



P-ISSN: 2349-8528

E-ISSN: 2321-4902

www.chemijournal.com

IJCS 2020; 8(4): 187-195

© 2020 IJCS

Received: 07-05-2020

Accepted: 11-06-2020

Sonu Get

Ph.D., Scholar, Division of Plant Breeding and Genetics, Rajasthan Agricultural Research Institute, Durgapura, SKNAU, Jobner, Jaipur, Rajasthan, India

Unveiling of allelic variants of crop plants from natural populations through allele mining strategies

Sonu Get

DOI: <https://doi.org/10.22271/chemi.2020.v8.i4c.9687>

Abstract

Currently crop production is affected by several emerging biotic and abiotic stresses that are ultimately insufficient to meet the food requirement of the growing human population. Plant genetic resources are a treasure for many novel alleles that play an important role in adjusting crop genotypes under the changing climatic landscape. Germplasm of natural populations and crop species around the world contain rare alleles that are no longer present in our current breeding line, therefore a large opportunity to characterize these genetic resources at the molecular level and identify key alleles that provide resistance to biotic and abiotic stresses. Allele mining is a recent and promising technology to identify the natural allelic variation at candidate genes controlling important agronomic traits that has great role in improvement of crop species. This approach proved its potential applications in plant breeding by detecting the new superior alleles for various traits like disease resistance, drought tolerance, salinity tolerance, quality etc in many crops. Therefore, allele mining can be efficiently used to unclosethis hidden allelic variation in natural population as well as in germplasm collection.

Keywords: Allele mining, TILLING, Eco-TILLING, haplotype, synteny, association mapping

Introduction

Plant breeding relies on the availability of genetic diversity and their identification in existing gene pool of the crop that can give enormous functional variants of genes. Introduction, selection, hybridization and genetic engineering are major techniques in plant breeding that cause genetic variability in the current breeding lines of crops. Genetic engineering is an adjunct technique to conventional breeding techniques that makes it possible to improve crop genotypes by transferring genes from unrelated organisms. In course of agricultural development, early domesticated plants were gradually replaced first by landraces and farmers varieties, and later by genetically uniform modern cultivars. Despite of this, crop wild relatives which are the store house of superior and rare alleles, facing a serious problem of threat by natural calamities and anthropogenic factors (Tanksley, *et al.*, 1997) ^[50]. Therefore, gene bank collections of germplasm resources are essential to store and conserve this biodiversity and thus pay big dividends to improvement of crops when used efficiently (Johnson, 2008) ^[25]. Till date, the progress in plant breeding in terms of development of superior and high yielding varieties of agricultural crops is possible by accumulation of desirable alleles from vast plant germplasm resource existing worldwide (Yugandhar, *et al.*, 2017) ^[63]. Collecting a large number of germplasms is not sufficient, it is essential to identify the available and hidden genetic variability in that collections to rectify the undesirable agronomic traits and developing resistant genotypes to several biotic and abiotic stresses. Although several studies depicted the utilization of germplasm resources in plant breeding (Yu, *et al.*, 2008) ^[62], but the global germplasm collections are unexplored for a number of crop species. Hence, these vast germplasm collections need to be relooked for novel alleles to further enhance the genetic potential of crop varieties for various agronomic traits. Identification and access to these novel alleles can be used in breeding programs to produce high yielding crop varieties in order to address the changing climatic scenario. Now, the challenge is how to unlock this variation. Allele mining is a research field aimed at finding the allelic variation of relevant traits within germplasm resources collections. The strategy of detecting alleles of a known locus that are involved in expression of a particular trait phenotype and their variants in other genotypes or

Sonu Get

Ph.D., Scholar, Division of Plant Breeding and Genetics, Rajasthan Agricultural Research Institute, Durgapura, SKNAU, Jobner, Jaipur, Rajasthan, India

identifying novel, superior and beneficial alleles from the germplasm or natural population is known as allele mining. Allele mining is a lately developed technique that promises to find out naturally occurring allelic variation in the candidate genes and their introgression in to agronomically adapted genotypes or varieties. Allele mining is a relatively underexplored method to identify new alleles at a known locus (Bhullar, *et al.*, 2009) [4]. There are two techniques of this approach i.e. 1) Modified TILLING (Eco-TILLING) based allele mining & 2) Sequenced based allele mining.

TILLING (Targeting Induced Local Lesions IN Genomes) technique is suitable for most plants (McCallum *et al.* 2000a) [34]. For TILLING, mutations are created by treatment with the same chemical mutagens that have been successfully employed in mutation breeding programs for decades. By using chemical mutagens that induce primarily random point mutations at high density, allelic series of missense and truncation mutations can be discovered with TILLING (Greene *et al.* 2003) [20]. Thus with only a small population, multiple alleles may be obtained regardless of the size of the gene. Gene regions are targeted for mutation discovery using PCR and standard SNP discovery methods. The use of general techniques for the generation and discovery of mutations means that the method should be applicable to a wide variety of organisms. TILLING methodology can be used to disclose the natural nucleotide variation linked to important phenotypic traits, a process termed Eco-TILLING (Comai *et al.* 2004) [13]. Both these techniques uses gene and genome sequences available in public database and PCR strategies are devised to identify and isolate novel and superior alleles for key agronomic traits from a wide range of crop species to suitably organize the development of improved cultivars.

Existing and emerging tools of genomics will be used to assign the function of unknown genes, to relate genotype to phenotype, and to identify the genes governing important traits. Reverse genetics is a powerful technique for elucidation of gene function through generation of mutations in candidate genes that can then be assayed for phenotypes. Many reverse genetic techniques have been developed for plants (An, *et al.* 2005; Burch-Smith, *et al.* 2004; Henikoff and Comai 2003; Kusaba, 2004) [2, 6, 21, 31]. Some involve targeting genes one at a time, such as transformation with hairpin constructs for RNAi-mediated *knock down*, which can be used to dominantly *knock down* multiple homologous genes with a single construct and CRISPR Cas-9 for targeted editing of genes either by *knock out* or by *knock in*. Other techniques involve creating a mutagenized population that can be screened for generally recessive DNA lesions, including insertions, deletions and point mutations. Genomic deletions have also been exploited for reverse genetics. This method has the advantage to excise the target blocks of randomly repeated genes. To discover a deletion, PCR is performed with primers that hybridize to the genome.

Genetic variation is caused by allelic diversity at the genetic loci contributing to a particular trait. The genetic variation is a prerequisite for any genetic improvement in crop varieties. Therefore, in any breeding program, this is always the first step unless genetic variation pre-exists in the breeding population. It can be created by several breeding methods among them mutation is an evolutionary driving force which cause existing allelic diversity in any crop species. Generally, mutation occur in coding and/or regulatory regions of genome either as Single Nucleotide Polymorphism (SNP) or as Insertion and Deletion (InDel) may have a tremendous effect on changing phenotype of trait by changing the amino acid

sequences and/or their function in protein encoded by targeted DNA segment where as those occur in the non-coding regions will have a silent effect on the phenotype of the trait. In other words, it is an important source for creating a new alleles or causing variation in the preexisting allele and allelic combinations. For instance, the long growth habit of wild rice is governed by a single gene PROG1 while the mutant allele of this wild allele produce an erect growth habit, greater grain number and higher grain yield in most of the cultivated rice (Tan *et al.*, 2008) [48]. On contrary to its beneficial effects in creating genetic variation, most of the mutations are deleterious in their effect but in general 0.1% of the mutations are super-vital leading to alteration in gene function which may be highly necessary for the survival of the plant (Singh, 2005) [45]. Allele mining can be potentially employed in the identification of new haplotypes, allele specific marker, allelic synteny and evolutionary relationship and effectively utilized in expression study as well as gene prediction through promoter mining.

TILLING (Targeting Induced Local Lesions IN Genomes)

TILLING is a non-transgenic reverse genetics approach to study gene functions in candidate genes follows screening of induced mutations in individuals (McCallum *et al.*, 2000b) [35]. It helps in the direct identification of allelic variation or polymorphism resulting from induced mutations in targeted genes by heteroduplex analysis (Till *et al.*, 2003) [52]. Hence, it is a powerful reverse genetics tool for functional genomics where knockout approaches cannot be applicable. This approach consists of three main principles; Development of a mutagenized population, DNA preparation and pooling, and mutation discovery or search for variant (Fig. 1).

Development of a mutagenized population

TILLING is ideally suitable to a number of plants. The most suitable plant part of TILLING is seed or pollen grain. For seed mutagenesis, a single M2 plant per line is typically included in the TILLING population. The M1 generation is chimeric for mutations and is unsuitable for TILLING. DNA and seed are collected from the M2 generation. When mutations are identified, the M3 seed can be germinated for phenotypic analysis. When mutagenizing pollen, the M1 generation is not chimeric and can be screened for TILLING. All M1 progeny from a single cross should carry distinct heterozygous mutations as each pollen grain will have accumulated mutations randomly. Seed from a self-cross of the M1 is collected for subsequent phenotypic analysis.

The ideal mutagen for TILLING is one that randomly induces single nucleotide substitutions, or small insertions/deletions of about 30 nucleotides at a high frequency in throughout the genome. The ethyl methane sulfonate (EMS) is chemicals that have multiple effects on DNA. It induce all types of mutations, including transitions, transversions, frameshifts, and even chromosome aberrations, with relative frequencies. The EMS generates mostly SNPs, and can be controlled to produce a high density of point mutations, causing a variety of lesions including nonsense and missense mutations and cause G:C → A:T transitions (Greene, *et al.* 2003, Koornneef, *et al.* 1982) [20, 26]. Indeed, EMS has been the mutagen of choice for most plant TILLING program. The seed mutagenesis is most practical for many plants. Seed are soaked in 0.3-1% ethyl methane sulfonate solution chemical mutagen for about 14–18 hrs. The resulting adult M₁ generation plant will contain different chimeric tissues due to multicellular nature of embryo (Henikoff and Comai, 2003)

[21]. Thus, mutations present in the somatic tissue will not match those in the germinal tissues, and this generation is not suitable for TILLING screens. M_1 generation mutations will be heterozygous, and therefore M_2 generation plants from a self cross of the M_1 should segregate mutations in 1:2:1 Mendelian ratio. The single seed descent approach provides a predictable ratio of mutant to wild-type alleles in a single individual either 1:1 for heterozygotes or 2:0 for homozygous alleles and allows for straightforward segregation analysis of the M_3 generation (Colbert, *et al.*, 2001) [11]. Pollen mutagenesis was used in Maize TILLING program (Till, *et al.*, 2004) [53]. After crossing the mutagenized pollen onto a non-mutagenized ear, each developed kernel contains unique heterozygous mutations. The M_1 plants are therefore suitable to use in a TILLING screen and hundreds of unique lines are possible from a single ear. Pollen mutagenesis therefore requires fewer field resources than the seed mutagenesis approach. A potential additional advantage of pollen mutagenesis is that the nearly quiescent pollen might be less sensitive to cytotoxic effects from chemical mutagens. For many organisms, the appropriate measure of mutagenesis will not be known in advance and can only be determined through careful observation of phenotypes, followed by TILLING screens to determine the density of induced mutations.

DNA preparation and pooling (Arraying the population for TILLING)

The sample pooling will also directly affect the efficiency and cost of mutation discovery. The screening of four samples pooled together will take approximately twice as long and cost twice as much as screening a pool of eight samples. Factors that affect the pooling of sample include the quality of genomic DNA, the accuracy of sample quantification, and the method used for SNP discovery. Isolate the genomic DNA from M_2/M_3 individual plants and standardized the DNA concentration of each sample. To construct pools it is required to take equal quantities of DNA from individual samples and from which robust PCR results can be obtained. Sample pools are arrayed in a 96-well microtiter plate and allowing up to 8-fold pooling of individual plants to increase TILLING throughput. Two basic pooling strategies have been most often used. For large scale services, we typically use a one-dimensional pooling strategy where each individual sample is represented in only one pool. When a mutation is identified in a pool of eight individuals, each member of the pool is then screened independently to identify the individual sheltering the mutation (Colbert, *et al.*, 2001) [11]. The other approach is to pool samples two-dimensionally such that each sample is present in two unique pools. This pooling strategy used for smaller scale projects and for the larger scale Maize TILLING service. Though, duplicating each sample decrease the throughput of detection in pools by half, the sample sheltering the mutation is unambiguously determined in the pool screening, so there is no need to screen individual samples as is done in the one dimensional strategy. Also, because two-dimensional pooling involves screening each sample with two-fold coverage, potential false positive and false negative errors are minimized at the initial screening step, rather than when individuals are screened in the second step with one-dimensional pooling. Before moving to a large-scale operation such as a public TILLING service, the advantage of higher throughput using one-dimensional pooling is weighed against the advantage of one-step determination and a decision is made on a case-by-case basis.

Search for variant (mutation discovery)

Single nucleotide polymorphism discovered using various technologies that includes array-based methods, denaturing HPLC, mass spectroscopy, denaturing gradient capillary electrophoresis and enzymatic mismatch cleavage (Comai and Henikoff 2006) [12]. In practical, the SNP discovery method must be both robust and cost effective and give the highest density of induced point mutations yet reported for TILLING a diploid species is about one mutation/250 kb (Greene, *et al.*, 2003) [20]. Screening several thousand mutant individuals will likely be required to ensure a high probability of identifying at least one deleterious mutation. Because of this limitation, whole genome scanning methods such as SNP-chips and polony based sequencing are too error-prone to be cost-competitive at the present time. Targeted scanning methods allow screening resources to be spent only on SNP discovery in candidate genes and thus provide a large cost savings over whole genome methods. Sanger sequencing, denaturing gradient capillary electrophoresis (DGCE), denaturing HPLC and enzymatic mismatch cleavage have all been used as SNP discovery methods for reverse-genetic screens (Colbert, *et al.*; 2001; McCallum, *et al.* 2000b; Slade and Knauf, 2005; Wienholds, *et al.* 2003) [11, 35, 46, 60]. The most common method used for TILLING has been enzymatic mismatch cleavage and resolution on polyacrylamide gels to detect the cleaved fragments (Figure 1).

In a typical reaction, a ~1.5-kb gene target is amplified by PCR with gene specific primers. Primers are end-labeled with fluorescent dyes for downstream visualization. The forward primer is 5'-end-labeled with a fluorescent dye that is detected at 700 nm (IRDye 700) and the reverse primer is 5'-end-labeled with the IRDye 800 nm (Till, *et al.*, 2006) [54]. After PCR, the amplified products are denatured and annealed, generally by heating and cooling to create heteroduplexes between wild-type and mutant DNA strands. Amplified products are incubated with an endonuclease such as CEL1 or celery juice extract (Oleykowski, *et al.* 1998) [39]. CEL1 is a member of the S1 nuclease family of single strand-specific nucleases which cleaves the 3' side of mismatched DNA where the heteroduplex between the wild-type and the mutant strands of DNA loops out while homoduplexes are remains intact (Till, *et al.*, 2004) [53]. Cleaved products are visualized using denaturing polyacrylamide gel electrophoresis and a gel readout platform such as automated sequencing gel apparatus LI-COR DNA analyzer. Gel images are analyzed with the aid of a standard commercial image- processing program. The lanes that have a mutation in the pool, a band will be visible below the wild type band on 'IR dye 700' infra-red dye image. A counterpart band will be visible in the same line on the 'IR dye 800' infra-red dye from the complementary DNA strand.

The sum of the length of two counterpart bands is equal to the size of amplicon, which distinguish mutation from amplification artifacts. It is noteworthy that the choice of enzyme and readout platform can potentially affect the optimal level of sample pooling. After detection of a mutation in a pool, the individual DNA samples in the pool are screened again to find out which of the eight pooled samples from the crossed population has the mutation. Once mutations are discovered, they are sequenced to determine the precise change in base sequence. An important advantage of the mismatch cleavage system is that the location of each mutation is determined within a few nucleotides, unlike methods such as denaturing HPLC, which can detect a

mismatch but does not identify where it lies in the sequence. By pinpointing the location of the putative mutation, the mismatch cleavage method allows for confident identification of each mutation, whether heterozygous or homozygous, with a single sequencing run, priming with the nearer of the amplifying primers.

TILLING has potential application to examine both natural as well as induced variation. It is suitable for most plants and enables the identification of single-base-pair allelic variation

in candidate genes in a high-throughput manner. TILLING has several benefits over other techniques such as single-strand conformation polymorphism (SSCP), denaturing gradient gel electrophoresis (DGGE) which is used to detect single-base pair polymorphisms (DeFrancesco and Perkel, 2001) [16]. Similarly, this technique also superior over the Array based hybridization techniques (Borevitz, *et al.*, 2003) [5] and RNAi-based gene silencing (Sabetta, *et al.*, 2011) [42] in identifying SNPs in candidate genes.

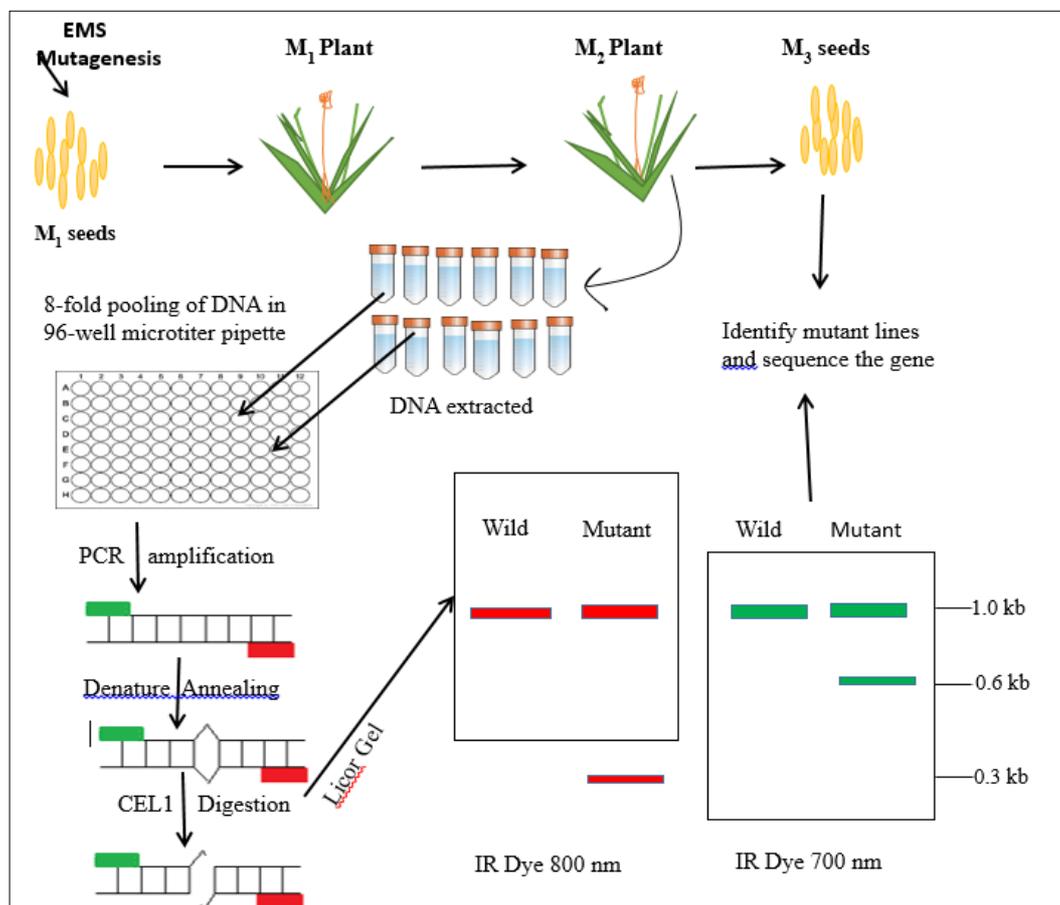


Fig 1: TILLING overview. Seeds are mutagenized and grown into M_1 plants, which are allowed to self fertilize. Leaf tissue is harvested from the M_2 plants for DNA extraction. The M_2 self fertilizes and the resulting M_3 seed is saved for mutant analysis. DNA from eight M_2 plants is pooled into a well on a 96-well plate. The target region is amplified by PCR with gene-specific primers that are end-labeled with fluorescent dyes. Following PCR, samples are denatured and annealed to form heteroduplexes that become the substrate for enzymatic mismatch cleavage.

Cleaved bands representing mutations or polymorphisms are visualized using denaturing polyacrylamide gel electrophoresis. Plants with mutations predicted to affect protein function can be carefully analyzed for phenotypic alterations.

Eco-TILLING (Modified TILLING)

Eco-TILLING is modified TILLING method that mine natural polymorphisms and reported in 2004 from work in *Arabidopsis thaliana* (Comai, *et al.* 2004) [13]. Eco-TILLING is similar to TILLING, except that it identifies natural genetic variation as opposed to induced mutations in TILLING method. Several crop species are not amenable to chemical mutagenesis; so, Eco-TILLING can assist in the discovery of natural variants and their putative gene function. Its main aim is to discover single nucleotide polymorphisms and haplotyping of gene in natural population. The principle of Eco-TILLING is alike to that of TILLING where it also depends on the enzymatic cleavage of heteroduplexed DNA with a single strand specific nuclease followed by detection through automated sequencing gel apparatus LI-COR DNA analyzer.

Eco-TILLING is suitable for the identification of natural variance within populations or germplasm without mutagenesis (Comai, *et al.*, 2004) [13]. It identify DNA

polymorphism in the form of SNPs, small insertions and deletions (InDels), haplotypes in candidate genes and variation in microsatellite (SSR) repeat number (Comai, *et al.*, 2004 and Till, *et al.*, 2006) [13, 54]. This is cost effective because only one individual for each haplotype need to be sequenced (Simsek and Kacar, 2010) [44] and applicable to any organism i.e., heterozygous or polyploidy and has practical application in the searching for resistance to new race of pathogen to create genetic variability (Kurowska, *et al.*, 2011) [30].

Sequencing based allele mining

Sequencing based allele mining involves PCR amplification of alleles of a gene in different genotypes and then DNA sequencing to recognize nucleotide variance in the alleles. The procedure includes collection of sample, DNA extraction, primer design, PCR amplification and sequencing and identification of allelic variation. Sequencing based allele mining would help to analyze individual haplotype structure

and diversity and such analysis is expected to simplify genetic association studies in plants. With the development of new generation sequencing (NGS) technologies, sequencing at lower cost would soon become possible. Technological developments also continue to increase sample throughput, which will facilitate large-scale genotyping of germplasm resources. Therefore, sequencing based-allele mining would result in faster generation of allelic data at a cheaper cost. It is very effective method to expand the rice blast R gene source and manage damaging blast disease (Ashkani, *et al.*, 2015) [3]. It would help to recognize the effect of mutations on gene structure and the sequences are analyzed for the location of point mutations or SNPs and insertions or deletions (InDel) to construct haplotypes. Unlike TILLING and Eco-Tilling, it does not require much sophisticated equipment or involve tedious steps (Ramkumar, *et al.*, 2010) [41].

Association Mapping-Based Allele Mining

Association mapping-based allele mining depicts more fine advances discovery of allelic variations. The strategy is used to establish regions of the genome associated with critical phenotypes by association or linkage-disequilibrium mapping (Kumar, *et al.*, 2016) [27]. The approach relies on the assumption that alleles responsible for a phenotype, along with the markers which flank the locus, are inherited as a block. DNA markers have been assisting this strategy as a means to identify useful alleles in the vast reservoirs of genetic diversity.

Importance of Allele Mining In Crop Improvement

The continuous improvement of crop plants is essential for agriculture in the coming decades and relies on the use of genetic variability through breeding. Thus, progress in plant breeding has been made possible by the accumulation of desirable alleles from the vast plant genetic resource existing worldwide. The rapid accumulation of sequence data and expression data in genomic databases has made genetic prospecting possible through high throughput PCR strategy for beneficial alleles of key genes from a wide range of species conferring resistance to biotic and abiotic stresses, greater nutrient use efficiency, enhanced yield and improved quality (Latha *et al.*, 2004) [32]. Although most mutations are deleterious, approximately 0.1% of mutations lead to alteration in gene function that may be essential for the survival of the plant. Hence, allele mining helps in detecting the naturally occurring allelic variation in candidate genes and tracing their evolution. Allele mining exploits the DNA sequence of one genotype to isolate useful alleles from related genotypes of the species. Also helps in identification of new haplotypes and development of allele-specific markers for use in marker-assisted selection (MAS). It can also provide insight into the molecular basis of novel trait variations and identify the base sequence changes associated with superior alleles.

Resistance to biotic stresses

Developing genetic resistance to pathogens is a main aim for plant breeding. Natural variability issued from the crop wild relatives constitutes a large reservoir of such resistances but methods have yet to be developed in order to efficiently tap for alleles of interest. In agriculture, biotic stress is a major cause of pre- and postharvest losses. The genetic basis of the defense mechanisms against biotic stresses is stored in the plant's genetic code. Plant genomes encode hundreds of biotic stress resistance genes. However, we have just started to

uncover the molecular mechanisms and networks controlling biotic stress resistance in crop plants. In several crops, the translation initiation factors eIF4E are susceptibility factors to many RNA viruses and allelic variants from those genes are a source of genetic resistances and transcriptomic data can be mined for resistance alleles of a target gene. (Lebaron, *et al.*, 2016) [33]. A group of researchers isolated a new eIF4E resistance allele against Potato Virus-Y using RNAseq data from an accession related to the cultivated tomato. It can be extended to mine for new resistances alleles, by looking at the many susceptibility host genes on which the virus rely to successfully infect plants (Lebaron, *et al.*, 2016) [33]. A large-scale allele-mining was done for a set of 1,320 bread wheat landraces from a database of 16,089 accessions, using the focused identification of germplasm and isolate 7 resistance alleles of previously undescribed *Pm3* alleles the powdery mildew resistance. Lines with new functional specificities were further tested using virus-induced gene silencing to find out if only the *Pm3* allele is involved in the observed resistance, or if other *Pm* genes are additionally present (VIGS, Scofield *et al.*, 2005) [43]. From above said set of accessions, 56 landraces were found completely resistant to at least one powdery mildew *Pm3* isolate (Bhullar, *et al.* 2009) [4]. One important way of yield security is the prevention of losses due to diseases and pests, which will have to be achieved by resistance breeding rather than by chemical pest control. Race-non-specific pathogen resistance is of high importance to plant breeders due to its expected durability. However, it is usually controlled by quantitative trait loci (QTL) and therefore difficult to handle in practice. In a study, 212 spring barley accessions were assembled from worldwide collections to mine the candidate genes for race-NR to the powdery mildew fungus *Blumeria graminis* f.sp. *Hordei* (Bgh). This led to the identification of 11 associated genes with converging evidence for an important role in race-NR and among assembled accessions, 30 susceptible and 35 resistant accessions were identified (Spies *et al.*, 2012) [47]. The high level of variability in the blast pathogens, necessitate the isolation of many resistance genes as resources to defend against various pathogens to develop durable blast resistance rice varieties (Datta, *et al.*, 2002) [15]. Allele mining approaches have been intended to identify superior alleles of rice blast resistance genes such as *Pita* (Huang, *et al.*, 2008; Ramkumar, *et al.*, 2010; Wang, *et al.*, 2008a; Yang, *et al.*, 2008) [23, 41, 58, 61], *Pikh* (Ramkumar, *et al.*, 2010) [41], *Pi54* (Kumari, *et al.*, 2013) [29], and *Pi-2* (Hittalmani, *et al.*, 2013) [22] from different cultivated rice varieties and wild species. Allelic variants of the broad-spectrum blast resistance gene, *Pi9* have been analyzed in 338 Indian rice landraces that are phenotyped for rice blast (Imam, 2016) [24]. The *Pi9* locus contains at least six known resistance genes specific to the fungal pathogen *M. oryzae* and three R-genes from this locus (*Pi9*, *Pi2*, and *Piz-t*) have been cloned (Qu, *et al.*, 2006; Zhou, *et al.*, 2006) [40, 65].

Tolerance to abiotic stresses

Abiotic stress has main effect on the crop productivity worldwide, reducing average yields for major crop plants. These abiotic stresses are interconnected as osmotic stress, resulting in the disruption of ion distribution and homeostasis in the cells of tissues (Kumar, 2013) [28]. Abiotic stress such as cold, drought, salt, and heavy metals mostly influences plant growth and crop productivity. A PCR-based allele mining was successfully performed on identified accessions of rice and related germplasm. The international rice genome sequencing

project sequence the genome of *Oryza sativa* L cv. Nipponbare will make allele mining possible for all genes of rice and possibly related cereals. Allele mining was based on three important rice genes conferring stress tolerance, which is a calmodulin gene encoding a signal-transduction pathways centered on calcium ion, a gene encoding a late embryogenesis-associated protein, and salt-inducible rice gene produced as part of the primary response to salt and drought tolerance stress (Latha, *et al.* 2004) [32]. DREB is a transcription factor produced under osmotic stresses such as salt, moisture stress and freezing. Allele mining for DREB2A gene associated with drought tolerance in Indian rice collections' was carried out in rice genotypes with the objective of identifying desirable alleles for DREB2A, if any, and validating their expression under moisture deficit stress. In a study, allele mining was conducted for a set of 96 barley genotypes using Eco-TILLING analyzing approximately 1.5 million base pairs in barley a total number of 94 verified unique haplotypes were identified in 18 amplicons designed for 9 genes. Overall, 185 single nucleotide polymorphisms (SNPs) and 46 insertions/deletions (INDELs) were found out (Cseri, *et al.* 2011) [14]. An integrated database developed from allele mining studies contains information about the genes, genotypes, and haplotypes and natural variation in the candidate genes providing a significant asset for further mapping studies dealing with the traits. Two loci, namely *PSTOL1* and *PupK20-2* reported for low P tolerance in rice genotypes. The discovery of novel SNPs across both *PSTOL1* and *PupK20-2* suggests the existence of novel haplotypes in genotypes adapted to acidic soil conditions (Yumnam, *et al.*, 2017) [64]. Low temperature stress is one of the major limiting factors affecting productivity of many crops in higher altitudes. A sequence analysis of 20 rice genotypes DREB1A and DREB1B, are two transcription factors that have been reported to play key regulatory role in low temperature tolerance. So far, data concerning the physiological role and regulation of DREB1 protein in different genetic background are very limited therefore it is to be expected that they will be studied extensively in the near future (Challam, *et al.*, 2015) [7].

Greater nutrient use efficiency

Increasing population levels and per capita consumption will require further significant gains in production to meet food and non-food demand, but such gains will also impact the environment (Tilman, *et al.*, 2002) [55]. Nitrogen (N) is the most important plant nutrient and is most commonly applied element to agricultural crops. In many cases nitrogen fertilizer is the main input cost for farmers, and in addition, is a major pollutant from agricultural activity. Most plants only take up 50% or less of the applied N, and most of the remainder is lost to leaching into ground water and waterways, or volatilized into the atmosphere. Reducing nitrogen inputs while maintaining high seed yield levels is necessary to sustain agricultural production in the future. Hence, improving the nitrogen responsiveness of crops is crucial for food security and environment safety, and breeding for Nitrogen use efficient (NUE) crop varieties. Reverse genetics approaches was used to examine allelic variation in two key N metabolism genes in the genomes of 44 genetically diverse sorghum genotypes and identified a nitrate reductase and a glutamate synthase gene (NADH-GOGAT) that were under balancing selection in improved sorghum cultivars (Diatloff, 2017) [17]. Limiting nitrogen supply universally resulted in less shoot and root growth, increased root weight ratio, reduced

tissue nitrate and nitrogen concentrations, and reduced NADH-GOGAT expression. It is concluded that reverse genetics that take advantage of rapidly expanding genomic databases contributes towards a systematic approach for developing N efficient crops.

Enhanced yield

To meet the food requirements several billion people living on this planet, the production of high-quality food must increase with reduced input usage, but this accomplishment will be particularly challenging the global environmental scenario. Plant breeders need to focus on the traits with the maximum potential to increase crop production. Hence, new technologies must be developed to achieve this goal and accelerate breeding through improving genotyping and phenotyping methods and by effective utilization of the available genetic diversity in breeding germplasm (Tester and Langridge, 2010) [51]. Mining elite alleles contributing for grain yield is of utmost importance for the improvement of cultivated crops and selection for market demand. Association mapping study was performed for grain traits (grain length, grain width, grain thickness, grain length to width ratio, and 1000-grain weight) on 628 rice cultivars using 262 SSRs. In total, 2953 alleles were detected with an average of 11.3 alleles per locus. The accessions containing elite alleles for grain traits mined in this study could be used for breeding rice cultivars and cloning the candidate genes (Edzesi, *et al.*, 2016) [18]. Promising alleles of six key genes *DEP1*, *Ghd7*, *Gn1a*, *GS3*, *qSW5* and *sd1* influencing the yield were uncovered by targeted resequencing of the 200 rice genotypes. From this study, 225 novel alleles were identified and only 91 were beneficial alleles that showed significant association with the yield traits (Vemireddy, 2019) [56]. Association mapping based on linkage disequilibrium is an effective approach for dissecting the inheritance of complex traits. An association mapping was performed for yield traits on 172 Upland cotton (*Gossypium hirsutum* L.) cultivars using polymorphic simple sequence repeat (SSR) markers and seven yield traits such as seed cotton yield, lint yield, boll number per plant, boll weight, lint percentage, lint index and seed index were identified. Furthermore, elite alleles for yield traits were also mined. The association mapping based allele mining data can provide useful information for further understanding the genetic basis of yield traits, and facilitate high-yield breeding by molecular design in crops (Li, *et al.*, 2017) [34].

Improved quality

Plants are the ultimate source of nutrients for humans and livestock. The first challenge for a breeder is to determine which trait is most importance. In a subsistence situation, focusing on content of key nutrients that are deficient in a diet that may lack quantity and diversity may be of utmost importance. A genome-wide association mapping (GWAS) was performed on a selected sample of 462 rice accessions using 262 simple sequence repeat markers for eating and cooking quality in rice. Researchers identified 10 loci and 27 favorable alleles for gelatinization temperature, gel consistency and amylose content. There existed multiple gene loci underlying grain eating and cooking qualities in the natural population of rice. Among the 10 loci detected for grain GT, GC and AC, 6 loci were newly detected. Within the novel loci, the favorable allele RM3600-90bp on chromosome 9 could significantly reduce gelatinization temperature, RM6712-95bp on chromosome 3 could significantly increase

gel consistency, and RM6327-230bp on chromosome 11 could significantly reduce amylose content in hybrid japonica rice mixed rice samples (Wang, *et al.*, 2019) [57]. Chiapparino *et al.* 2006 [9] reported Bmy1 (β amylase-I) allele encoding starch break down enzyme in barley through allele mining. In rice, aBadh2 allele conditioning the fragrance of rice grain was discovered by Amarawathi *et al.* in 2008 [1].

Application of Allele Mining

Allele identification

Allele mining largely applied for the discovery of beneficial alleles through 'mining' the gene of interest from diverse genetic resources. Identification of previously undescribed functional alleles of *Pm3* locus in wheat against powdery mildew disease (Bhullar, *et al.*, 2009) [4].

Allelic variation characterization

Allelic diversity existed in gene banks are characterized after mining of novel descriptors of germplasm of crop species utilizing the molecular markers and identify the nucleotide sequences changes associated with trait of interest.

Haplotype identification and analyzing haplotype diversity

Haplotype is a set of single nucleotide polymorphisms present on a chromosome. Therefore, it can be potentially employed in the identification of nucleotide sequence variation in candidate gene associated with phenotypic variation for a trait. Through this, one can evaluate the frequency, type and the extent of occurrence of new haplotypes and the resulting phenotypic changes. Using Genome-wide association study (GWAS), haplotype-based allele mining was performed in rice and determined the loci controlling the glutinous endosperm and culm length traits (Ogawa, *et al.* 2018) [38].

Development of Allele-specific markers

Identification of sequence variation with observed phenotypic changes will pave the way to develop allele-specific marker assay for precise introgression of the identified desirable alleles into suitable genetic background. Such as, comparison of nucleotide sequences of Waxy gene codes for a granule-bound starch synthase revealed the presence of five different alleles with the observed phenotypic variations (Mikami, *et al.*, 2008) [37].

Allelic synteny and evolutionary relationship

Sequence data obtained from allele mining studies assist in finding syntenic relationship of the cultivated species with their wild relatives and other crop species of related taxa. On the basis of syntenic allele sequence information from wheat, alleles for aluminum tolerance in rye and agronomically superior alleles in *Phaseolus vulgaris* and other grasses were isolated (Fontecha, *et al.*, 2007) [19]. Genes/QTL controlling quantitative resistance to the blast pathogen, *Magnaporthe grisea* showed high similarity in their sequence and as well as genetic/physical position in both rice and barley, suggesting a common evolutionary origin for this resistance gene (Chen, *et al.*, 2003) [8].

Promoter mining

It can be effectively utilized in expression study as well as gene prediction through promoter mining. For example: The expression of some genes namely, Xa 13 conferring bacterial blight resistance gene in rice (Chu, *et al.*, 2006) [10] and GIF1 (incomplete grain filling gene) in rice (Wang, *et al.*, 2008) [59]

and maize have been studied through changes in nucleotide sequence changes in the promoter region of those genes.

Conclusion

The use of genetic diversity is limited due to the resources which are at hand for characterization of all these lines. The sheer number of germplasm collections should be screened at phenotypic and genotypic level through molecular characterization to draw valid results and conclusions. There should be efficient technique such as FIGS (Focused identification of germplasm strategy) to handle the vast number of germplasm lines to construct core collection sets and this reduces the duplication of germplasm. Allele mining can be effectively used for discovery of superior alleles, through 'mining' the gene of interest from germplasm. It can also provide insight into genome wide association studies and molecular basis of novel trait variations and identify the nucleotide sequence changes associated with beneficial alleles. In addition, allele mining made important emphasis to determine the rate of evolution of alleles; allelic similarity and dissimilarity at a candidate gene, allelic relationship with other members of the family and introgression of novel alleles through Marker Assisted Selection paving the allele mining.

Prospects and Future Consideration for Allele Mining

Allele mining possesses good potential to be applied in molecular plant breeding. With the ever-increasing sequence data in Gene Bank and ever-expanding crop gene banks, it is highly essential to develop novel and efficient mining strategies to screen Gene Bank collections more efficiently for DNA sequence variation and the management of genome resources. Allele mining will provide a way to discover terminal heat stress tolerance genotypes in wheat through sequence-based characterization of wheat genetic resources collected from high temperature regions. Drought and mineral toxicity are the major stresses in the Indian context, which decline crop production, so mining for these stress tolerant alleles or genes is a large challenge to combat these stress. The success of allele mining mainly depends on the type of genetic materials used for screening and should be as diverse as possible. To this end, wild relatives and local landraces are used because they are reservoirs of useful alleles hidden in their phenotype (Tanksley, *et al.*, 1996) [49]. Hence, we need to construct more polished sampling methods for the efficient selection of genotypes to distinguish valuable and novel alleles underlying particular traits.

References

1. Amarawathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR *et al.* Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). Mol. Breeding. 2008; 21:49-65.
2. An G, Jeong DH, Jung, KH, Lee S. Reverse genetic approaches for functional genomics of rice. Plant Mol Biol. 2005; 59:111-12.
3. Ashkani S, Yusop MR, Shabanimofrad M, Aziz GP, Latif MA. Molecular progress on the mapping and cloning of functional genes for blast disease in rice (*Oryza sativa* L.): current status and future considerations. Critical reviews in biotechnology. Curr Issues Mol Biol. 2015; 17:57-74.
4. Bhullar NK, Streetb K, Mackayc M, Yahiaouia N, Keller B. Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles

- at the Pm3 resistance locus. Proceedings of the National Academy of Sciences. 2009; 106(23):9519-9524.
5. Borevitz JO *et al.* Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Research*. 2003; 13:513-523.
 6. Burch-Smith TM, Anderson JC, Martin GB, Dinesh-Kumar SP. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *Plant J*. 2004; 39:734-746.
 7. Challam C, Ghosh T, Rai M, Tyagi W. Allele mining across DREB1A and DREB1B in diverse rice genotypes suggest a highly conserved pathway inducible by low temperature. *J. Genet.* 2015; 94:231-238.
 8. Chen H, Wang S, Xing Y, Xu C, Hayes PM, Zhang Q *et al.* Comparative analyses of genomic locations and race specificities of loci for quantitative resistance to *Pyricularia grisea* in rice and barley. Proceedings of the National Academy of Sciences. 2003; 100:2544-2549.
 9. Chiapparino E, Donini P, Reeves J, Tuberosa R, Sullivan DM. Distribution of β -amylase I haplotypes among European cultivated barleys. *Mol. Breeding*. 2006; 18:341-54.
 10. Chu Z *et al.* Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* 2006; 20(10):1250-1255.
 11. Colbert T *et al.* High-throughput screening for induced point mutations. *Plant Physiol.* 2001; 126:480-484.
 12. Comai L, Henikoff S. TILLING: practical single-nucleotide mutation discovery. *The Plant Journal*. 2006; 45:684-694.
 13. Comai L *et al.* Efficient discovery of DNA polymorphisms in natural populations by Eco-TILLING. *Plant J*. 2004; 37 778-786.
 14. Cseri A *et al.* Allele mining and haplotype discovery in barley candidate genes for drought tolerance. *Euphytica*. 2011; 181(3):341-356.
 15. Datta K, Baisakh N, Thet KM, Tu J, Datta S. Pyramiding transgenes for multiple resistance in rice against bacterial blight, yellow stem borer and sheath blight. *Theoretical and applied genetics*. 2002; 106:1-8.
 16. DeFrancesco L, Perkel JM. Lab consumer-in search of genomic variation. *Scientist-the Newspaper for the Science Professional*. 2001; 15:24-27.
 17. Diatloff E. Genome wide analysis and allele mining of diverse sorghum genes involved in nitrogen use efficiency. MPhil Thesis, School of Agriculture and Food Sciences, The University of Queensland, 2017.
 18. Edzesi WE, Dang X, Liang L, Liu E, Zaid IU, Hong D *et al.* Genetic diversity and elite allele mining for grain traits in rice (*Oryza sativa* L.) by association mapping. *Frontiers in Plant Science*. 2016; 7(787):1-13.
 19. Fontecha G, Silva-Navas J, Benito C, Mestres MA, Espino FJ, Hernandez-Riquer MV *et al.* *Theor Appl Genet*. 2007; 114:249-60.
 20. Greene EA *et al.* Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*. 2003; 164:731-740.
 21. Henikoff S, Comai L. Single-nucleotide mutations for plant functional genomics. *Annu Rev Plant Biol*. 2003; 54:375-401.
 22. Hittalmani S, Kahani F, Dhanagond SM, Rao AM. DNA marker characterization for allele mining of blast and bacterial leaf blight resistant genes and evaluation for grain yield. *African Journal of Biotechnology*. 2013; 12:2331-2340.
 23. Huang CL, Hwang SY, Chiang YC, Lin TP. Molecular evolution of the Pi-ta gene resistant to rice blast in wild rice (*Oryza rufipogon*). *Genetics*. 2008; 179:1527-1538.
 24. Imam J, Mandal NP, Variar M, Shukla P. Allele mining and selective patterns of pi9 gene in a set of rice landraces from India. *Frontiers in Plant Science*. 2016; 7:1-9.
 25. Johnson RC. Gene banks pay big dividends to agriculture, the environment and human welfare. *PLoS Biol*. 2008; 6:e148.
 26. Koornneef M, Dellaert LW, van der Veen JH. EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutat Res*. 1982; 93:109-123.
 27. Kumar A, Babu BK, Yadav S, Agrawal PK. Allele mining for resistance gene analogs (RGAs) in crop plants: A special emphasis on blast resistance in finger millet (*Eleusine coracana* L.). *Indian Journal Genetics*. 2016; 76:1-9.
 28. Kumar M. Crop plants and abiotic stresses. *J Biomol Res Ther*. 2013; 3:e125.
 29. Kumari A, Das A, Devanna B, Thakur S, Singh P, Singh N *et al.* Mining of rice blast resistance gene Pi54 shows effect of single nucleotide polymorphisms on phenotypic expression of the alleles. *European Journal of Plant Pathology*. 2013; 137:55-65.
 30. Kurowska M *et al.* *J App Genet*. 2011; 52:371-390.
 31. Kusaba M. RNA interference in crop plants. *Curr Opin Biotechnol*. 2004; 15:139-143.
 32. Latha R, Rubia L, Bennett J, Swaminathan MS. Allele mining for stress tolerance genes in *Oryza* species and related germplasm. *Mol. Biotech*. 2004; 27:87-96.
 33. Lebaron C, Rosado A, Sauvage C, Gauffier C, German-Retana S, Moury B *et al.* A new eIF4E1 allele characterized by RNAseq data mining is associated with resistance to PVY in tomato albeit with a low durability. *J Gen. Virol*. 2016; 97(11):3063-3072.
 34. Li C, Dong N, Fu Y, Sun R, Wang Q. Marker detection and elite allele mining for yield traits in Upland cotton (*Gossypium hirsutum* L.) by association mapping. *The Journal of Agricultural Science*. 2017; 155(4):613-628.
 35. McCallum CM, Comai L, Greene EA, Henikoff S. Targeted screening for induced mutations. *Nat Biotechnol*. 2000a; 18:455-457.
 36. McCallum CM, Comai L, Greene EA, Henikoff S. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiol*. 2000b; 123:439-442.
 37. Mikami I, Uwatoko N, Ikeda Y, Yamaguchi J, Hirano HY, Suzuki Y *et al.* *Theor Appl Genet*. 2008; 116:979-989.
 38. Ogawa D *et al.* Haplotype-based allele mining in the Japan-MAGIC rice population. *Scientific Reports*. 2018; 8:4379.
 39. Oleykowski CA *et al.* Mutation detection using a novel plant endonuclease. *Nucleic Acids Research*. 1998; 26:4597-4602.
 40. Qu S *et al.* The broad-spectrum blast resistance gene Pi9 encodes a nucleotide-binding site-leucine-rich repeat protein and is a member of a multi gene family in rice. *Genetics*. 2006; 172:1901-1914.
 41. Ramkumar G *et al.* Identifying novel alleles of rice blast resistance genes *Pikh* and *Pita* through allele mining. *Plant breeding*, 2010.

42. Sabetta W, Alba V, Blanco A, Montemurro C. SunTILL: a TILLING resource for gene function analysis in sunflower. *Plant Methods*. 2011; 7:20.
43. Scofield SR, Huang L, Brandt AS, Gill BS. Development of a Virus-Induced Gene-Silencing system for hexaploid wheat and its use in functional analysis of the Lr21-mediated leaf rust resistance pathway. *Plant Physiology*. 2005; 138:2165-2173.
44. Simsek O, Kacar YA. Discovery of mutations with tilling and ecotilling in plant genomes. *Science and Research Essays*. 2010; 5:3799-3802.
45. Singh BD. A text book of plant breeding. Kalyani publishers, 2005, 171-172.
46. Slade AJ, Knauf VC. TILLING moves beyond functional genomics into crop improvement. *Transgenic Res*. 2005; 14:109-115.
47. Spies A *et al.* Allele Mining in Barley Genetic Resources Reveals Genes of Race-Non-Specific Powdery Mildew Resistance. *Frontiers in Plant Science*. 2012; 1:1-23.
48. Tan L, Li X, Liu F, Sun X, Li C, Zhu Z *et al.* *Nat Genet*. 2008; 40:1360-1364.
49. Tanksley S *et al.* Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. *Theoretical and applied genetics*. 1996; 92:213-224.
50. Tanksley SD, Mc Couch SR. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science*. 1997; 277:10630-1066.
51. Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. *Science*. 2010; 327(5967):818-22.
52. Till BJ *et al.* Large-scale discovery of induced point mutations with high throughput TILLING. *Genome Research*. 2003; 13:524-530.
53. Till BJ, Burtner C, Comai L, Henikoff S. Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Research*. 2004; 32:2632-2641.
54. Till BJ, Zerr T, Bowers E, Grene EA, Comai L, Henikoff S *et al.* High throughput discovery of rare human nucleotide polymorphisms by Eco-TILLING. *Nucleic Acids Research*. 2006; 34:1479.
55. Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S. Agricultural sustainability and intensive production practices. *Nature*. 2002; 418(6898):671-677.
56. Vemireddy LR *et al.* Uncovering of natural allelic variants of key yield contributing genes by targeted resequencing in rice (*Oryza sativa* L.). *Sci Rep*. 2019; 9:8192.
57. Wang H *et al.* Favorable alleles mining for gelatinization temperature, gel consistency and amylose content in *Oryza sativa* by association mapping. *BMC Genet*. 2019; 20:34.
58. Wang J *et al.* Analysis of *Pin a* and *Pin b* alleles in the micro-core collections of Chinese wheat germplasm by Ecotilling and identification of a novel *Pin b* allele. *Journal of Cereal Science*. 2008a; 48:836-842.
59. Wang X, Jia Y, Shu QY, Wu D. *Phytopathology*. 2008; 98:1305-1311.
60. Wienholds E, van Eeden F, Kusters M, Mudde J, Plasterk RH, Cuppen E *et al.* Efficient target selected mutagenesis in zebrafish. *Genome Research*. 2003; 13:2700-2707.
61. Yang S, Gu T, Pan C, Feng Z, Ding J, Hang Y *et al.* Genetic variation of NBS-LRR class resistance genes in rice lines. *Theoretical and Applied Genetics*. 2008; 116:165-177.
62. Yu J, Bai G, Cai S, Dong Y, Ban T. New Fusarium head blight resistant sources from Asian wheat germplasm. *Crop Sci*. 2008; 48:1090-1097.
63. Yugandhar G, Nishant BA, Kailash C, Singh MN, Hemantaranjan A. Allele mining—an approach to discover allelic variation in crops. *Jour Pl Sci Res*. 2017; 33(2):167-180.
64. Yumnam J, Rai M, Tyagi W. Allele mining across two low-P tolerant genes PSTOL1 and PupK20-2 reveals novel haplotypes in rice genotypes adapted to acidic soils. *Plant Genetic Resources*. 2017; 15(3):221-229.
65. Zhou B *et al.* The eight amino-acid differences within three leucine-rich repeats between Pi2 and Piz-t resistance proteins determine the resistance specificity to *Magnaporthe grisea*. *Mol. Plant Microbe Interact*. 2006; 19:1216-1228.