ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of DEVELOPMENT RESEARCH

International Journal of Development Research
Vol. 4, Issue, 2, pp. 300-305, February, 2014

Full Length Research Article

ALLELE MINING IN CROP IMPROVEMENT

[1,*]Amaranatha Reddy, M., [1]Vilas Bhagwat Gambhire and [2]Rajeswar Reddy, T.

[1]Department of Plant Breeding and Genetics, COH, KAU, Vellanikkara, India
[2]Department of Plant Pathology, COH, KAU, Vellanikkara, India

ARTICLE INFO

Article History:
Received 07th November, 2013
Received in revised form
11th December, 2013
Accepted 28th January, 2014
Published online 21st February, 2014

Key words:
Accumulation,
Deoxy-ribonucleic acid,
Granule-Bound Starch Synthase,
Allele-specific markers.

ABSTRACT

Development of superior and high yielding varieties made possible by accumulation of beneficial alleles from vast plant genetic resources existing worldwide. Still, a significant portion of these beneficial/ superior alleles were not used as these were left behind during evolution and domestication. Introducing novel alleles from wild crop plants to cultivated varieties have clearly demonstrated that certain alleles and their combinations potentially make dramatic changes in trait expression. Hence, the vast germplasm resources need to be relooked for novel alleles to further enhance the genetic potential of crop varieties for various agronomic traits. Alleles are alternative forms of a gene occupying a given locus on a chromosome. Allele mining exploits the deoxy-ribonucleic acid (DNA) sequence of one genotype to isolate useful alleles from related genotypes. It helps in tracing the evolution of alleles, identification of new haplotypes and development of allele-specific markers for use in marker-assisted selection. Allele mining is a way to find out the superior alleles from related genotypes. This is made possible as enormous sequence information is available in public databases as a result of sequencing of diverse crop genomes. It is important to use this genomic information for the identification and isolation of novel and superior alleles of agronomically important genes from crop gene pools to suitably organize for the development of improved cultivars. Alleles such as *Sh* (grain shattering), *Rc* (grain pericarp color), *Wx* for Granule-Bound Starch Synthase (GBSS) and *GS* (grain size) have led to significant improvements in rice. In allele mining, different softwares are used for identifying the nucleotide variation and prediction of amino acid changes which is responsible for encoding protein structure and functions. Some of the bioinformatics softwares used are ClustalW, DCPD, Fast PCR and Plant CARE. Allele mining can be effectively used for discovery of superior alleles are through 'mining' the gene of interest from diverse genetic resources. It can also provide insight into molecular basis of novel trait variations and identify the nucleotide sequence changes associated with superior alleles. It will help to trace the evolution of alleles, identification of new haplotypes and development of allele-specific markers for use in marker-assisted selection. Realizing the immense potential of allele mining, concerted allele mining efforts are underway in many international crop research institutes.

INTRODUCTION

Progress in plant breeding in terms of development of superior and high yielding varieties of agricultural crops is possible by accumulation of beneficial alleles from vast plant genetic resources existing worldwide. But still, a significant portion of these superior alleles cannot be used, because those alleles are left behind during evolution and domestication. This genetic variation existing in wild relatives and land races of crop plants and it can be used for development of agronomically superior cultivars. Introgressions of novel alleles occur from wild crop plants into cultivated varieties. It clearly demonstrated that certain alleles and their combinations potentially make dramatic changes in character expression when it moved to a suitable genetic background. Hence, the vast germplasm resources need to for novel alleles to further enhance the genetic potential of crop varieties for various agronomic traits. Enormous progress has been made in the last 15 years in depositing an amount of sequence information into GenBank. With rapid accumulation of sequence and expression data in various genomic databases accelerate discovery of new genes. It is easy to development of allele-specific markers. Based on gene and genome sequences,

***Corresponding author: Amaranatha Reddy, M.,** Department of Plant Breeding and Genetics, COH, KAU, Vellanikkara, India*

Polymerase Chain Reaction (PCR) strategies are use to isolate Superior alleles of genes from a wide range of species. This key alleles imparting resistance to biotic and abiotic stresses, greater nutrient use efficiency enhanced yield and improved quality. Using novel genomic tools, similar alleles responsible for a given trait and their variants in other genotypes can be identified. This is often referred to as 'dissection of naturally occurring variation at candidate genes/loci' or simply 'allele mining'. Identification of allelic variants from germplasm collections not only provides new germplasm for delivering novel alleles to targeted trait improvement but also categorizes the germplasm entries for their conservation (Latha *et al*., 2004).

Evolution of new alleles

It is alternative forms of gene. In other words, genes occur at similar loci of homologous chromosomes. Mutation is an evolutionary driving force which causes existing allelic diversity in any crop species. For creation of new alleles or causing variations in the existing allele and allelic combinations, generally mutations will occur in the genic regions of the genome either as Single Nucleotide Polymorphism (SNP) or as Insertion and Deletion (InDel) are important. The mutations in coding regions and/or regulatory regions may have tremendous effect on the phenotype by changing the encoded protein structure and/or function while those that occur in noncoding regions of a gene. It has been silent without any effect on the phenotype. Even though most of the mutations are deleterious, in genera l0.1% of the mutations are vital leading to changes will occur in gene function which may be highly necessary for the survival of the plant (Singh, 2005).

a) How mutation does these variations occur?

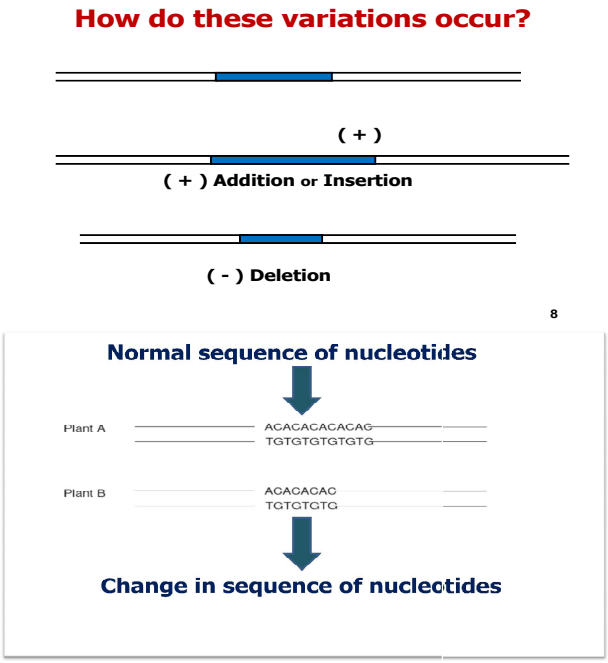

Fig.1. Insertion and deletion in nucleotide sequence

Mutation is defined as sudden heritable changes occur in DNA segment of individuals. This change will occur in normal sequence of nucleotides in individuals. For example, the long growth habit of wild rice is controlled by a single gene PROG1. The wild type allele is replaced by mutant allele (prog1) in most of the Oryza sativa cultivars, which inactivate the PROG1, gene function and leading to erect growth, greater grain number and higher grain yield in cultivated rice (Tan *et al*., 2008). This phenomenon is observed as the wild rice has an allele their expression or its effect inactivate in cultivated rice due to mutation. In other words change will occur in nucleotides.

b) How new alleles are formed?

Mutations will happen in both genomic region such as coding region as well as non-coding regions of gene. That creates heritable change in individuals.

What is allele mining?

Mining is extraction of any non-renewable or highly valuable resources such as Petroleum, natural gas, minerals, or even water etc. It is a finding of superior allele from the natural population. In other words introgression of novel or superior allele from wild relatives into cultivated one.

Allele mining in crop improvement

It is a strategy of finding valuable and unknown alleles at a known locus, is called as allele mining. These valuable and unknown novel alleles can be used for further crop improvement such as resistance to biotic and abiotic stresses, increase greater nutrient use efficiency, enhance yield in crops and improve quality such as cooking, protein, starch including human nutrition (Latha *et al*., 2004). Allele mining is conduct on specific genes that are involved in the particular mechanism of stress tolerance express by identify accessions. In Cultivated rice and its relatives it will give an opportunity to test the evolutionary range over which PCR-based allele mining can be successful. Nipponbare is a japonica cultivar that belongs to isozyme group VI of *O. sativa (1)*. For each gene, the Nipponbare allele will be most closely related to the alleles of other japonica cultivars and then progressively less closely related to alleles of (1) the isozyme groups I–V of *O. sativa,* (2) the other AA genome species, (3) the non-AA genome species of genus *Oryza,* (4) related grass genera such as *Porteresia,* and (5) the other cereals. Vaughan *(2)* describes the genus *Oryza* and related grasses (subfamily Oryzoideae), whereas Kellogg *(3)* summarizes the evolutionary relationships among the grasses in general, including the cereals (Glaszmann, (1987).

Disease resistance alleles in rice: Role of mutations in the promoter region of the candidate gene for xa13, a recessive bacterial blight resistance gene, imparting disease resistance in rice crop (Chu *et al*., 2006).

Table.1. Disease resistance alleles in rice crop

Sr.No.	Gene	Allele	Species	Reference
1	*Xa*	*Xa* 5,13, 21	*O.longistaminata*	(Kush *et al*., 1991)
2	*Pi*	*Pi* 5,9	*O. minuta*	(Sitch *et al*., 1989)

Approaches for allele mining

There are two major approaches. They are:

TILLING -based allele mining: It is nothing but a (Targeting Induced Local Lesions In Genomes), to determine variation in individual through artificially changed mutation (Comai *et al*., 2004). It is a powerful reverse genetics tool for functional genomics where knockout methodologies cannot be applied. Tilling allows the identification of allelic variation of target gene in a high-throughput manner. The use of the Tilling technique to survey natural variation in genes is called Ecotilling. Tilling make use of chemical mutagens to introduce random mutation. Seeds are mutagenized with EMS, which causes G/C-to-A/T point mutations (Nagy, *et al*., 2003). M_1 seeds are selfed to produce M_2 seeds. M_2 progeny from single seed descent are used for screening. For screening, DNAs are pooled eightfold to maximize the efficiency of mutation detection. PCR is performed using 5'-end-labeled gene-specific primers to target the desired locus, and heteroduplexes are formed by heating and cooling the PCR products. CEL I nuclease is used to cleave at base mismatches, and products representing induced mutations are visualized with denaturing polyacrylamide gel electrophoresis. The detailed procedure has been given by Till, *et al*. (2003).

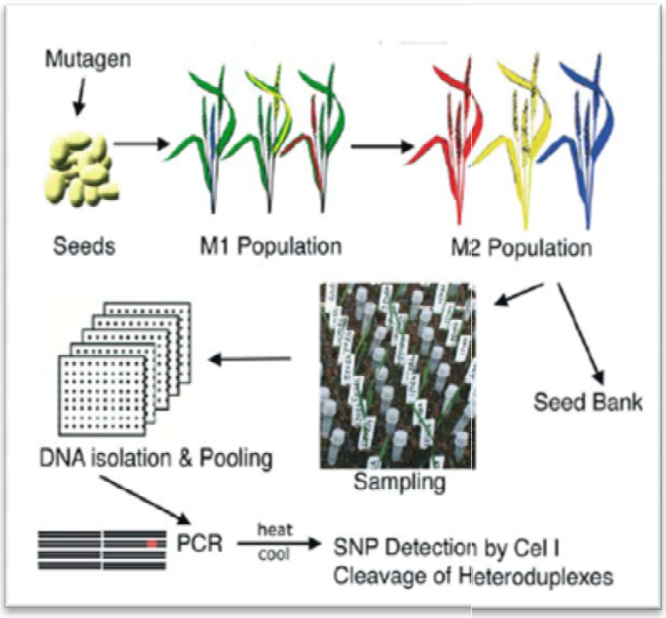

Fig.2. TILLING –based allele mining

Sequencing based allele mining: It is amplification of alleles in diverse genotypes through PCR and identify nucleotide variation by DNA sequencing. In other words multiplication of specific segment of deoxy-ribose nucleotides as well as, we can identify various allele among the cultivars through sequence based allele mining. Unlike Eco Tilling, sequencing-based allele mining does not require much sophisticated equipment or involve tedious steps, but involves huge costs of sequencing. A comparison between these two procedures is given below. Eco Tilling also depend on the enzymatic cleavage of heteroduplexed DNA (formed due to single nucleotide mismatch in sequence between reference means genome sequence available and test genotype) with a single strand specific enzymes use nuclease (i.e. Cel-1,S1nuclease) under specific conditions followed by detection through Li-Cor genotypers at point mutations, there will be a cleavage by the nuclease to produce two cleaved products whose sizes will be equal to the size of full length product.

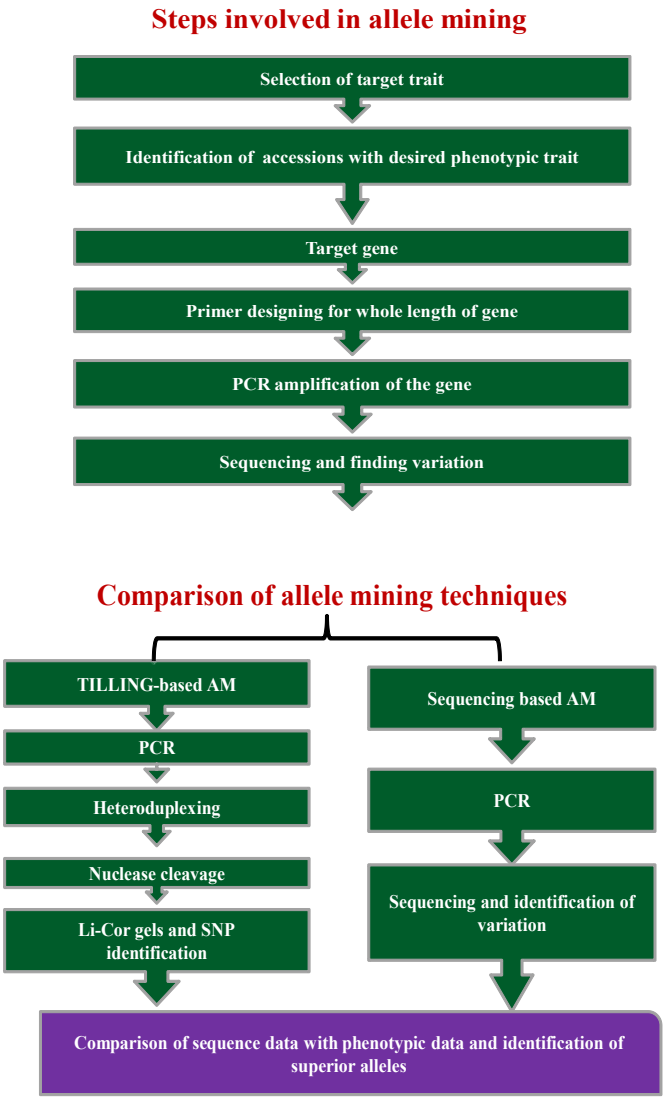

Fig.3. Steps involved in allele mining

The presence, type and location of point mutation or SNP will be confirmed by sequencing the amplicon from the test genotype that carry the mutation. TILLING and EcoTilling has been proposed cost effective approaches for haplotyping and SNP discovery, these techniques require more sophistication and involve several steps starting from making DNA pools of reference and test genotypes, specific conditions for efficient cleavage by nuclease, detection in polyacrylamide gels using Li- Cor genotyper and confirm through sequencing (Raghavan *et al*., 2007).

Next generation sequencing technologies in allele mining

a) NGS (Next generation sequencing): Sequencing is a technique which enables us to understand arrangement of this nucleotide in DNA molecules. In the last few years, 'massively parallel' methods have also emerged and lead to the development of 'next generation' sequencing platforms with increased throughput and accuracy. These methods used for resequencing, alignment of the sequence data and their comparison with reference genome. The first of this type was commercialized by 454 Life Sciences and this technique relied on pyrosequencing while eliminating the need for cloning. With this 454 sequencing platform, it is possible to produce

100 Mb of sequence with 99.5% accuracy and increase read length averaging over 250 bases (Margulies *et al*., 2005). Another massively parallel sequencer Illumina/Solexa genome analyzer has been developed and this is capable of sequencing one billion bases (1 Gb) of 30–40 base sequence reads in a single run in a short timer period of 3–4 days. In addition to the above, Applied Biosystems Inc (USA) has also developed a massively parallel sequencer using supported oligonucleotide ligation and detection system (SOLiD). This system, its shorter read length of about 25–35, can generate 2–3 Gb of sequences per run.

b) Different sequencing platforms

1. Illumina/Solexa genome analyzer: It is capable to sequencing one billion bases (1 Gb) of 30–40 base sequence reads in a single run in a short timer period of 3–4 days.

2. Supported oligonucleotide ligation and detection system (SOLiD): This system, is useful for to reads its shorter read length of about 25–35, can generate 2–3 Gb of sequences per run.

3. 454 sequencing analyzer: It is possible to produce 100 Mb of sequence with 99.5% accuracy and increased read length averaging over 250 bases.

Tools required for allele mining

Software tools are useful for handling the complex nucleotide data, prediction of functional or structural components of complex macromolecules, prediction of transcription factor binding sites, identification of sequence polymorphisms and to predict the amino acid changes which are responsible for changes in encoded protein structure and function.

Bioinformatics tools for allele mining: Allele mining requires various sophisticated bioinformatic tools viz., PLACE, plantCARE, TRANSFAC, JASPAR, MEME, Plantprom DB, DCPD, SCPD, BioEdit, ClustalW etc.,. These tools useful for sequence alignment means to compare our genome sequence to reference genome i.e, sequenced genome data.

Considerations for allele mining

Successful and efficient allele mining activity might be depended mainly on the existing genetic base and the availability of genome and gene sequences information of a particular crop species. As well as other important factors include availability of efficient and essay phenotyping techniques; genomic resources; availability of high throughput techniques for quick and easy generation of allelic data points; cost-effective sequencing platforms; efficient bio-informatic tools for identification of nucleotide variation and molecular marker construction technologies for Marker-Assisted selection (MAS). Considering these, a set of criteria for allele mining is detailed in the Table. Based on these criteria, allele mining can be initiated for any crop with sequence information and even for a related member of its family.

Sl.No.	Activities	Criteria for selection
1.	Germplasm identification	Reference collection representing maximum genetic diversity in a minimum possible number.
2.	Crop species	Availability of comparative map/ sequence colinearity
3.	Target trait	Traits/Characters of interest
4.	Phenotypic characterization	Correct phenotyping using efficient protocols in suitable environmental condition
5.	Primer design	Primer pairs with a minimum overlap of 100– 200 bp to maximize their utility for SNP detection and sequencing
6.	Sequencing analysis for detection of SNPs and InDels	Sequencing of single and clear amplicons Amplicons of genotypes with high trait intensity in desired direction

Applications of allele mining

Allele mining can be effectively used for discovery of superior alleles, through 'mining' the gene of interest from diverse genetic resources. It can also provide insight into molecular basis of novel trait variations and identify the nucleotide sequence changes associated with superior alleles. In addition, the rate of evolution of alleles; allelic similarity/dissimilarity at a candidate gene and allelic relation with other members of the family can also be study. Allele mining may also give way for molecular difference among related species, development of allele-specific molecular markers, for essay introgression of novel alleles through MAS or through genetic engineering (GE).Those applications as fallows.

Characterization of allelic diversity in gene banks

Identification and allelic variation that affects the plant phenotype is importance for the utilization of genetic resources in crop improvement. Potential of allele mining in Plant Genetic Resources (PGR) management applications, many of the international crop research institutes which are maintaining crop germplasm collections have initiate studies to characterize the allelic diversity of crop plants.

Table 2. Bioinformatics tools for allele mining

Sl. No.	Name	Utility	Available at	Reference
1	AGRIS	TF database	Arabidopsis.med.ohiostate.edu	(Davuluri *et al*., 2003)
2	FastPCR	Nucleotide sequence analysis	-	(Kalender, 2009)
3	BioEdit	Nucleotide sequence analysis	www.mbio.ncsu.edu.	-
4	Clustal W	Sequence alignment	www.ebi.ac.uk	-

Identification of new haplotypes

Allele mining can be potentially employed in the identification of nucleotide variation at a genomic region (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. Many case studies are available for analyzing the haplotype diversity and identification of new haplotypes at candidate genes. Knowledge on the most common haplotype changes and their frequency in the populations will form the basis for association mapping studies.

Development of allele-specific markers for MAS

Identification of sequence variation will pave the way to develop allele-specific marker assay for precise introgression of the identified 'superior and/or novel' alleles to suitable genetic background. In recent years, several case studies are being reported which demonstrate the existence of sequence variation at key genes while some studies have demonstrated the utility of these variations through development of allele-specific molecular markers for MAS. For example, in rice, comparison of nucleotide sequences of Waxy gene (codes for a granule-bound starch synthase) in 18 different accessions revealed the presence of five different alleles, which are characterized with a unique replacement, frame shift or site mutations. All the alleles are clearly associated with the observed phenotypic alterations (Mikami *et al*., 2008).

Allelic synteny and evolutionary relationship

Using the sequence information obtained from allele mining studies, syntenic relationships can be assessing among the identified loci/genes across the species/genera. In tomato, Rose *et al*. (2005) amplified and sequenced the alleles of Pto gene (conferring resistance to bacterial pathogen) in 49 genotypes and tested a subset of these alleles for its function, identified several nucleotide changes responsible for pathogen recognition and hypersensitive resistance response and these changes to the natural variation in resistance to *Pseudomonas syringae pv. tomato* (*Pst*) strains.

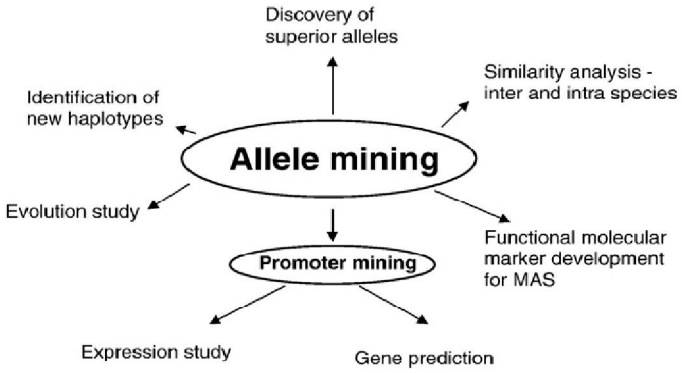

Fig.4. Applications of allele mining

Challenges and solutions in allele mining

1. Selection of genotypes: The most important challenge in unlocking the existing variation is the selection of germplasm to be 'mined'. Given a reliable protocol for characterizing a gene, screening the entire collection would certainly be helpful to find rare alleles, but this is an enormous and inefficient way of screening. Hence, we need to find a way forward from genotyping the whole composite collections to efficiently screening the accessions to discover new alleles.

2. Development of core/mini core collections: Handling the entire germplasm is a difficult task, be it for conventional plant breeding or for allele mining and hence requires sampling strategies to narrow down to a manageable size while maintaining the variability. Development of core and mini collections out of the entire collection has been proposed to simplify the conservation of germplasm resources and effective utilization of the existing variation in gene banks.

3. Accurate phenotyping methods: Phenotyping is a key for success in short listing the genotypes for allele mining as phenotypes are considered as the best clues for genotypes. Increase in precision in phenotyping techniques maximizes the chances of 'mining' the prospective genotypes. Also,with the development of trait-specific subsets obtained from phenotyping experiments the number of genotypes for 'mining' is significantly reduced. Hence, there is a need to refine phenotyping protocols to increase the efficiency of allele mining.

4. Handling genomic resources: To keep pace with rapid accumulation of nucleotide and gene expression data, computational tools need to be developed for analyzing the functional nucleotide diversity and to predict specific nucleotide changes responsible for altered function. Exploiting the developments in allele mining, association genetics and comparative genomics by combining expertise from several disciplines, including molecular genetics, statistics and bioinformatics is the suggested way forward.

5. Higher sequencing costs: One of the important challenges is to minimize the time and efforts require while reducing the cost per data point. These challenges may minimum be overcome by resorting to cheaper and faster sequencing platforms for high output detection of allelic variations.

6. Flexible computational tools: Computational tools are useful for selection of desirable alleles for rapid expansion of Gene Bank.

Conclusion

The use of genetic diversity is limited due to the resources which are at hand for characterization of all these 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 underlying agronomically important traits 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, through 'mining' the gene of interest from germplasm. It can also provide insight into molecular basis of novel trait variations and identify the nucleotide sequence changes associated with superior alleles. In addition, the rate of evolution of alleles; allelic similarity and dissimilarity at a candidate gene and allelic relationship with other members of the family. Allele mining may also pave way for introgression of novel alleles through Marker Assisted Selection.

REFERENCES

Chu, Z., Yuan, M., Yao, J., Ge, X., Yuan, B. and Xu, C. 2006. Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev*. 20:1250–1255.

Comai, L., Young, K., Till, B. J., Reynolds, S. H., Greene, E. A. and Codomo, C. A. 2004. Efficient discovery of DNA polymorphisms in natural populations by EcoTilling. *Plant J.* 37:778-86.

Davuluri, R. V., Sun, H., Palaniswamy, S. K., Matthews, N., Molina, C. and Kurtz, M. 2003. AGRIS: Arabidopsis gene regulatory information server, an information resource of *Arabidopsis* cis-regulatory elements and transcription factors. *BMC Bioinform*. 4:25-6.

Glaszmann, J. C. 1987. Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* 74: 21-30.

Kalendar, R. 2009. Fast PCR software for PCR primer and probe design and repeat search. www.biocenter.helsinki.fi/bi/programs/fastpcr.htm.

Khush, G. S., Bacalangco, E. and Ogawa, T. 1991. A new gene for resistance to bacterial blight from *O. longistaminata*. *Rice Genet. Newslett*. 7:121-122.

Latha, R., Rubia, L., Bennett, J. and Swaminathan, M. S. 2004. Allele mining for stress tolerance genes in *Oryza* species and related germplasm. *Mol. Biotech.*27: 101-108.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S. and Bemben, L. A. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376-80.

Mikami, I., Uwatoko, N., Ikeda, Y., Yamaguchi, J., Hirano, H. Y. and Suzuki, Y. 2008. Allelic diversification at the wx locus in landraces of Asian rice. *Theor. Appl. Genet.* 116:979-89.

Nagy, A., Perrimon, N., Sandmeyer, S. and Plasterk, R. 2003. Tailoring the genome: The power of genetic approaches. *Nat. Genet.,* 33: 276-284.

Raghavan, C., Naredo, M. E. B., Wang, H., Atienza, G., Liu, B. and Qiu, F. 2007. Rapid method for detecting SNPs on agarose gels and its application in candidate gene mapping. *Mol. Breed.* 19:87-101.

Rose, L. E., Langley, C. H., Bernal, A. J. and Michelmore, R. W. 2005. Natural variation in the Pto pathogen resistance gene within species of wild tomato (*Lycopersicon*). *Genet*. 171:345-57.

Singh, B. D. 2005. A text book of plant breeding. Kalyani publishers. 171–172pp.

Sitch, L. A., Amante, A. D., Dalmacio, R. D. and Leung, H. 1989. Oryza minuta, a source of blast and bacterial blight resistance for rice improvement. In: Mujeeb-Kazi A, Sitch LA, editors. *Review of advances in plant biotechnology*. Mexico and IRRI, Philippines: CIMMYT. 315–322pp.

Tan, L., Li, X., Liu, F., Sun, X., Li, C. and Zhu, Z. 2008. Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet*. 40:1360-1364.

Till, B.J., Reynolds, S.H., Grene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E,, Burtner, C., Odden, A.R., Young, K., Taylor, N.E., Henikoff, J.G., Comai, L. and Henikoff, S. 2003. Large scale discovery of induced point mutations with high-throughput TILLING. *Genome Res.,* 13: 524-530.

****