isolated. Analysis of individuals for haplotype structure and study of diversity to determine genetic association in plants can also be carried out with the help of this method. It is important that the primers used must provide specific amplification without unduly compromising the evolutionary range over which allele mining can be conducted. Alleles are generally amplified using candidate gene specific, long range PCR amplification which can be followed by a nested long range PCR in presence of a high fidelity polymerase.

For mining of complete alleles which also include promoters and terminators by PCR amplification based approach, primer walking is advisable. The evolutionary distance over which PCR based allele mining succeeds is dependent strongly on the location of PCR primers within the gene.

Examination of the feasibility of allele mining of coding sequences using PCR primers based on 5'- and 3'- untranslated regions in rice and demonstrated that primers based on 5'- and 3'- UTR are sufficiently allele specific and conserved as compared to primers that are located with the coding regions as close as possible to the NC termini of the protein. Another important aspect of note is that true allele mining must include all the functional segments of the gene in the amplicon and so the location of the primers should be upstream of the promoter and downstream of the terminator (Latha *et al.,* 2004)

In order to analyze nucleotide variations in candidate genes and their regulatory sequences a number of different techniques can be used, but none is devoid of any limitation. Sequencing which is considered as the most accurate approach is relatively expensive when multiple loci in a large number of individuals are to be analyzed (Cseri *et al.,* 2011) The first step after the accessions have been carefully screened and selected for positive response to drought tolerance is extraction of the genomic DNA from them.

Genomic DNA extraction from leaf samples is generally done using the CTAB method (Murray and Thompson, 1980). Other methods used by researchers include methods given by (Dellaporta *et al.,* (1983), Törjék *et al.,* (2006) and Cuc *et al.,* (2008). A schematic representation of the two main methods followed or allele mining is given below (Fig 3).

Applications of allele mining

There are numerous applications of allele mining highlighted, of which the most important is the discovery of superior allele, SNPs and In Dels. These are helpful in functional molecular marker development for Marker assisted selection (MAS). The identified superior allele may also be directly transferred to agronomically superior but drought sensitive genotypes using genetic engineering approaches. Allele mining helps in evolutionary studies, discovery of superior haplotypes and promoter. Allele mining also helps in characterizing the huge number of accessions stored in germplasm collections. These can be later used for breeding purposes. Apart from these using the sequence information obtained from allele mining studies, syntenic relationships can be assessed among the identified loci/genes across the species/genera.

The most practical application of an allele mining experiment is to predict allelic selection on the drought tolerant genes and then to use MAS based on SNPs within the gene themselves to transfer the new alleles from wild or unadapted landraces into modern cultivars. Comparison of QTL and microarray data is difficult due to low number of sequence based markers in genetic map of crops such as wheat. To overcome this problem SNP discovery is very important.

In maize SNP variation is closer to 2% per site (Tenaillon *et al.,* 2001), in rice SNPs are estimated at about 3 to 4 per 1000 bases depending on the chromosomal region examined (Fleury *et al.,* 2010). On applied level, this very high density of SNPs has turned them into molecular markers of choice for fine mapping studies by most researchers (Rizhsky *et al.,* 2004).

Table.1 Genes with potential to serve as candidates in allele mining programmes
for drought tolerance

Sl.	Category	Name	Activity	Function
1	Transcription factors	DREB, bZIP, MYC, MYB, NAC, AP2-Domain, NF-Y, ERF, WRKY, Zinc Fingers, Others	TF induced regulation	Regulation of gene expression
2	Histone modifier	HDA	Histone deacetylase	Regulation of gene expression
3	Chromatin	MH, PDH	Stress responsive helicase	Regulation of gene expression
4	Post translational modifier	AIRP, ATL	Ubiquitin ligase	Regulator of abscisic acid-dependent response to drought stress
5	Protein kinase and Protein Phosphatases	AHK, CBPK, CIPK, CPK, MKK, NPK. ABI, HAB	Enzymes that cause conformational change from an inactive to an active form of the protein	Proteins of signaling cascades that help in Signal transduction
6	Hormone signaling	NCED, AAO,DSM	ABA Biosynthesis	Regulation of physiological processes ranging from stomatal opening to protein storage and TF induction
7	Detoxification	P5CS, GolS1, TPS, FSPD, CMO	Osmolyte production	Physiological Adaptation to retain the water potential, cell turgor, and membrane stability
8	Protection factors	LEA, HVA, TAS	Late embryogenesis abundant proteins	Functional proteins that protect the cellular membranes and other proteins
9	ROS scavenger	GPX, APX, SOD, GSTU, MT3a	Removal of ROS	Overexpression decreases sensitiveness to drought
10	Root development	EVP	Modulation of root system architecture	Increases number of root hairs

Table.2 List of bioinformatic tools/web resources/databases useful for allele/promoter mining for drought tolerance

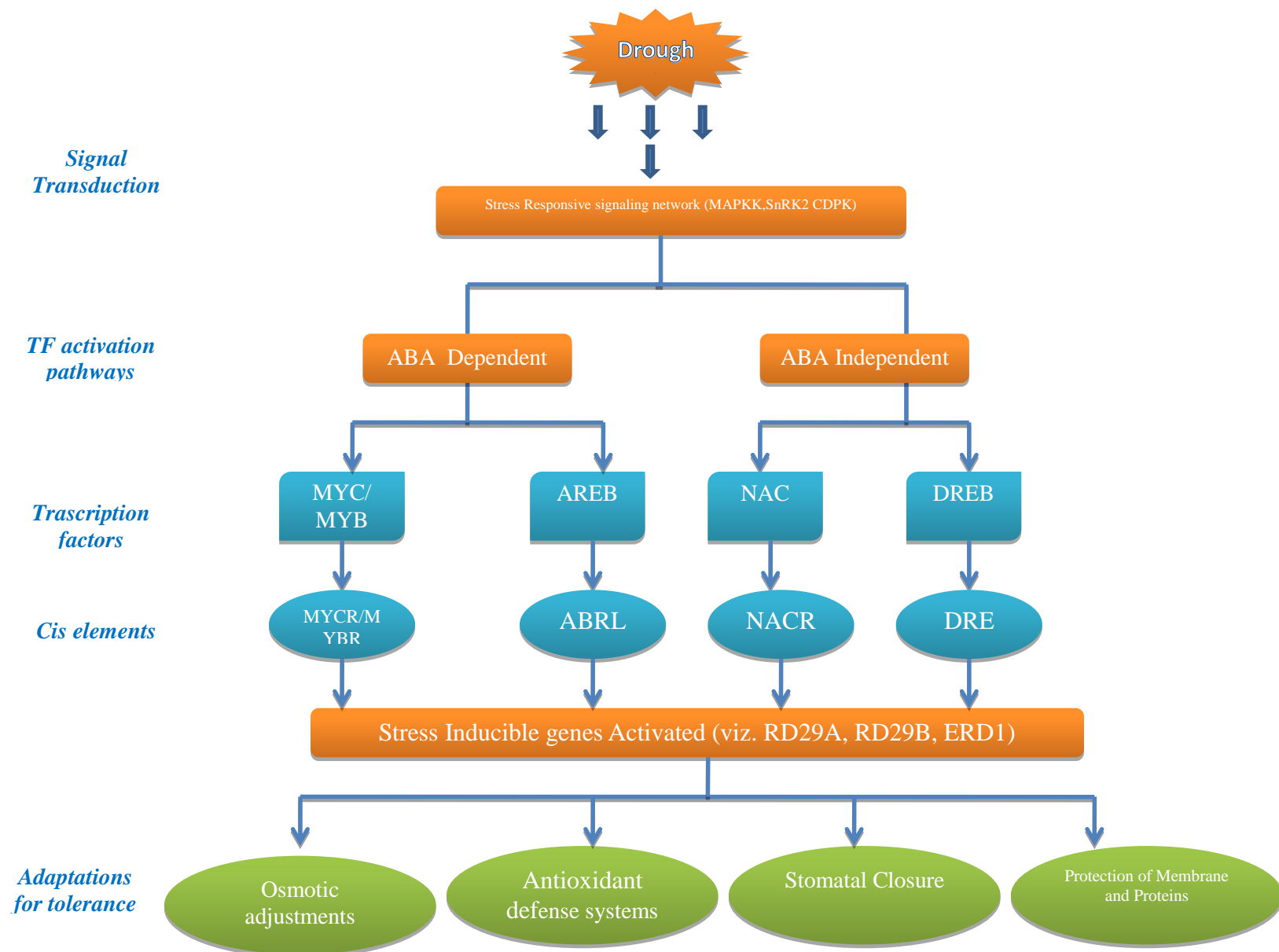| Sl.	Name	Use	Website	References
1	DroughtDB	compilation of molecularly characterized genes that are involved in drought stress response	http://pgsb.helmholtz-muenchen.de/droughtdb/	Alter *et al.,* 2015
2	PlantTFDB	Plant Transcription Factor Database	http://planttfdb.cbi.pku.edu.cn/	Tian *et al.,* 2020
3	GRAMENE	data resource for comparative functional genomics in crops and model plant species	http://www.gramene.org/	Tello-Ruiz *et al.,* 2018
4	JASPAR	database of curated, non-redundant transcription factor (TF) binding profiles	http://jaspar.genereg.net/	Fornes *et al.,* 2020
5	AGRIS	Database of Arabidopsis promoter sequences, transcription factors and their target genes	https://agris-knowledgebase.org/	Yilmaz *et al.,* 2010
6	PlantPromDB	Database of Plant Promoter Sequences	http://linux1.softberry.com/berry.phtml?topic=plantprom&group=data&subgroup=plantprom	Shahmuradov *et al.,* 2003
7	PlantCARE	a database of plant *cis*-acting regulatory elements	http://bioinformatics.psb.ugent.be/webtools/plantcare/html/	Magali *et al.,* 2002
8	PLACE	database of motifs found in plant cis-acting regulatory DNA elements	http://www.dna.affrc.go.jp/PLACE/	Higo *et al.,* 1999
9	EPD	collection of eukaryotic POL II promoters	https://epd.epfl.ch//index.php	Dreos *et al.,* 2017
10	MUSCLE	Nucleotide sequence analysis	https://www.ebi.ac.uk/Tools/msa/muscle/	Madeira *et al.,* 2019
11	Clustal omega	Multiple Sequence alignment	https://www.ebi.ac.uk/Tools/msa/clustalo/	Madeira *et al.,* 2019
12	MEGA	Multiple Sequence alignment	https://www.megasoftware.net/	Kumar *et al.,* 2018

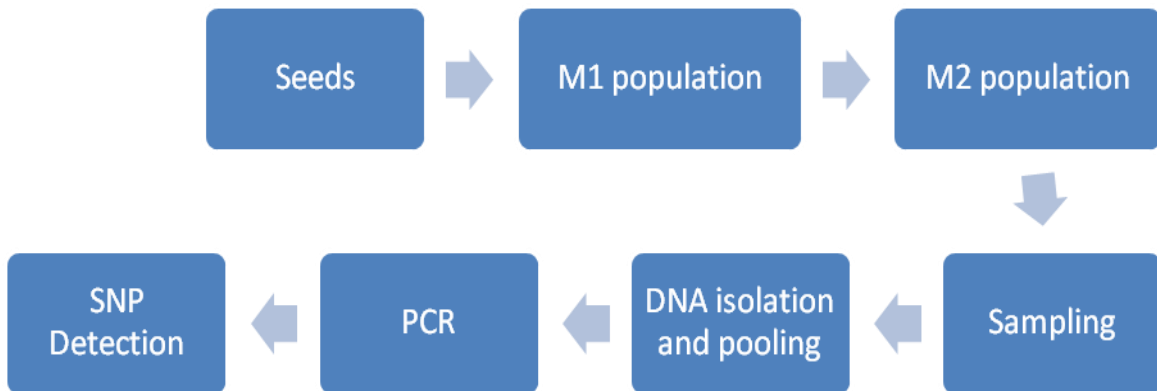**Fig.1** Schematic representation of molecular response to drought stress

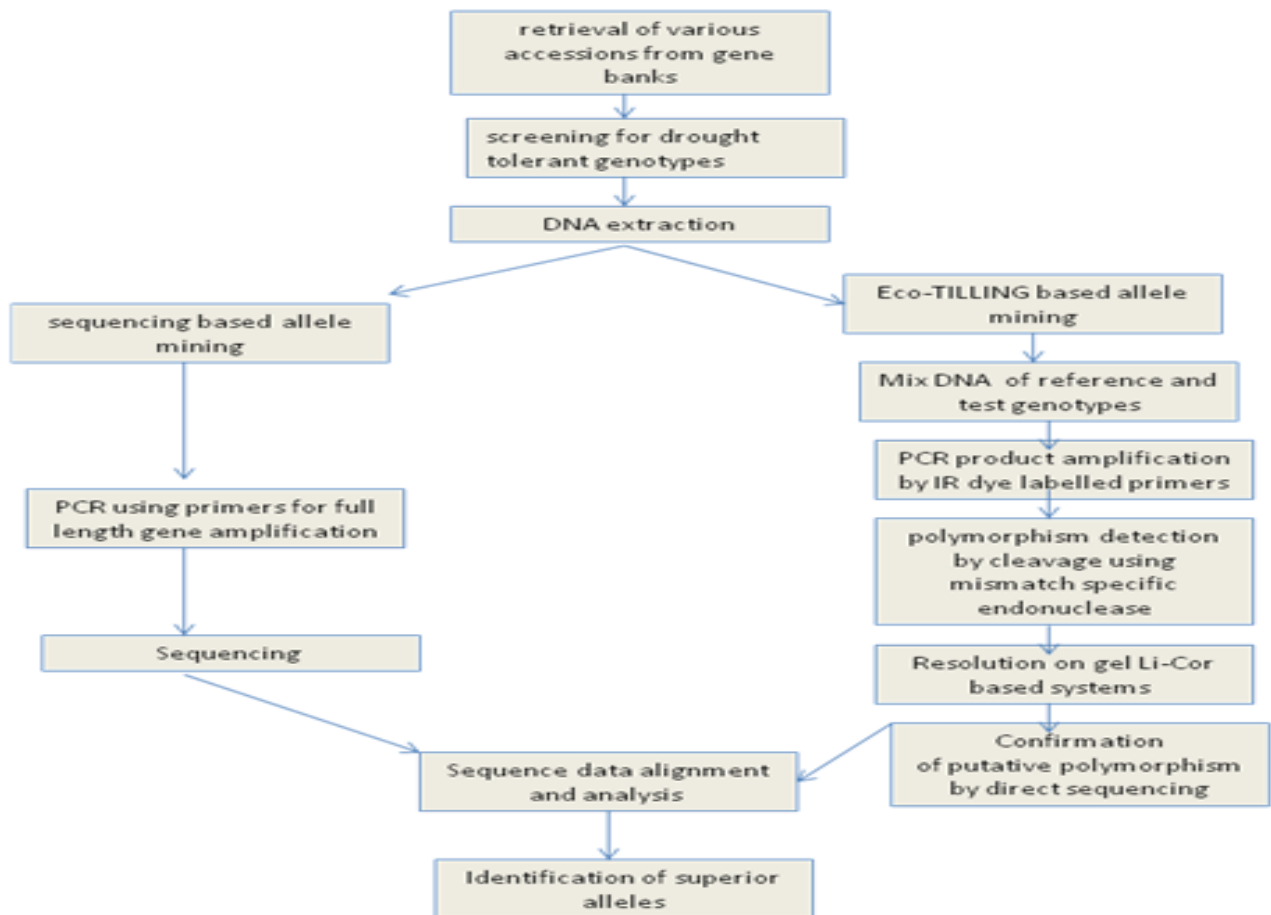**Fig.2** Schematic representation of TILLING based allele minning

Fig.3 Flowchart depicting the overview of allele mining steps

Apart from this SNP study provides a framework for examining how population history, selection and breeding systems affect variation at genetic loci. This helps to delineate the mechanism that lead to evolutionary diversification of genomes (Nordborg and Innan, 2011; Palaisa *et al.,* 2004). SNPs provide the ultimate anchor to relate all forms of polymorphisms, including biochemical, metabolic, physiological and phenotypic performance (McNally *et al.,* 2006). In a study, the gene for DREB1F, a potent drought tolerance transcription activator was re-sequenced for carrying out allele mining and association study in a set of 136 wild rice accessions and 4 cultivated varieties and identified 22 SNPs with 8 haplotypes. By association studies it was revealed that 3 coding SNPs were significantly associated with drought tolerance (Singh *et al.,* 2015). Some varieties withstood a long term directional selection and changes that improved drought resistance were accumulated. InDels thus obtained can be used in marker anchored genetic map for identification of major QTLs governing candidate genes for drought tolerance.

Cseri *et al.,* (2011) conducted allele mining on a panel of drought related candidate genes in a set of 96 barley genotypes using EcoTILLING. 185 single nucleotide polymorphisms (SNPs), 46 insertions/deletions (INDELs) and94 verified unique haplotypes were detected. Based on overlapping haplotype sequences, markers were developed for four candidate genes.

One of the major limitations for crop production is optimal availability of water. Concerns about water accessibility have always accompanied crop production in dry areas, which are on the other hand the most extensive areas for agriculture on earth. As a consequence, one has to develop agricultural strategies to cope with water shortage,

growing the plants during the short climatic interval of water availability and selecting plants possessing a relatively superior tolerance to water deficiency. Several molecular networks involved in stress perception, signal transduction and stress responses in plants have been elucidated so far. Transcription factors are major players in water stress signaling. Various studies suggest the role of different myb transcription factor genes in drought tolerance/response. So to find out the major genes responsible for the regulation of these transcription factors allele mining is the strategy which can be utilized.

The use of genetic diversity is limited due to the resources which are at hand for characterization of all the available lines. Therefore, we need to (i) develop strategies to assemble focused sets of material for specific traits based on criteria for selection of the lines but also (ii) to identify genes having traits of agronomic importance and (iii) establish the molecular tools for rapid characterization of new alleles. Allele mining is a promising approach to dissect naturally occurring allelic variation at candidate genes controlling key agronomic traits which has potential applications in crop improvement programs. Allele mining can be effectively used for discovery of superior alleles by 'mining' the gene of interest from diverse genetic resources.

It can help in accessing and determining the change in nucleotide sequence linked with superior alleles as well as develops understanding of phenotypic changes associated with novel traits at molecular level. Candidate genes for stress tolerance may be used in crop improvement programs through identification of linked SNPs. Single nucleotide polymorphisms (SNPs) have gained much popularity in assessing the diversity because of automation and abundance.

SNP is of great importance if it affects gene function and the function of the gene in stress response is known and the SNP is associated with differences in plant performance. Functional significance of the mined superior alleles can be confirmed and evaluated by transformation in another plant or introducing their full length cDNA clones in an expression vector.

References

Adunga, A. 2007. The role of introduced sorghum and millets in Ethiopian agriculture. *SAT J.* 3, pp. 1-4.

Alter, S., Bader, K.C., Spannagl, M., Wang, Y., Bauer, E., Schön, C.C. and Mayer, K.F., 2015. DroughtDB: an expert-curated compilation of plant drought stress genes and their homologs in nine species. *Database*, *2015*.

Apel, K. and Hirt, H., 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.*, *55*, pp.373-399.

Bates, L.S., Waldren, R.P. and Teare, I.D., 1973. Rapid determination of free proline for water-stress studies. *Plant and soil*, *39*(1), pp.205-207.

Bhullar, N.K., Zhang, Z., Wicker, T. and Keller, B., 2010. Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene Pm3: a large scale allele mining project. *BMC plant biology*, *10*(1), p.88. doi: 10.1186/1471-2229-10-88.

Cai, Y., Agmon, N., Choi, W.J., Ubide, A., Stracquadanio, G., Caravelli, K., Hao, H., Bader, J.S. and Boeke, J.D., 2015. Intrinsic biocontainment: multiplex genome safeguards combine transcriptional and recombinational control of essential yeast genes. *Proceedings of the National Academy of Sciences*, *112*(6), pp.1803-1808.

Chen, B.J., Wang, Y., Hu, Y.L., Wu, Q. and Lin, Z.P., 2005. Cloning and characterization of a drought-inducible MYB gene from Boeacrassifolia. *Plant Science*, *168*(2), pp.493-500.

Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R. and Henikoff, S., 2004. Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *The Plant Journal*, *37*(5), pp.778-786.

Cortés, A.J., This, D., Chavarro, C., Madrinán, S. and Blair, M.W., 2012. Nucleotide diversity patterns at the drought-related DREB2 encoding genes in wild and cultivated common bean (Phaseolus vulgaris L.). *Theoretical and Applied Genetics*, *125*(5), pp.1069-1085.

Cseri, A., Cserháti, M., Von Korff, M., Nagy, B., Horváth, G.V., Palágyi, A., Pauk, J., Dudits, D. and Törjék, O., 2011. Allele mining and haplotype discovery in barley candidate genes for drought tolerance. *Euphytica*, *181*(3), p.341.

Cuc, L.M., Mace, E.S., Crouch, J.H., Quang, V.D., Long, T.D. and Varshney, R.K., 2008.Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (Arachishypogaea). *BMC plant biology*, *8*(1), pp.55.

Dellaporta, S.L., Wood, J. and Hicks, J.B., 1983. A plant DNA minipreparation: version II. *Plant molecular biology reporter*, *1*(4), pp.19-21.

Ding, Z., Taneva, S.G., Huang, H.K., Campbell, S.A., Semenec, L., Chen, N. and Cornell, R.B., 2012. A 22-mer segment in the structurally pliable regulatory domain of metazoan CTP: phosphocholinecytidylyltransferase

facilitates both silencing and activating functions. *Journal of Biological Chemistry*, *287*(46), pp.38980-38991.

Dreos, R., Ambrosini, G., Groux, R., CavinPérier, R. and Bucher, P., 2017. The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic acids research*, *45*(D1), pp.D51-D55.

Ergen, N.Z., Thimmapuram, J., Bohnert, H.J. and Budak, H., 2009. Transcriptome pathways unique to dehydration tolerant relatives of modern wheat. *Functional & integrative genomics*, *9*(3), pp.377-396.

Fleury, D., Jefferies, S., Kuchel, H. and Langridge, P., 2010. Genetic and genomic tools to improve drought tolerance in wheat. *Journal of experimental botany*, *61*(12), pp.3211-3222.

Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D. and Santana-Garcia, W., 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic acids research*, *48*(D1), pp.D87-D92.

Gorantla, M., Babu, P.R., Reddy Lachagari, V.B., Reddy, A.M.M., Wusirika, R., Bennetzen, J.L. and Reddy, A.R., 2007. Identification of stress-responsive genes in an indica rice (Oryzasativa L.) using ESTs generated from drought-stressed seedlings. *Journal of experimental botany*, *58*(2), pp.253-265.

Guo, P., Baum, M., Varshney, R.K., Graner, A., Grando, S. and Ceccarelli, S., 2008. QTLs for chlorophyll and chlorophyll fluorescence parameters in barley under post-flowering drought. *Euphytica*, *163*(2), pp.203-214.

Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T., 1999. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic acids research*, *27*(1), pp.297-300.

Huang, Q., Wang, Y., Li, B., Chang, J., Chen, M., Li, K., Yang, G. and He, G., 2015. TaNAC29, a NAC transcription factor from wheat, enhances salt and drought tolerance in transgenic Arabidopsis. *BMC plant biology*, *15*(1), p.268.

Ibiza, V.P., Cañizares, J. and Nuez, F., 2010. EcoTILLING in Capsicum species: searching for new virus resistances. *BMC genomics*, *11*(1), p.631.

Jung, C., Seo, J.S., Han, S.W., Koo, Y.J., Kim, C.H., Song, S.I., Nahm, B.H., Do Choi, Y. and Cheong, J.J., 2008. Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic Arabidopsis. *Plant physiology*, *146*(2), pp.623-635.

Kingsbury, R.W. and Epstein, E., 1984. Selection for Salt- Resistant Spring Wheat 1. *Crop Science*, *24*(2), pp.310-315.

Kumar, G.R., Sakthivel, K., Sundaram, R.M., Neeraja, C.N., Balachandran, S.M., Rani, N.S., Viraktamath, B.C. and Madhav, M.S., 2010. Allele mining in crops: prospects and potentials. *Biotechnology advances*, *28*(4), pp.451-461.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, *35*(6), pp.1547-1549.

Ladizinsky, G., 1985. Founder effect in crop-plant evolution. *Economic Botany*, *39*(2), pp.191-199.

Latha, R., Rubia, L., Bennett, J. and Swaminathan, M. S. 2004. Allele mining for stress tolerance genes in *Oryza* species and related germplasm. *Molecular biotechnology*, *27*(2) pp.101.

Lenka, S.K., Katiyar, A., Chinnusamy, V. and Bansal, K.C., 2011. Comparative analysis of drought- responsive transcriptome in Indica rice genotypes with contrasting drought tolerance. *Plant biotechnology journal*, *9*(3), pp.315-327.

Lescot, M., 2002. Patrice D hais, GertThijs, Kathleen Marchal, Yves Moreau, Yves Van de Peer, Pierre Rouz and StephaneRombauts. *Nucleic Acids Res*, *30*(1), pp.325-327.

Li, Z.M., Zheng, X.M. and Ge, S., 2011. Genetic diversity and domestication history of African rice (Oryzaglaberrima) as inferred from multiple gene sequences. *Theoretical and Applied Genetics*, *123*(1), pp.21-31.

Liang, Y.K., Dubos, C., Dodd, I.C., Holroyd, G.H., Hetherington, A.M. and Campbell, M.M., 2005. AtMYB61, an R2R3-MYB transcription factor controlling stomatal aperture in Arabidopsis thaliana. *Current Biology*, *15*(13), pp.1201-1206.

Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K. and Shinozaki, K., 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought-and low-temperature-responsive gene expression, respectively, in Arabidopsis. *The Plant Cell*, *10*(8), pp.1391-1406.

Lucas, S., Durmaz, E., Akpınar, B.A. and Budak, H., 2011. The drought response displayed by a DRE-binding protein from Triticumdicoccoides. *Plant Physiology and Biochemistry*, *49*(3), pp.346-351.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R., Potter, S.C., Finn, R.D. and Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, *47*(W1), pp.W636-W641.

Martignago, D., Rico-Medina, A., Blasco-Escaméz, D., Fontanet-Manzaneque, J.B. and Caño-Delgado, A.I., 2019. Drought resistance by engineering plant tissue-specific responses. *Frontiers in Plant Science*, *10*, pp.1676.

McNally, K.L., Bruskiewich, R., Mackill, D., Buell, C.R., Leach, J.E. and Leung, H., 2006. Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiology*, *141*(1), pp.26-31.

Mittler, R., Vanderauwera, S., Suzuki, N., Miller, G.A.D., Tognetti, V.B., Vandepoele, K., Gollery, M., Shulaev, V. and Van Breusegem, F., 2011. ROS signaling: the new wave?. *Trends in plant science*, *16*(6), pp.300-309.

Murray, M.G. and Thompson, W.F., 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic acids research*, *8*(19), pp.4321-4326.

Nordborg, M. and Innan, H., 2002. Molecular population genetics. *Current opinion in plant biology*, *5*(1), pp.69-73.

Palaisa, K., Morgante, M., Tingey, S. and Rafalski, A., 2004. Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *Proceedings of the National Academy of Sciences*, *101*(26),

pp.9885-9890.

Qasim, M. and Ashraf, M., 2006. Time course of ion accumulation and its relationship with the salt tolerance of two genetically diverse lines of canola (Brassica napus L.). *Pakistan Journal of Botany*, *38*(3), p.663-672.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C.Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R. and Creelman, R., 2000. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*, *290*(5499), pp.2105-2110.

Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., Davletova, S. and Mittler, R., 2004. When defense pathways collide. The response of Arabidopsis to a combination of drought and heat stress. *Plant physiology*, *134*(4), pp.1683-1696.

Seo, P.J., Lee, S.B., Suh, M.C., Park, M.J., Go, Y.S. and Park, C.M., 2011. The MYB96 transcription factor regulates cuticular wax biosynthesis under drought conditions in Arabidopsis. *The Plant Cell*, *23*(3), pp.1138-1152.

Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., Bramley, P.M. and Solovyev, V.V., 2003. PlantProm: a database of plant promoter sequences. *Nucleic acids research*, *31*(1), pp.114-117.

Shinozaki, K. and Yamaguchi-Shinozaki, K., 2007. Gene networks involved in drought stress response and tolerance. *Journal of experimental botany*, *58*(2), pp.221-227.

Simmonds, N.W. 1976. Evolution of Crop Plants. Longman, London. pp.339.

Singh, B.P., Jayaswal, P.K., Singh, B., Singh, P.K., Kumar, V., Mishra, S., Singh, N., Panda, K. and Singh, N.K., 2015. Natural allelic diversity in OsDREB1F gene in the Indian wild rice germplasm led to ascertain its association with drought tolerance. *Plant cell reports*, *34*(6), pp.993-1004.

Solomon, S., Manning, M., Marquis, M. and Qin, D., 2007. *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC* (Vol. 4). Cambridge university press.

Sun, Y., Ning, T., Liu, Z., Pang, J., Jiang, D., Guo, Z., Song, G. and Yang, D., 2015. The OsSec18 complex interacts with P0 (P1-P2) 2 to regulate vacuolar morphology in rice endosperm cell. *BMC plant biology*, *15*(1), p.55.

Tello-Ruiz, M.K., Naithani, S., Stein, J.C., Gupta, P., Campbell, M., Olson, A., Wei, S., Preece, J., Geniza, M.J., Jiao, Y. and Lee, Y.K., 2018. Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic acids research*, *46*(D1), pp.D1181-D1189.

Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F. and Gaut, B.S., 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). *Proceedings of the National Academy of Sciences*, *98*(16), pp.9161-9166.

Tian, F., Yang, D.C., Meng, Y.Q., Jin, J. and Gao, G., 2020. PlantRegMap: charting functional regulatory maps in plants. *Nucleic acids research*, *48*(D1), pp.D1104-D1113.

Till, B.J., Zerr, T., Comai, L. and Henikoff, S., 2006. A protocol for TILLING and Ecotilling in plants and animals. *Nature protocols*, *1*(5), p.2465.

Törjék, O., Witucka-Wall, H., Meyer, R.C., Von Korff, M., Kusterer, B.,

Rautengarten, C. and Altmann, T., 2006. Segregation distortion in Arabidopsis C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theoretical and Applied Genetics*, *113*(8), pp.1551-1561.

Triques, K., Sturbois, B., Gallais, S., Dalmais, M., Chauvin, S., Clepet, C., Aubourg, S., Rameau, C., Caboche, M. and Bendahmane, A., 2007. Characterization of Arabidopsis thaliana mismatch specific endonucleases: application to mutation discovery by TILLING in pea. *The Plant Journal*, *51*(6), pp.1116-1125.

Uauy, C., Paraiso, F., Colasuonno, P., Tran, R.K., Tsai, H., Berardi, S., Comai, L. and Dubcovsky, J., 2009. A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC plant biology*, *9*(1), p.115.

Varshney, R.K., Bansal, K.C., Aggarwal, P.K., Datta, S.K. and Craufurd, P.Q., 2011. Agricultural biotechnology for crop improvement in a variable climate: hope or hype?. *Trends in plant science*, *16*(7), pp.363-371.

Wang, X., Niu, Q.W., Teng, C., Li, C., Mu, J., Chua, N.H. and Zuo, J., 2009. Overexpression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in Arabidopsis. *Cell research*, *19*(2), pp.224-235.

Yang, S., Vanderbeld, B., Wan, J. and Huang, Y., 2010. Narrowing down the targets: towards successful genetic engineering of drought-tolerant crops. *Molecular plant*, *3*(3), pp.469-490.

Yilmaz, A., Mejia-Guerra, M.K., Kurz, K., Liang, X., Welch, L. and Grotewold, E., 2010. AGRIS: the Arabidopsis gene regulatory information server, an update. *Nucleic acids research*, *39*(suppl_1), pp.D1118-D1122.

Yu, S., Liao, F., Wang, F., Wen, W., Li, J., Mei, H. and Luo, L., 2012. Identification of rice transcription factors associated with drought tolerance using the ecotilling method. *PLoS One*, *7*(2) e30765.

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., Li, C., Liu, B., Luo, Y., Zhu, Y., Zhang, N., He, S., He, C., Wang, H., Chen, D. 2015. N(6)-methyladenine DNA modification in Drosophila. Cell 161(4), pp. 893-906.

Zhang, X., Ju, H. W., Chung, M. S., Huang, P., Ahn, S. J. and Kim, C. S. 2011. The R-R-type MYB-like transcription factor, AtMYBL, is involved in promoting leaf senescence and modulates an abiotic stress response in *Arabidopsis*. *Plant Cell Physiol.,* 52, pp. 138–148.