

Original Research Article

<https://doi.org/10.20546/ijcmas.2019.806.020>

## Utilization and Characterization of Genome-wide SNP Markers for Assessment of Ecotypic Differentiation in *Arabidopsis thaliana*

Astha Gupta<sup>1,2,3\*</sup>, Archana Bhardwaj<sup>1,2</sup>, Samir V. Sawant<sup>1,2</sup>  
and Hemant Kumar Yadav<sup>1,2</sup>

<sup>1</sup>CSIR-National Botanical Research Institute, Rana Pratap Marg,  
Lucknow, UP, India -226001

<sup>2</sup>Academy of Scientific & Innovative Research (AcSIR), New Delhi, India – 110 025

<sup>3</sup>Department of Botany, University of Delhi, New Delhi, India - 110 007

\*Corresponding author

### ABSTRACT

Development of SNPs (Single Nucleotide Polymorphisms) marker is an important step to initiate the molecular breeding and genetic based studies. Identification and validation of polymorphic SNP will be valuable resource for gene tagging through linkage mapping/QTL mapping. In present study, two ecological ecotypes of *Arabidopsis thaliana* i.e. Col-0 and Don-0 exhibited variation at phenotypic level (leaf, flower, siliques and root related traits) and genotypic level (SNPs). Out of 500 SNPs, total 365 polymorphic SNPs were validated on Sequenome MassARRAY. These polymorphic SNPs would be very useful for genotyping of Col-0 and Don-0 mapping population to explore the quantitative trait loci for desired trait in future studies. Detailed analysis of selected SNPs gives the idea of their distribution in genome includes location with their nature. Location (coding and non-coding region) and nature (synonymous and non-synonymous) of SNPs may also create the phenotype diversity by regulation of genes in cis and trans regulatory mechanism and/or modulation of metabolic process and pathway. Identified non-synonymous deleterious SNPs (G/C) may associate with biomass trait because it encodes a plastid-localized Nudix hydrolase that has FAD pyrophosphohydrolase activity (control growth and development). In addition, this SNP can alter the protein function by controlling riboflavin metabolism, purine metabolism and their related metabolic pathways which ultimately may responsible for phenotypic differences. Result suggested that SNP may lead phenotypic variability and associate with particular traits. Later, SNPs genotyping and QTL mapping would be helpful for candidate gene tagging and marker-assisted breeding in *Arabidopsis*.

#### Keywords

Genome-wide SNP  
Markers,  
*Arabidopsis  
thaliana*

#### Article Info

Accepted:  
04 May 2019  
Available Online:  
10 June 2019

### Introduction

Single nucleotide polymorphisms (SNPs) are sequencing-based marker and very informative to explore the genetic variation

that influence the phenotype (Bokharaeian *et al.*, 2017). SNP may originated because of single nucleotide alternation (deletion, insertion or transition and transversion substitution) during evolution for adaptation

under unfavourable conditions. SNPs are distributed throughout the genome i.e. coding and non-coding region which may alter metabolic pathway processes and lead to phenotypic change (Zhou *et al.*, 2012; Zhao *et al.*, 2016; Massonnet *et al.*, 2010). SNPs presence in non-coding region may alter the binding sites of transcription factor, regulator, enhancer, silencer, splice sites and other functional site for transcriptional regulation (Reumers *et al.*, 2007). In coding region, SNPs are further categorized into synonymous (no change in protein nature) and non-synonymous SNPs (alteration in protein structure and function) and affect the function of protein which can be visualized by SNPviz tool (Seitz *et al.*, 2018). In 1001 Genomes Project, several ecotypes of *Arabidopsis* have been sequenced including Col-0 and Don-0 and approximately 711,668 unique SNPs were identified between these two ecotypes of *Arabidopsis* (Cao *et al.*, 2011) which can be utilized for diversity analysis, allele mining, gene discovery, functional genomics or marker assisted selections/breeding. Although it is observed that SNPs contributed in phenotypic variation and were associated with trichome density, days to flowering, level of leaf serration in *Arabidopsis* (Lee and Lee 2018). Therefore, there is need to identify the association between identified polymorphic SNPs with particular traits due to presence and availability of unique SNPs in genome of Don-0. As one report suggested that Don-0 ecotype contain unique SNPs and identified novel active allele associated with trait (Mendez-Vigo *et al.*, 2016). Establishment of association (SNPs marker and trait) would be useful for detection of novel allelic contribution involved in phenotypic variations, metabolic pathways and processes. In present study true SNPs will be validated between Col-0 and Don-0 on Sequenome MassARRAY followed by detection of functional impact of SNPs. In addition to that,

phenotypic variation of novel and less studied Don-0 ecotype of *Arabidopsis* would be explore with widely studied Col-0 ecotype which would be further useful for molecular biology and genetics studies.

## **Materials and Methods**

Two ecotypes of *Arabidopsis* i.e. Col-0 and Don-0 were chosen for present study which located in Columbia and Donana with different longitude of -92.3 and -6.36 respectively (Table 1). Previous research suggested that selected ecotypes were different at ecological and molecular level (Wang *et al.*, 2012; Cao *et al.*, 2011) due to their presence in different geographical conditions.

## **Growth conditions and procedure**

Col-0 and Don-0 seeds were procured from Arabidopsis Biological Resource Centre (ABRC), Ohio State University (<https://abrc.osu.edu/>) and grown under the glasshouse conditions at CSIR-NBRI, Lucknow. Seeds were sown in pot commercial soil mix containing soilrite (Keltech Energies Ltd., Bengaluru, India) and vermiculite (3:1) at 22<sup>0</sup>C with particular growth conditions (16 hr light/8hr dark photoperiod, 200  $\mu\text{mol m}^{-2} \text{s}^{-1}$  light intensity and 80% relative humidity). Pots were kept in tray (with 1inch of filled Osgrel Somerwhile solution media) at 4<sup>0</sup>C for 3 days stratification and covered with plastic wrap followed by transferred to glasshouse for proper growth.

## **Evaluation of phenotypic variations**

Seeds were germinated and developed in to plant under glasshouse conditions. It was observed that plants of Col-0 and Don-0 showed phenotypic diversity. Therefore, phenotypic data was recorded between Col-0 and Don-0 (average of six plants) for some

phenotypic traits includes bolting and flowering days, differences in leaf morphology and structure, trichome density, flower diameter, plant height and seed length and root related traits etc.

### **Selection of polymorphic SNP from 1001genomes**

Genome sequence data of Col-0 and Don-0 ecotypes was available 1001 Genomes-A Catalog of *Arabidopsis thaliana* Genetic Variation (<http://1001genomes.org/>). Therefore, the SNP sequence data (working variants with reference) was downloaded and a set of 100 SNPs were selected from each chromosome (total 500 SNPs: almost uniformly distribute on the five chromosomes of *Arabidopsis*). In this way, a set of 500 sequences were extracted for designing SNP assay. We retrieved the 200 bases upstream and downstream from each of selected SNP sites, which were used to design SNP specific primers by MassARRAY Assay Design 3.0 software.

### **Validation of true polymorphic SNP**

DNA was isolated from the leaf of Col-0 and Don-0 through DNAzol method (manufacturer's protocol; Invitrogen) and checked on 0.8% agarose gel using  $\lambda$  DNA (Invitrogen, Carlsbad, CA, USA). Extracted genomic DNA was normalized to 10 ng/ $\mu$ l for further PCR amplification and SNP genotyping.

The SNP genotyping was performed on Sequenom<sup>TM</sup> MassARRAY platform (available at CSIR-NBRI, Lucknow) using iPLEX<sup>TM</sup> protocol as described by the manufacturer (Oeth *et al.*, 2005). True polymorphic SNPs were screened between Col-0 and Don-0 after peak analysis on Sequenom<sup>TM</sup> MassARRAY platform. SNPs exhibited missing data were eliminated for further analysis.

### **Functional impact of SNPs**

SnEff software (Cingolani *et al.*, 2012) was used to annotate the effect of SNPs (synonymous and non-synonymous). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) have been performed for SNPs encoding genes using Kobas web server (<http://kobas.cbi.pku.edu.cn/home.do>). Non-synonymous SNPs were used for analysis of deleterious SNP on the basis of functional effect of amino acid substitution on corresponding proteins through PANTHER23 (tolerance index score of  $\leq 0.05$ ; Thomas *et al.*, 2003).

### **Results and Discussion**

#### **Evaluation of phenotypic diversity**

Germination rate of Col-0 (100%) was higher than Don-0 (66-75%) under glasshouse conditions.

It was observed that Col-0 and Don-0 exhibited variations for several phenotypic traits (Figure 1). Col-0 showed early bolting (31 days) and flowering (41.3 days) as compared to Don-0 (76.3 bolting days and 85.3 flowering days). At maturity, rosette diameter was high in Don-0 (7.7 cm) as recorded in Col-0 (10.9 cm). Maximum number of rosette leaf was counted in Don-0 (87 leaves) as compared to Col-0 (63 leaves). Rosette leaf length of Col-0 (2.54 cm) was less than Don-0 (3.18 cm) but width was high (Col-0: 1.88 cm and Don-0: 1.73 cm). Trichome density was analysed in mature leaf (3 leaf: average of 9 square box of 0.5 cm<sup>2</sup> leaf area) which was high in Col-0 (26 trichomes) as observed in Don-0 (17 trichomes). In addition, Col-0 exhibited serration in rosette leaf margin in contrast to Don-0 (smooth leaf margin). Number of cauline leaf (stem leaf) was high in Don-0 (93 leaves) as counted in Col-0 (51 leaves; single

leaf appeared on each node) at maturity. Maximum plant height of Col-0 and Don-0 was measured 33.90 cm, 39.70 cm as measured at 69 days, 118 days respectively. Flower diameter of Col-0 was large i.e. 0.4 cm as recorded in Don-0 (0.3 cm). At maturity, average silique length (total 36 siliques: 6 siliques / plant of each ecotype) was high for Don-0 (1.4 cm) as compared to Col-0 (1.1 cm). Initially root length and number of secondary roots of Don-0 (4.9 cm and 7.1) was lesser than Col-0 on MS agar media (9.1 cm and 15.4) up to 20 days but high at maturity. Under soil condition, root length and root biomass of Don-0 (26 cm and 47.7 mg) was high as compared to Col-0 (21 cm and 13.6 mg) at 121 days. Visualization of root hairs under confocal microscope interpreted that Don-0 contained high number of root hairs.

### **Validation of polymorphic SNPs**

Out of 500 SNP, 365 polymorphic SNPs (73%) were successfully screened on Sequenom<sup>TM</sup> MassARRAY platform and used for further analysis (list of polymorphic SNP: supplementary Table 1). Rest of 27% (135 SNPs) were not validated between Col-0 and Don-0 as detected previously (1001 genome project) due to missing data or wrong allele call during analysis. During SNP analysis, particular SNP primer showed homozygous call for both ecotypes for example: peak of 'CC' allele in Col-0 and 'AA' allele in Don-0 (Figure 2).

### **Classification of SNPs based on their impact on gene functionality**

Total validated 365 SNPs were annotated and classified into three categories depending upon SNP impact on gene functionality using SnpEff tool (Cingolani *et al.*, 2012). All the selected SNPs were classified into three classes named as low (8.8 %) moderate

(12.6%) and modifier (78.6%) SNPs. Approximately 20% SNPs (73 SNPs) were found in coding region includes synonymous (27 SNPs) and non-synonymous (46 SNPs) (Table 2).

SNPs code for same nature of amino acid (hydrophobic/hydrophilic) through alteration of single nucleotide change which showed less effect on gene functionality comes under low impact synonymous SNPs. We found total 27 synonymous SNPs for example: leucine-rich repeat receptor kinase (AT1G31420), succinate dehydrogenase assembly factor 2 (AT5G51040), TATA-binding related factor (AT2G28230), histone acetyl transferase (AT5G50320). Interestingly, one of SNP showed start codon gain (SNP A/G) effect in 5' UTR of unknown gene AT3G26440 which may have some specific function and might be involved particular molecular pathways or processes. In present results, three SNPs (G/A, T/A and A/T SNPs) were identified as splice variants that effected following genes: polynucleotidyl transferase (AT5G61090), LIM proteins (AT1G10200) and ubiquitin-specific protease 8 (UBP8; AT5G22030). These splice variants might play role in diversity as it could lead to production of multiple proteins of different functions.

Non-synonymous SNPs were observed under the moderate type of impact on gene functionality which altered the protein structure and function (due to change in amino acid; hydrophobic to hydrophilic and vice versa) by nucleotide substitutions. Although, aspartyl protease family protein (AT5G48430) contained T/G non-synonymous SNP and change Lysine to Asparagine amino acid at 202 position (Lys202Asn). It was investigated that missense non-synonymous SNPs were found in phloem protein 2-B1 (AT2G02230, F-box domain, C/A SNP) and putative transcription

factor -MYB59 (G/C SNP; AT5G59780) which altered amino acid Val116Leu and Phe191Leu correspondingly.

Maximum number of SNPs (222 SNPs: 61%) were lies in upstream region followed by downstream region (35 SNPs) found in modifier class. In modifier class SNPs affects the gene functionality due to presence in binding site of transcription factors (upstream region: promoter) and miRNA (5` and 3` UTR). A/T-SNP was identified in 5` and 3` UTR that encode UDP-glucosyl transferase 71C1 (AT2G29750) and Chromatin Assembly Factor-1 (AT5G64630) which is involved in metabolic process of the shoot and root apical meristem (Kaya *et al.*, 2001). The Homeobox-leucine zipper family protein (HD-ZIP IV; AT1G05230) was found in modifier SNP (G/T) related to trichome development (Marks *et al.*, 2009).

In addition to that upstream region SNP (T/C) encodes CLAVATA1-related receptor kinase-like protein (AT4G20270) and C/T SNP was found in gene SNF1-related protein kinases (SnRK2; AT3G50500) which control leaf morphology (DeYoung *et al.*, 2006), root growth and seed germination (Fujii *et al.*, 2007) correspondingly. Downstream SNP-C/T and upstream SNP-G/T were consist of ACTIN-RELATED PROTEIN6 (ARP6: chromatin-remodeling complex, AT3G33520) and zinc finger domain (AT2G33835) respectively that regulate flowering in *Arabidopsis* (Choi *et al.*, 2005, 2011).

### **Gene ontology and KEGG analysis**

Annotations of selected SNPs would provide a valuable resource for investigating specific processes, functions, and pathways underlying variations between Col-0 and Don-0. Alteration of pathways and molecular processes might be combination of alleles/SNPs and their position on genome

which lead phenotype or traits modifications. Gene ontology and pathway analysis of SNP containing genes were conducted using KOBAS server. All genes were assigned to at least one term in GO molecular function, cellular component and biological process categories with best hits (Figure 3). All selected genic SNPs were further classified into 42 functional subcategories, providing an overview of ontology content. However, cellular component was most highly represented groups (GO term: 246) followed by biological process (GO term: 143) and molecular function (GO term: 91). In cellular component category, cell and cell parts were the most highly represented functional subcategories which may involved for variations of biomass between both plants. Cellular process, metabolic process and binding, catalytic activity were dominating functional subcategories of biological process and molecular function respectively which might be involved for phenotypic variation of Col-0 and Don-0. Therefore, GO terms served as indicators of different biological and cellular processes takes place in cells of plant. As a result, It was found that 8 genes showing significant enriched GO term i.e response to stress (P value <0.05) which are following AT2G01440, AT1G35515, AT4G36150, AT1G33590, AT5G59780, AT5G58670, AT3G05640, AT2G35000 (Figure 4). Pathway-based analysis was performed for same set of SNPs sequences using the KEGG pathway database to identify metabolic pathways in which eight genes were participating under nine pathways for example: glutathione metabolism, riboflavin metabolism, N-glycan biosynthesis, homologous recombination, ribosome biogenesis in eukaryotes, purine metabolism, RNA transport, plant hormone signal transduction and metabolic pathways. Three genes (AT2G42070, AT4G30910 and AT1G16900) were involved in metabolic pathways followed by two genes in

glutathione metabolism (AT4G30910 and AT2G29460). Therefore, further study was focus on these genes. PANTHER (Protein analysis through evolutionary relationships) was used to categorized these SNPs into tolerable and deleterious based on tolerance index score of  $\leq 0.05$  and found that genes containing SNP: AT4G30910 (SNP G/C), AT5G41190 (T/C), AT2G29460 (A/G) and AT1G16900 (G/T) were tolerant except AT2G42070 (G/C) which was deleterious non-synonymous SNP. Interestingly it was observed that AT2G42070 gene was involved in multiple pathways includes riboflavin metabolism (Figure 5), purine metabolism

and metabolic pathways (supplementary file 1). Due to nucleotide substitution of non-synonymous SNP, amino acid alteration takes place from polar to polar AA (Tyr62His and Ser192Tyr), hydrophobic to hydrophobic AA (Ile90Val) and polar to charged AA (Gln494Glu) indicated four tolerable SNPs. Deleterious non-synonymous SNP AT2G42070 (G/C) showed Thr28Ser AA change with P-Value: 0.02 (score: 0.00) that can affect the protein function which encodes a plastid-localized Nudix hydrolase that has FAD pyrophosphohydrolase activity (Maruta *et al.*, 2012).

**Table.1** Basic information of Col-0 and Don-0 ecotypes

| Descriptions    | Information of selected ecotypes                         |  |
|-----------------|--|--|
| Name            | Col-0  | Don-0  |
| Ecotype ID      | 6909   | 9944   |
| CS Number       | CS76778  | CS76411  |
| Country         | United States of America (USA)                           | Spain  |
| Location        | Columbia   | Donana   |
| Latitude        | 38.3   | 36.83  |
| Longitude       | -92.3  | -6.36  |
| Sequencing year | 2000   | 2010   |
| Sequenced by    | Gregor Mendel Institute of Molecular Plant Biology (GMI) | Max Planck Institute for Developmental Biology (MPI) |

**Table.2** SNPs distribution and their mode of action

| SNP effect | Location and nature of SNP                     | count |
|------------|--|-------|
| Low        | splice_region_variant                          | 1     |
|            | 5_prime_UTR_premature_start_codon_gain_variant | 1     |
|            | splice_region_variant&intron_variant           | 3     |
|            | synonymous_variant                             | 27    |
| MODERATE   | missense_variant (non-synonymous)              | 46    |
| MODIFIER   | intergenic_region                              | 2     |
|            | non_coding_transcript_exon_variant             | 3     |
|            | 5_prime_UTR_variant                            | 5     |
|            | 3_prime_UTR_variant                            | 20    |
|            | downstream_gene_variant                        | 35    |
|            | upstream_gene_variant                          | 222   |
| Total      |  | 365   |

Fig.1



Fig.2

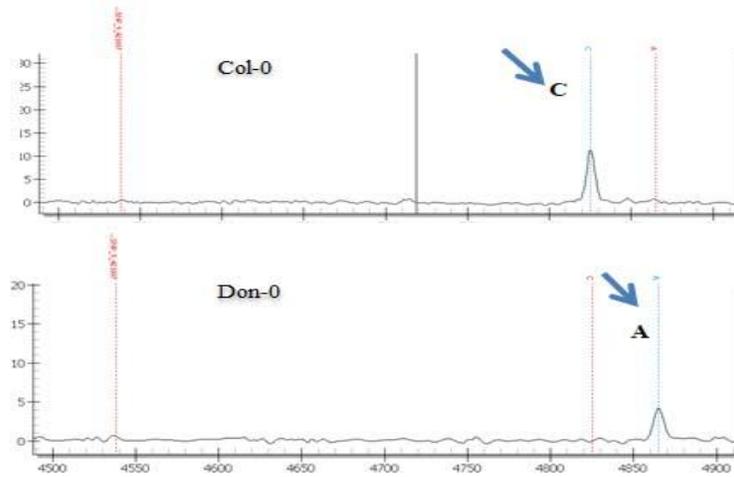


Fig.3

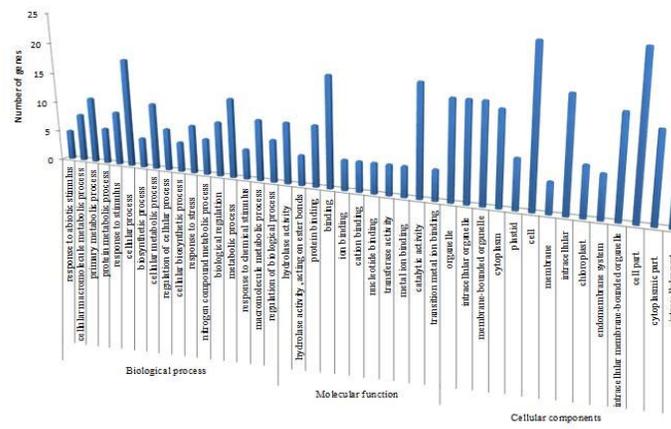


Fig.4

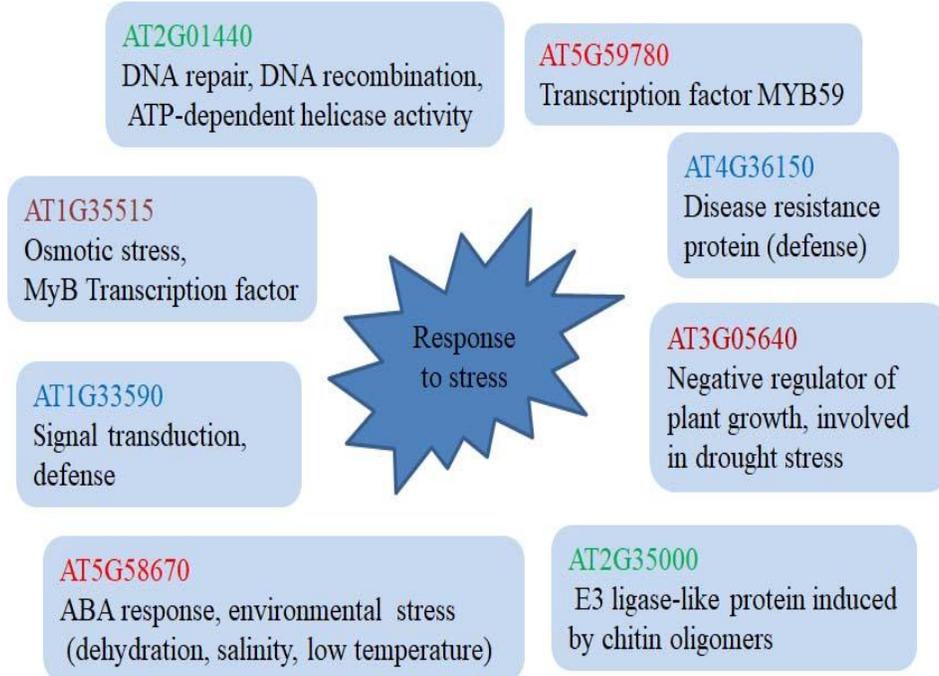
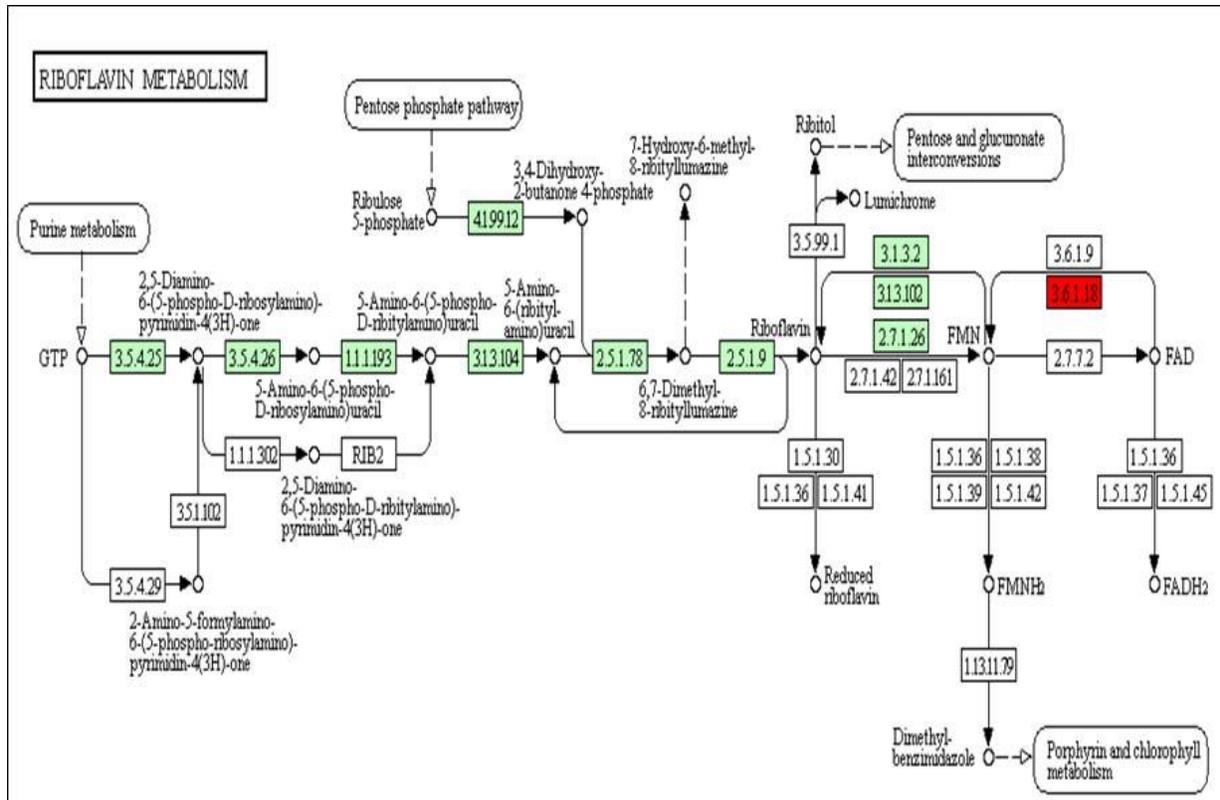


Fig.5





In present study, phenotypic diversity has been explored between Col-0 and Don-0 under glasshouse conditions. Although, bioinformatically detected in-silico SNPs were also validated through wet-lab experiments on Sequenom MassARRAY. Successfully identified and polymorphic SNPs (365 SNP) might be associated with particular phenotypic traits that can regulate metabolic pathway and processes as analysis predicted. However, phenotypic traits were analysed between Col-0 and Don-0 which showed visual variations for rosette size, leaf structure, morphology, trichome and root traits, flower size, flowering days, bolting days and silique related traits. In addition, genetic variations were also detected between Col-0 and Don-0 which has been explored through SNP markers screening. We can hypothesized that these SNP may govern particular traits directly (cis-regulation) or indirectly (trans-regulation) depending upon their location within genome.

After annotation through SnpEff tool (Cingolani *et al.*, 2012), maximum number of SNPs were located in non-coding region (hetero-chromatin, as explained in Table 2) that may associated with epigenetic contribution of DNA methylation, histone modifications and gene expression which would lead epigenetic regulation of phenotypic variations (Fujimoto *et al.*, 2012; Groszmann *et al.*, 2011; Shen *et al.*, 2012; Zhu *et al.*, 2016; Zhu *et al.*, 2017). Non-coding region may also involve indirectly for phenotypic variation by regulation of protein binding factor (transcription factor and regulator) on promoter binding (upstream region). In previous studies, SNP polymorphism is also reported in promoter, UTRs that regulates gene expression which create natural morphological variations (Guyon-Debast *et al.*, 2010). Presence of SNPs in 5' UTR or 3' UTR, intronic region and splice site may affects the mRNA

stability and translation that leads the different protein and consequently altered phenotypic traits (Gardner *et al.*, 2016; Zhao *et al.*, 2016; Rodgers-Melnick *et al.*, 2016). For instance, candidate drought-QTL of *Arabidopsis* was associated with two SNPs found in 5' UTR and promoter of same gene i.e. AT5G0425 (Bac-Molenaar *et al.*, 2016).

Phenotypic variation between Col-0 and Don-0 for shoot, root biomass traits might be existence of two SNPs in UTR region that is UDP-glucosyl transferase 71C1 (AT2G29750; SNP A/T) and Chromatin Assembly Factor-1 (AT5G64630; SNP A/T) related to shoot, root traits (Kaya *et al.*, 2001). Less number of trichome (mature leaf) and poor seed germination of Don-0 (as compared to Col-0) may associate with SNP G/T of Homeobox-leucine zipper family protein (AT1G05230: HD-ZIP IV) and SNP C/T of SNF1-related protein kinases (SnRK2: AT3G50500) genes correspondingly or their interactions with other regulatory elements. However, HD-ZIP IV and SnRK2 genes regulate trichome development and seed germination, dormancy respectively (Marks *et al.*, 2009; Nakashima *et al.*, 2009). Although, SNP (T/C) encodes CLAVATA1-related receptor kinase-like protein (AT4G20270) which play role in development of leaf shape, size and symmetry (DeYoung *et al.*, 2006) and might be correlated for variation in leaf morphology between Col-0 and Don-0. Downstream gene variant (SNP C/T) of actin-related protein 6 (ARP6: chromatin-remodeling complex, AT3G33520) may alters the expression of FLC, MAF4, MAF6 genes by histone acetylation and methylation of the FLC chromatin in *Arabidopsis* (Choi *et al.*, 2005). As previous research suggested that C/T transition led to distorted and unstable hairpin structure of miRNA (Singh *et al.*, 2017) which play important role in the post transcription regulation of gene expression. The Zinc finger domain (AT2G33835; SNP

G/T) might be responsible for delay flowering of Don-0 because it acts with FRI to repress flowering (Choi *et al.*, 2011).

Occurrence of SNPs in coding regions may be responsible for particular phenotype (Keurentjes *et al.*, 2007; Zhou *et al.*, 2012) of Col-0 or Don-0 by altering gene product, proteins and metabolites. Don-0 exhibited slow growth (less root length and shoot biomass in early stage of life) which might be regulation of coding synonymous SNP (G/T) which encode histone acetyl transferase (AT5G50320) because this gene reduced cell division rate that may lead reduced plant growth (Fina *et al.*, 2015). However, splice region SNPs were also identified (polynucleotidyl transferase, LIM proteins, ubiquitin-specific protease 8) which may introduced premature termination codon by creation of new splicing branchpoint (Guyon-Debast *et al.*, 2010) that would be responsible for diversity or may create new phenotype.

Coding synonymous SNP (T/G) of TATA-binding related factor (AT2G28230) may altered mRNA stability (Duan *et al.*, 2003; Capon *et al.*, 2004) and gene expression through RNA polymerase II transcription mediator activity. Identified Leucine-rich repeat receptor kinase, succinate dehydrogenase assembly factor 2 involved in cell wall biosynthesis and root elongation (Torii 2004). These coding synonymous SNP may have some effect on appearance of phenotypic diversity because of their occurrence in expressed region. Non-synonymous SNP highly affects the phenotype (Ramensky *et al.*, 2002) due to change in protein structure and confirmation. However, non-synonymous SNP (G/C) was identified in putative transcription factor-MYB59 (AT5G59780) that expressed in leaves and seedlings which is also known for alternatively generated spliced transcript (Horstmann *et al.*, 2000; Guo *et al.*, 2017)

that play important roles in many developmental processes, defence responses of plants (Li *et al.*, 2006).

After analysis it was predicted that, non-synonymous SNP were involved in multiple molecular processes and pathways which control phenotypes (Massonnet *et al.*, 2010) through regulation of metabolic pathways. In present analysis these SNPs were participated in selected pathways like metabolic pathways, glutathione metabolism, riboflavin metabolism, N-glycan biosynthesis, homologous recombination, ribosome biogenesis, purine metabolism, RNA transport, plant hormone signal transduction. Therefore, PANTHER tool was used to detect the nature of SNPs which regulates specific pathways. Later on deleterious SNP was identified (AT2G42070) with amino acid substitutions of G to C which may altered the biological functions of a target protein (Singh *et al.*, 2017; Bhardwaj *et al.*, 2016). In present investigation, substitution from Isoleucine to Valine at position 90 (Ile90Val) may associate with particular phenotypic trait as predicted in DNA repair genes (Ile658Val: DNA double-strand break repair protein associated with lung cancer risk; Sakiyama *et al.*, 2005). Similar study was performed using genome-wide analysis of branched-chain amino acid levels (isoleucine and valine) was performed using SNP marker and found their association with seed traits in *Arabidopsis* (Angelovici *et al.*, 2013). Likewise, non-synonymous deleterious SNP (may affect protein function) carried amino acid substitution at position 28 from threonine to serine (Thr28Ser) that encodes a plastid-localized Nudix hydrolase, distributed in plastids and has FAD pyrophosphohydrolase activity (hydrolyze FAD to produce FMN and AMP in plastids; Maruta *et al.*, 2012). Further analysis suggested that it regulates the ratio of FMN and FAD in whole plant cells and play diverse roles in wide range of physiological

processes (Ogawa *et al.*, 2008), pathogen resistance (Deng *et al.*, 2011). It is previously investigated that AtNUDX23 showed FAD pyrophosphohydrolase activity in *Arabidopsis* leaves (Maruta *et al.*, 2012) and pea plants (plastids and mitochondria, Sandoval *et al.*, 2008). However, FMN and FAD (part of flavoprotein) are essential cofactors for a variety of enzymes that involved in several metabolic processes and pathways: photosynthesis and mitochondrial electron transport (Sandoval *et al.*, 2008). Furthermore, Photosynthesis and photorespiration is very important process for plant growth and development which is directly associated for high biomass as studied in *Arabidopsis* (Liu *et al.*, 2016; Simkin *et al.*, 2017). So further exploration of these polymorphic SNP may involve in QTL mapping (linkage or association between polymorphic marker with traits) and other molecular breeding programmes for trait improvement. Therefore it would be direct contribution to detect the desired candidate gene for particular traits.

In conclusion, present study explained the significance of SNPs and their annotation information using two contrasting ecotypes of *Arabidopsis thaliana*. Nucleotide substitution (transition and transversion) would also play very important role to regulate the molecular process, pathways and phenotypic diversity. However, non-coding SNPs (cytosine modification) may be associated with chromatin modeling (i.e. hetero-chromatin and euchromatin by methylation process) for their active participation in phenotypic variability. Identified upstream and downstream SNPs might be associated with regulatory phenomenon of genes which will control the expression of concern genes through post transcription regulation. SNP containing gene involved in particular metabolic pathway and process may be responsible for phenotypic differences.

Coding synonymous and non-synonymous SNPs governed the transcript and protein diversity which ultimately regulate metabolomics and lead phenotypic diversity. Further, Fine mapping can predict the closely linked loci with phenotypic traits by establishment the association of SNPs through QTL mapping to detect the contribution of active allele/genes.

### **Author contributions**

The experiment was designed by S.V.S. and H.K.Y. Experiments were conducted by A.G. and data analysis performed by A.G. and A.B. All the authors have read and approved the final manuscript.

### **Acknowledgements**

This research was financially supported by Council of Scientific & Industrial Research (CSIR), India (BSC 0204). All the experiments and analysis was performed at CSIR-National Botanical Research Institute (NBRI), Lucknow.

### **References**

- Angelovici R, Lipka AE, Deason N, *et al.*, Genome-wide analysis of branched-chain amino acid levels in *Arabidopsis* seeds. *Plant Cell* 2013; 25(12):4827-4843. <https://doi.org/10.1105/tpc.113.119370>.
- Bac-Molenaar JA, Granier C, Keurentjes JJ, *et al.*, Genome-wide association mapping of time-dependent growth responses to moderate drought stress in *Arabidopsis*. *Plant, cell environ* 2016; 39(1): 88-102. <https://doi.org/10.1111/pce.12595>.
- Bhardwaj A, Dhar YV, Asif MH, *et al.*, In Silico identification of SNP diversity in cultivated and wild tomato species: insight from molecular

- simulations. *Sci Rep* 2016; 6:38715. <https://doi.org/10.1038/srep38715>.
- Bokharaeian B, Diaz A, Taghizadeh N, *et al.*, SNPPhenA: a corpus for extracting ranked associations of single-nucleotide polymorphisms and phenotypes from literature. *J Biomed Semantics* 2017; 8(1):14. <https://doi.org/10.1186/s13326-017-0116-2>.
- Cao J, Schneeberger K, Ossowski S, *et al.*, Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 2011; 43(10):956-63. <https://doi.org/10.1038/ng.911>.
- Capon F, Allen MH, Ameen M, *et al.*, A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups. *Hum Mol Genet* 2004;13(20):2361-2368. <https://doi.org/10.1093/hmg/ddh273>.
- Choi K, Kim J, Hwang HJ, *et al.*, The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in *Arabidopsis*, by recruiting chromatin modification factors. *Plant Cell* 2011; 23(1):289-303. <https://doi.org/10.1105/tpc.110.075911>.
- Choi K, Kim S, Kim SY, *et al.*, SUPPRESSOR OF FRIGIDA3 encodes a nuclear ACTIN-RELATED PROTEIN6 required for floral repression in *Arabidopsis*. *Plant Cell* 2005; 17(10):2647-2660. <https://doi.org/10.1105/tpc.105.03548> 5.
- Cingolani P, Platts A, Wang LL, *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 2012; 6(2): 80-92. <https://doi.org/10.4161/fly.19695>.
- Deng B, Deng S, Sun F, *et al.*, Down-regulation of free riboflavin content induces hydrogen peroxide and a pathogen defense in *Arabidopsis*. *Plant Mol Biol* 2011; 77: 185–201. <https://doi.org/10.1007/s11103-011-9802-0>.
- DeYoung BJ, Bickle KL, Schrage KJ, *et al.*, The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in *Arabidopsis*. *Plant J* 2006; 45(1): 1-16. <https://doi.org/10.1111/j.1365-313X.2005.02592.x>.
- Duan J, Wainwright MS, Comeron JM, *et al.*, Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet* 2003; 12: 205–216. <https://www.ncbi.nlm.nih.gov/pubmed/12554675>.
- Fina JP, Casati P. HAG3, a histone acetyltransferase, affects UV-B responses by negatively regulating the expression of DNA repair enzymes and sunscreen content in *Arabidopsis thaliana*. *Plant Cell Physiol* 2015; 56(7): 1388-1400. <https://doi.org/10.1093/pcp/pcv054>.
- Fujii H, Verslues PE, Zhu JK. Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in *Arabidopsis*. *Plant Cell* 2007; 19(2):485-94. <https://doi.org/10.1105/tpc.106.04853> 8
- Fujimoto R, Taylor JM, Shirasawa S, *et al.*, Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc Natl Acad Sci* 2012; 109(18):7109-14.

- <https://doi.org/10.1073/pnas.1204464109>.
- Gardner L, Paterson E, Freidin A, *et al.* Variation in a single nucleotide polymorphism in the 5'UTR of growth and differentiation factor-5 (GDF-5) predicts clinical outcome at 3 months following acute knee joint injury. *Osteoarthritis and Cartilage* 2016; 24:S40-S41. <https://doi.org/10.1016/j.joca.2016.01.097>.
- Groszmann M, Greaves IK, Albertyn ZI, *et al.*, Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *Proc Natl Acad Sci* 2011; 108(6): 2617-2622. <https://doi.org/10.1073/pnas.1019217108>.
- Guo J, Ling H, Ma J, *et al.*, A sugarcane R2R3-MYB transcription factor gene is alternatively spliced during drought stress. *Sci Rep* 2017;7:41922. <https://doi.org/10.1038/srep41922>.
- Guyon-Debast A, Lécureuil A, Bonhomme S, *et al.*, A SNP associated with alternative splicing of RPT5b causes unequal redundancy between RPT5a and RPT5b among *Arabidopsis thaliana* natural variation. *BMC Plant Biol* 2010;10(1):158. <https://doi.org/10.1186/1471-2229-10-158>.
- Horstmann S, Ferrari S, Klempnauer KH. An alternatively spliced isoform of B-Myb is a transcriptional inhibitor. *Oncogene* 2000; 19: 5428–5434. <https://doi.org/10.1038/sj.onc.1203937>
- Kaya H, Shibahara KI, Taoka KI, *et al.*, FASCIATA genes for chromatin assembly factor-1 in *Arabidopsis* maintain the cellular organization of apical meristems. *Cell* 2001; 104(1):131-42. [https://doi.org/10.1016/S0092-8674\(01\)00197-0](https://doi.org/10.1016/S0092-8674(01)00197-0).
- Keurentjes JJ, Bentsink L, Alonso-Blanco C, *et al.*, Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* 2007; 175(2):891-905. <https://doi.org/10.1534/genetics.106.066423>.
- Lee T, Lee I. araGWAB: Network-based boosting of genome-wide association studies in *Arabidopsis thaliana*. *Sci Rep* 2018; 8(1): p.2925. <https://doi.org/10.1038/s41598-018-21301-4>.
- Li J, Li X, Guo L *et al.*, A subgroup of MYB transcription factor genes undergoes highly conserved alternative splicing in *Arabidopsis* and rice. *J Exp Bot* 2006; 57(6):1263-1273. <https://doi.org/10.1093/jxb/erj094>.
- Liu F, Zhao Q, Mano N, *et al.*, Modification of starch metabolism in transgenic *Arabidopsis thaliana* increases plant biomass and triples oilseed production. *Plant Biotechnol J* 2016; 14(3):976-985. <https://doi.org/10.1111/pbi.12453>.
- Marks MD, Wenger JP, Gilding E, *et al.*, Transcriptome analysis of *Arabidopsis* wild-type and gl3–sst sim trichomes identifies four additional genes required for trichome development. *Mol plant* 2009; 2(4):803-22. <https://doi.org/10.1093/mp/ssp037>.
- Maruta T, Yoshimoto T, Ito D, *et al.*, An *Arabidopsis* FAD pyrophosphohydrolase, AtNUDX23, is involved in flavin homeostasis. *Plant Cell Physiol* 2012; 53(6):1106-1116. <https://doi.org/10.1093/pcp/pcs054>.
- Massonnet C, Vile D, Fabre J, *et al.*, Probing the reproducibility of leaf growth and molecular phenotypes: a comparison

- of three *Arabidopsis* accessions cultivated in ten laboratories. *Plant Physiol* 2010; 152(4):2142-57. <https://doi.org/10.1104/pp.109.148338>.
- Méndez-Vigo B, Savic M, Ausín I, *et al.*, Environmental and genetic interactions reveal FLOWERING LOCUS C as a modulator of the natural variation for the plasticity of flowering in *Arabidopsis*. *Plant cell environ* 2016; 39(2):282-294. <https://doi.org/10.1111/pce.12608>.
- Nakashima K, Fujita Y, Kanamori N, *et al.*, Three *Arabidopsis* SnRK2 protein kinases, SRK2D/SnRK2. 2, SRK2E/SnRK2. 6/OST1 and SRK2I/SnRK2. 3, involved in ABA signaling are essential for the control of seed development and dormancy. *Plant Cell Physiol* 2009; 50(7):1345-1363. <https://doi.org/10.1093/pcp/pcp083>.
- Oeth P, Beaulieu M, Park C, *et al.*, iPLEX assay: Increased plexing efficiency and flexibility for MassArray system through single base primer extension with mass-modified terminators. *Sequenom Application Note* 2005; 27:8876-006. <http://www.usc.es/cegen/wp-content/uploads/2016/03/Sequenom-iplx-assay.pdf>
- Ogawa T, Yoshimura K, Miyake H, *et al.*, Molecular characterization of organelle-type Nudix hydrolases in *Arabidopsis*. *Plant Physiology* 2008; 148(3):1412-1424. <https://doi.org/10.1104/pp.108.128413>
- Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002; 30(17): 3894-3900. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC137415/>.
- Reumers J, Conde L, Medina I, *et al.*, Joint annotation of coding and non-coding single nucleotide polymorphisms and mutations in the SNP effect and PupaSuite databases. *Nucleic Acids Res* 2007; 36(suppl\_1), pp.D825-D829. <https://doi.org/10.1093/nar/gkm979>.
- Rodgers-Melnick E, Vera DL, Bass HW, *et al.*, Open chromatin reveals the functional maize genome. *Proc Natl Acad Sci* 2016;113(22):E3177-E3184. <https://doi.org/10.1073/pnas.1525244113>.
- Sakiyama T, Kohno T, Mimaki S, *et al.*, Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. *Int J Cancer* 2005; 114(5):730-737. <https://doi.org/10.1002/ijc.20790>.
- Sandoval FJ, Zhang Y, Roje S. Flavin nucleotide metabolism in plants monofunctional enzymes synthesize FAD in plastids. *J Biol Chem* 2008; 283(45):30890-30900. <https://doi.org/10.1074/jbc.M803416200>.
- Seitz A, Koch T, Nieselt K, *et al.*, SNPviz- Visualization of SNPs in proteins. *Genomics and Computational Biology* 2018; 4(1): e100048-e100048. <https://doi.org/10.18547/gcb.2018.vol4.iss1.e100048>
- Shen H, He H, Li J, *et al.*, Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell* 2012; 24(3):875-892. <https://doi.org/10.1105/tpc.111.094870>.
- Simkin AJ, Lopez-Calcagno PE, Davey PA, *et al.*, Simultaneous stimulation of sedoheptulose 1, 7-bisphosphatase, fructose 1, 6-bisphosphate aldolase and the photorespiratory glycine

- decarboxylase-H protein increases CO<sub>2</sub> assimilation, vegetative biomass and seed yield in *Arabidopsis*. *Plant Biotechnol J* 2017; 15(7):805-816. <https://doi.org/10.1111/pbi.12676>.
- Singh I, Smita S, Mishra DC, *et al.*, Abiotic stress responsive miRNA-target network and related markers (SNP, SSR) in *Brassica juncea*. *Front Plant Sci* 2017; 8:1943. <https://doi.org/10.3389/fpls.2017.01943>.
- Thomas PD, Campbell MJ, Kejariwal A, *et al.*, PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 2003; 13(9): 2129-41. <https://doi.org/10.1101/gr.772403>.
- Torii KU. Leucine-rich repeat receptor kinases in plants: structure, function, and signal transduction pathways. *Int Rev Cytol* 2004; 234:1-46. [https://doi.org/10.1016/S0074-7696\(04\)34001-5](https://doi.org/10.1016/S0074-7696(04)34001-5).
- Wang L, Si W, Yao Y, *et al.*, Genome-wide survey of pseudogenes in 80 fully re-sequenced *Arabidopsis thaliana* accessions. *PloS One* 7; 2012:e51769. <https://doi.org/10.1371/journal.pone.0051769>
- Zhao H, Fan D, Nyholt DR, *et al.*, Enrichment of SNPs in Functional Categories Reveals Genes Affecting Complex Traits. *Hum Mutat* 2016; 37(8):820-826. <https://doi.org/10.1002/humu.23007>.
- Zhou G, Chen Y, Yao W, *et al.*, Genetic composition of yield heterosis in an elite rice hybrid. *Proc Natl Acad Sci* 2012; 109(39): 15847-15852. <https://doi.org/10.1073/pnas.1214141109>.
- Zhu A, Greaves IK, Dennis ES, *et al.*, Genome-wide analyses of four major histone modifications in *Arabidopsis* hybrids at the germinating seed stage. *BMC genomics* 2017; 18(1): 137. <https://doi.org/10.1186/s12864-017-3542-8>.
- Zhu A, Greaves IK, Liu PC, *et al.*, Early changes of gene activity in developing seedlings of *Arabidopsis* hybrids relative to parents may contribute to hybrid vigour. *Plant J* 2016; 88(4): 597-607. <https://doi.org/10.1111/tbj.13285>.

**How to cite this article:**

Astha Gupta, Archana Bhardwaj, Samir V. Sawant and Hemant Kumar Yadav. 2019. Utilization and Characterization of Genome-wide SNP Markers for Assessment of Ecotypic Differentiation in *Arabidopsis thaliana*. *Int.J.Curr.Microbiol.App.Sci*. 8(06): 158-173. doi: <https://doi.org/10.20546/ijcmas.2019.806.020>