eTechnical Bulletin No. 3
(Ref.: No.F.PME/3-3/2021 dated 4.7.2021)

Genomics in Potato Improvement

Authors

Jagesh Kumar Tiwari
Sundaresha S
VU Patil
Hemant B. Kardile
Vinay Bhardwaj
Dalamu
Tanuja Buckseth
Vanishree G
Clarissa Challam
Rajesh Kumar Singh
SK Chakrabarti
Manoj Kumar

ICAR - CENTRAL POTATO RESEARCH INSTITUTE, SHIMLA (HP)

1. Genome sequencing

1.1 The Potato genome

Potato is the world's most important vegetable crop and the 3rd largest global food crop. The Potato Genome sequencing: ICAR-Central Potato Research Institute, Shimla had been the partner of the Potato Genome Sequencing Consortium (PGSC) comprising of 26 international institutes belonging to 14 countries. Dr. S.K. Chakrabarti was the country leader for the potato genome sequencing project. In India, the project was fully funded by the ICAR, New Delhi and was executed by ICAR-CPRI, Shimla. The consortium deciphered the complex genome of potato (tetraploid and highly heterozygous) that has been published in the high impact journal "Nature". This is the first genome of a plant belonging to Asterid clade of eudicot that represents 25% of flowering plant species. A total of 39,031 protein-coding genes have been predicted in the 840 Mb genome of potato. The potato genome data is freely available in public for research use at *http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml*.

Despite its importance as a food crop throughout the world, the genetics of many potato traits is poorly understood and is complicated by its highly heterozygous polyploid genome. Unlike diploid crops, most potato varieties have four copies of each of the 12 chromosomes. This makes it very difficult to follow inheritance patterns, especially in relation to the many complex traits with which breeders are compelled to work. Many important agronomic traits are poorly understood; genes affecting these traits remain largely undiscovered and their locations on the 12 chromosomes are often imprecise. The potential of modern genomics tools and a genome sequence in particular in alleviating such negative aspects of potato as a genetic system have long been realized. The sequencing and anchoring of the potato genome will provide a major boost to gaining a better understanding of potato trait biology and will underpin all future breeding efforts. Looking at the importance of this an international "Potato Genome Sequencing Consortium" (PGSC) was formed consisting of 14 countries including India. The PGSC intended to finish sequencing of 840 Mb long potato genome by 2010 and ICAR-CPRI represented India at the international consortium. The institute was assigned the task of completing chromosome 2 sequences, which is about 60.2 Mb long. A total of around 800-1,000 BAC clones each having an insert of ~ 120 kb were sequenced and analyzed by the institute. The task given to India represented by ICAR-CPRI was successfully taken up and completed under stipulated time. The first draft of the potato genome sequence was published in July, 2011. The salient features of the potato genome are as below:

- A high-quality draft potato genome sequence has been released. A unique and homozygous doubled-monoploid potato clone (DM) was employed to overcome the problems associated with genome assembly due to high levels of heterozygosity in potato. The sequenced and assembled genome amounts to the 86% of the 844-megabase potato genome.
- The potato genome contains ~39,000 protein coding genes. There is also evidence for alternative transcripts for as many as 60% of these genes.

- Sequenced potato genome present evidence for at least two genome duplication events indicative of a palaeopolyploid origin.
- The genome is highly duplicated and contains many short segments showing conserved synteny with other plants such as rice and grape.
- This is the first asterid genome to be sequenced and reveals 2,642 genes specific to this large angiosperm clade.
- The mapped 90% of the genome contains ~90% of the 39,031 predicted genes.
- A total of 917 superscaffolds have been anchored, ordered and assembled into pseudomolecules corresponding to the 12 potato chromosomes.
- Additional efforts have been put to improve the chromosome 4 coverage and assembly by selectively choosing and sequencing 82 DM BACs spanning gaps between chromosome 4 superscaffolds.
- Development of a highly-dense potato map linked to the genome assembly. This includes the development of over 4000 new sequence-based genetic markers, the vast majority of which are linked directly to the genome assembly.
- The DM Whole Genome Shotgun sequencing project has been deposited at DDBJ/EMBL/GenBank under the accession AEWC00000000. Genome sequence and annotation can be obtained and viewed at http://potatogenome.net.

1.2 Impact of potato genome sequencing

The sequencing of the potato genome has opened up new vistas for the way potato genetics and breeding are performed making potato plant more amenable to modern genetics and genomics tools. The potato collections can now be more efficiently mined for novel alleles and beneficial traits of economical and industrial importance. The integrated genome sequence and genetic reference map will allow trait phenotype loci or QTLs defined by sequence based markers to be linked to specific genetic and physical regions of the genome. Such regions can be then used to define further markers for fine-scale mapping, or candidate genes can be sought directly from the genome sequence and associated annotation data. This step change, facilitating sequence-based genomics and aiding molecular breeding in potato, would accelerate trait-gene discovery and gene isolation. This would further shorten the time to breed new varieties and also significantly improve parental genotypic assessment. Genome tagged molecular marker studies will be more meaningful and enable more accurate estimates of population genetic and LD parameters. The shift towards sequence based polymorphism rather than fragment based, will virtually replace centimorgan (cM) position by sequence co-ordinates and greatly increase the information output and accuracy of mapping procedures. The integrated potato genetic and physical reference map forms an important resource for linking to all current and future genetic mapping efforts by the potato community and will help to alleviate many of the complicating aspects of potato as a genetic system. With the release of the genome of the other economically important Solanaceous plant tomato, comparative linkage mapping and in depth sequence based synteny analysis among *Solanaceae* will be feasible. Given the biological and economic importance of many Solanaceous species and the diversity of their phenotypes/products (agriculturally useful parts tubers, berries etc., growth habits, wide geographical growing range, clonal propagation,

regeneration), comparative Solanaceous genomics will provide a fundamental framework for tackling both applied and basic questions.

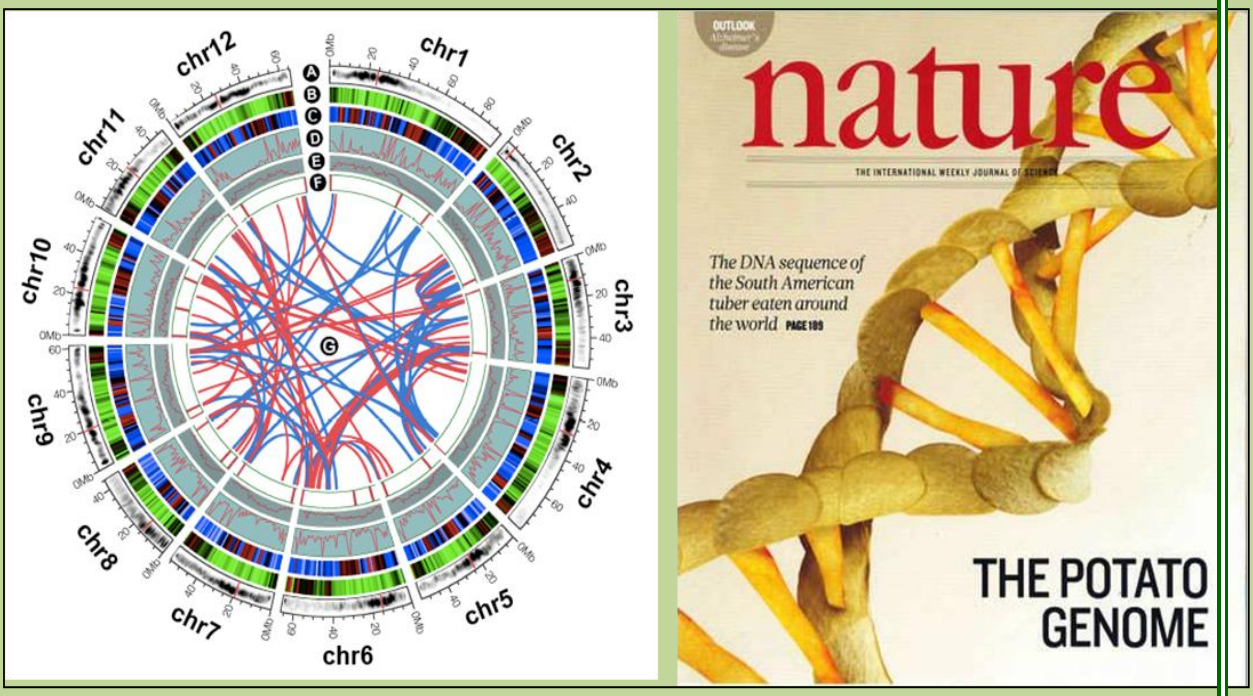


Fig. 1. A concise diagram of the potato genome & Cover page of Nature (2011) 475: 189-195

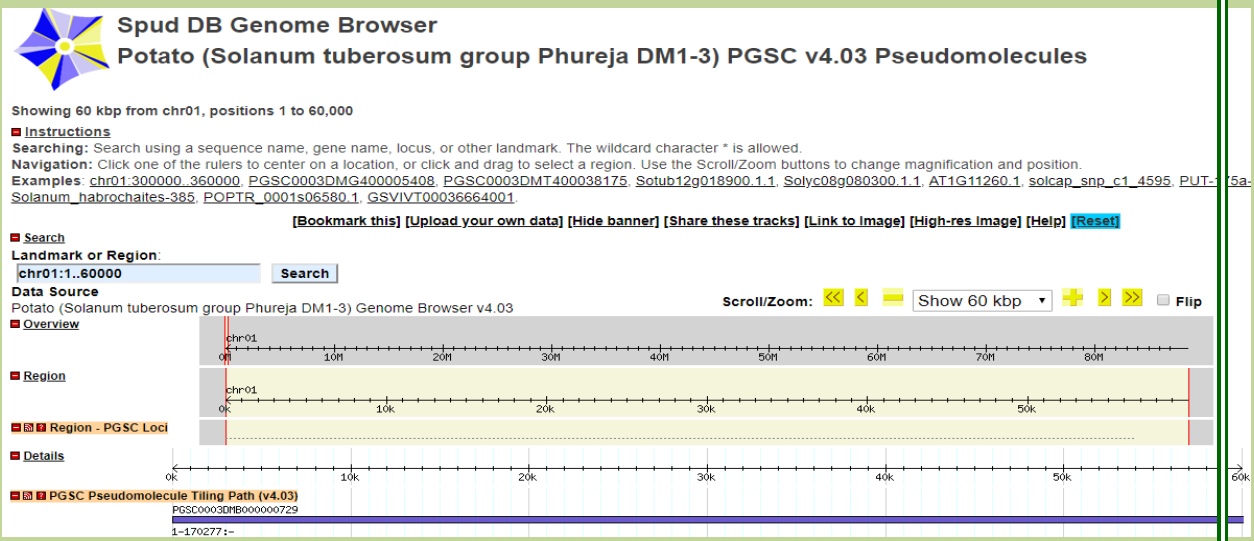


Fig. 2. Potato Genome browser http://solanaceae.plantbiology.msu.edu/cgi-bin/gbrowse/potato/

1.3 Genome sequencing of potato dihaploid C-13

Whole genome sequencing of androgenic potato dihaploid ‘C-13’ and identification of genomic variation (SNPs and Indels) with the reference potato genome: The *Solanum*

tuberosum dihaploid 'C-13' (2*n* = 2*x* = 24) was developed from potato cv. Kufri Chipsona-2 (2*n* = 2*x* = 48) by anther culture at our institute. The dihaploid 'C-13' has been used as fusion parent for development of somatic hybrids. We sequenced the whole genome of 'C-13' in house using Illumina approach. On average recovery of 80.68% reference assembly only with short reads data indicates higher quality of the assembly with 60X coverage, Phred score > Q30 and average GC content 44%. Overall, ~810 Mb genome size with 30,241 predicted genes have been identified in the C-13 genome. A total of 11,22,388 SNPs and 48,145 Indels were also identified in C-13.

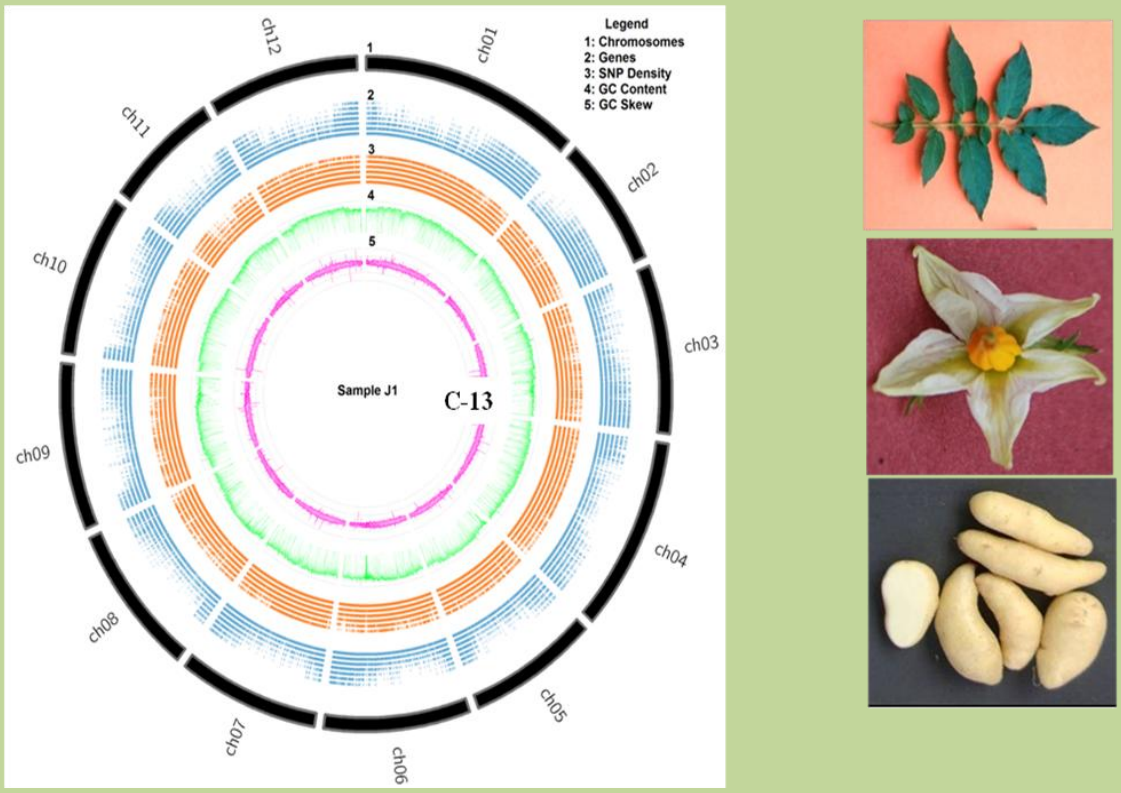


Fig. 3. Genome sequencing of potato dihaploid C-13 shown with circus plot

1.4 Bisulfite sequencing for DNA methylation analysis in potato

Whole genome bisulfite sequencing of potato cv. Kufri Bahar to analyze DNA methylation for tuber shape variation was analysed. DNA methylation pattern was analyzed in potato cv. Kufri Bahar to explain the observed tuber shape variation in tissue culture derived microplants. Methylation sites were analysed at CpG, CHG and CHH contexts where H is A, C or T. Differential methylation regions analysis showed more methylation in tissue culture-derived tubers (cytosine sites: 3622,290,882) than original (cytosine sites: 3109,549,323). Maximum differentially methylated sites were observed on chromosome 5. Differential methylation regions were CHG (9320) followed by CpG (4572) and HH (2453).

Fig. 4. DNA methylation profiling in potato cv. Kufri Bahar for tuber shape variation

1.5 Genome sequencing of late blight pathogen (*Phytophthora infestans*)

After the successful completion of potato genome sequencing, the institute undertook the responsibility of deciphering the genomes of important potato pathogens under different projects. Having the in-depth genome knowledge of pathogens would greatly help in understanding the pathogenicity and their evolution. The genome sequencing of major potato pathogens by the institute included the late blight pathogen *Phytophthora infestans* (A2 mating type) and complete mapping of its mitochondrial type revealed the recent evolution of the A2 mating type in India. Sequencing of mitochondrial genome of *P. infestans* (HP10-31)- belonging to haplotype-Ia and A_2 mating type- using Roche 454 resulted 40Mb data with coverage of ≅100X. Data was assembled using GS DeNova Assember and GS Ref with mitochondrial genome of Ia as reference, yielding single mega scaffold of 37,767 bp covering entire genome with 22.38% GC content. Total 53 protein coding genes were predicted on both strands which included 25 tRNA, 2 rRNA, and 18 respiratory protein coding genes.

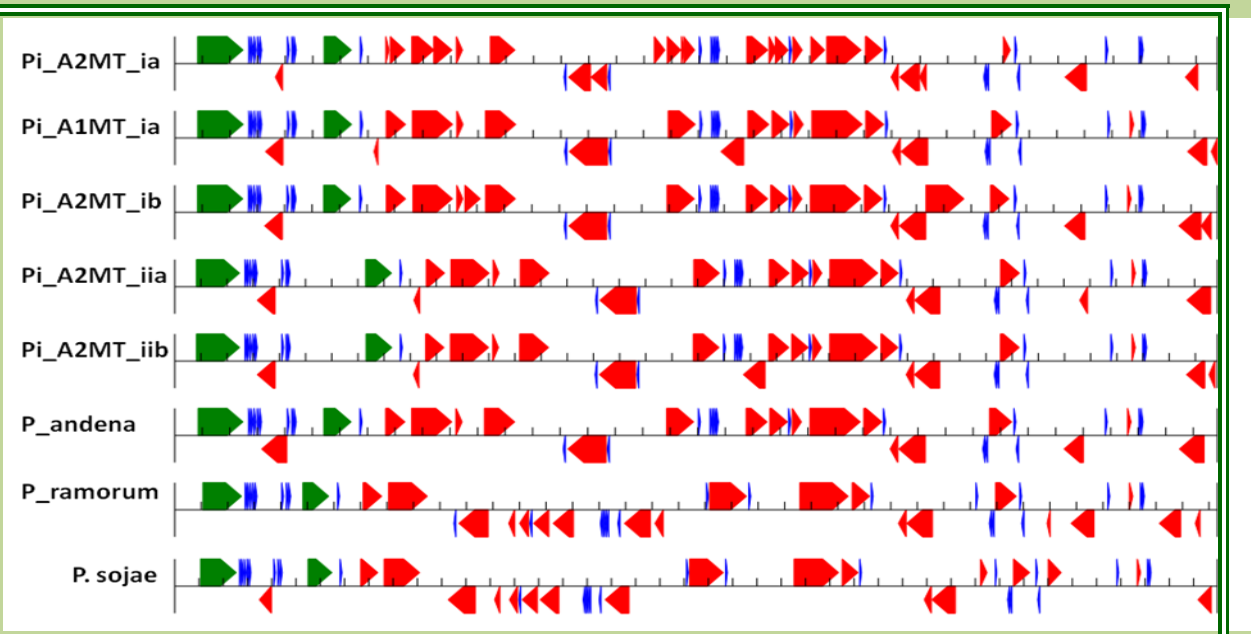


Fig. 5. Mapping of mitochondrial genome of potato late blight pathogen (*Phytophthora infestans* A_2 mating type) along with *P. andena, P. ramorum* and *P. sojae*

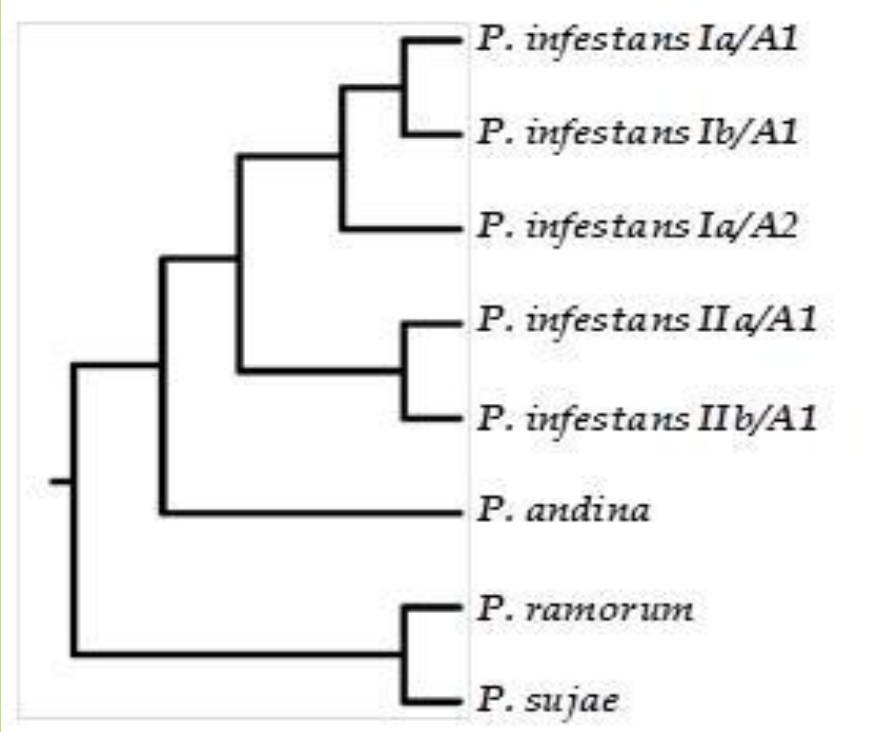


Fig. 6. Diversity in late blight pathogen (*Phytophthora infestans*) along with other related species

1.6 Genome sequencing of bacterial wilt pathogen (*Ralstonia solanacearum*)

The bacterial wilt pathogen of potato, *Ralstonia solanacearum* formally known as *Pseudomonas solanacearum* and *Burkholderia solanacearum* is a gram-negative, chemo-

organotroph, phytopathogenic ß-proteobacterium with an unusual broad host range was also sequenced (genome size: ~ 5.8 MB). Three of the four phylotypes of *R. solanacearum* are known to cause bacterial wilt of potato in India. Sequencing of four strains was completed namely RS2 (Phylotype II), RS25 and RS48 (Phylotype I) and RS75 (Phylotype IV), which were isolated from brown-rot infected potato tubers obtained from different parts of the country. The availability of the reference genomes of more and more strains of RsSC (*R. solanacearum* species complex) would greatly aid in epidemiological/quarantine studies and in gaining understanding on their origin, evolution, intra and inter-relationship within the complex and their interactions with plants.

Strain	Phylotype/ Biovar	Raw reads	HQ reads	Total Data (Mb)	Genome coverage (%)	Ref Genome (strain)
RS2	Phylotype IIB biovar 2	729,477	724,092	186	87.82	Po82
RS25	Phylotype I, iovar 4	566,754	561,461	223	90.03	GMI1000
RS48	Phylotype I, biovar 3	320,126	317,068	107	90.25	GMI1000
RS75	Phylotype IV, biovar 2T	1,265,491	1,264,138	459	90.00	PSI07

Fig. 7. Genome statistics of *Ralstonia solanacearum* and its disease symptom

1.7 Genome sequencing of dry rot pathogen (*Fusarium sambucinum*)

Fusarium sambucinum Fuckel [syns. *F. sulphureum* Schlechtlend., *F. roseum* var. *sambucinum* (Fuckel) SN. & H.) (teleomorph Gibberella pulicaris (Fr.:Fr.) Sacc.)], a cosmopolitan soil saprophyte that cause dry rot, an economically important postharvest fungal disease, affecting potato tubers worldwide was sequenced. The yield losses attributed to dry rot in storage range from 10 to 25%, with almost complete loss of stored tubers in some cases. High temperature (30-40°C) during harvesting in plains and humid condition caused by rain after June favours dry rot in potato storage. This is the first report of whole genome sequencing of *F. sambucinum* (F-4 strain) in the world and the availability of the

genome sequence is certain to be an important resource in epidemiological and quarantine studies. Its genome size is 52 MB and showed ~80% similarity with *F. graminearum*. A more detailed analysis of this genome and a comparative analysis with other *Fusarium* sp. genomes could expand our understanding of evolutionary relationships to obtain greater insights into their origins and pathogenicitiy.

Fusarium sp.	Strain	Chrs.	Sexual Stage	Genome size	(%) GC Content	Predicted genes	Total rRNAs	Reference
F. graminearum	**PH-1**	**4**	***Gibberella zeae***	**36.45**	**48.33**	**13,332**	**88**	**Li-Jun et al. 2010**
F. oxysporum	**4287**	**15**	**Asexual**	**61.36**	**48.40**	**17,735**	**121**	**Ma et al. 2010**
F. verticillioides	**7600**	**11**	***Gibberella moniliformis***	**41.78**	**48.70**	**14,179**	**75**	**Brown et al. 2008**
F. fujikuroi	**IMI58289**	**12**	***Gibberella fujikuroi***	**43.90**	**48.30**	**14,017**	**32**	**Jeaong et al. 2013**
F. sambucinum	**F-4**	**-**	***Gibberella pulicaris***	**52.00**	**42.81**	**12,845**	**64**	**This study**
F. solani	**77-13-4**	**17**	***Nectria haematococca***	**51.28**	**50.78**	**15,707**	**70**	**Coleman et al. 2009**

Fig. 8. Genome statistics of *Fusarium sambucinum* and its disease symptom

1.8 Genome sequencing of black scurf & stem canker pathogen (*Rhizactonia solani*)

The institute has completed the whole genome sequencing of the fungus *Rhizoctonia solani* Kühn (teleomorph Thanatephorus cucumeris ([A. B. Frank] Donk) (Deuteromycetes, Mycelia Sterilia). *R. solani* AG3-PT (The potato type) is a subgroup of AG3 and main causal agent of stem, stolon canker and black scurf of potato worldwide including India. It is both tuber and soil borne. Loss of quantity and in particular quality caused by black scurf disease reaches even up to 50%. All these pathogen genome studies and further, functional characterization of potential effector genes would greatly help in determining their roles in pathogenesis and ultimately better management of the diseases.

R. solani-AG	Strain	Host	Size (Mb)	Contigs	GC (%)	Genes	Reference
AG3-PT	RS-20	Potato	55.85	30,594	48.30	11,431	This study
AG1-IA	AG1IA	Rice	37.80	2,648	47.61	10,489	Zheng et al. 2013
AG1-IB	7/3/14	Rice	52.74	23,355	48.10	12,616	Wibberg et al. 2015
AG8	WAC10335	Potato	35.92		48.80	13,964	Hane et al. 2014
AG3	Rhs_1AP	Potato	52.50	6,040	48.40	12,726	Cubeta et al. 2014
AG2 IIIB	BBA69670	Sugar Beet	56.02	5,826	48.34	11,897	Wibberg et al. 2016

Fig. 9. Genome statistics of *Rhizactonia solani* and its disease symptom

1.9 Genome sequencing of glasshouse potato aphid *Aulacorthum solani*: *Aulacorthum solani* (Kaltenbach) is a glasshouse potato aphid or foxglove aphid is a vector of 45 different plant viruses including most dreaded viral diseases like PVY and PLRV. Draft genome sequence data of A. solani resulted into 89,260 scaffolds. The haploid genome size of A. solani was 316,399,729 bp with 42.63% GC content, and predicted 22,021 protein coding genes. Functional annotation of the predicted genes was performed using BLASTx program of the NCBI, which arrived at 20,359 known genes and 1,662 unknown genes. The high quality reads were mapped to M. persicae genome and identified a total of 17,485 filtered SNPs in A. solani. The draft whole genome sequencing of A. solani was deciphered to contribute to a better understanding of vector biology, genetics, aphid-plant interactions and management strategies of this aphid.

Fig. 10. Glasshouse potato aphid *Aulacorthum solani*

2. Functional Genomics

The term functional genomics can be referred to as the "development and application of global (genome-wide or system-wide) experimental approaches to assess gene function by making use of the information and reagents provided by structural genomics" It involves the use of high-throughput methods for the study of large numbers of genes (ideally the entire set) in parallel. Indirect information on cellular or developmental function can be obtained from spatial and temporal expression patterns; for example, the presence of mRNA and/or protein in different cell types, during development, during pathogen infection, or in different environments. The subcellular localization and posttranslational modifications of proteins can be informative as well. The institute has mastered in many of the functional genomics techniques by standardizing them for potato. Since, the potato genome sequence is available for the public it can very well be used for functional validation. The techniques used for functional genomics in potato by the institute include RNASeq and microarray at the whole genome level and reverse genetic approaches like gene knockout by RNAi (RNA interference) and VIGS (virus induced gene silencing) at the gene level.

With the completion of genome sequencing of more and more organisms, research focus has now been shifted from sequencing to delineating the biological functions of all genes coded within the genome of a particular organism. Methodologies of biological research are evolving from "one gene in one experiment" to "multiple genes in one experiment" paradigm. Molecular biology research evolves through the development of the

technologies used for carrying them out. It is not possible to research on a large number of genes using traditional methods. Earlier DNA microarray and now RNA sequencing are the such technologies which enables the researchers to investigate and address issues which were thought to be non traceable. One can analyze the expression of many genes in a single reaction quickly and in an efficient manner. It can also be used to understand the fundamental aspects underlining the heat tolerant potato, disease resistance, tuberization and other physiological responses occurring in the functioning of the potato cells.

2.1 Late blight resistance

2.1.1 Identification of novel genes by microarray: A typical microarray experiment involves the hybridization of an mRNA molecule to the DNA template from which it is originated. Many DNA samples are used to construct an array. The amount of mRNA bound to each site on the array indicates the expression level of the various genes. This number may run in thousands. All the data is collected and a profile is generated for gene expression in the cell. Microarray analysis for late blight resistant Indian potato cv. Kufri Girdhari revealed up-regulation of 2,344 genes post-inoculation compared to pre-inoculation stage. Molecular chaperones played critical role in controlling resistance in Kufri Girdhari. Besides, identified late blight resistance genes in somatic hybrid [C-13 (+) S. pinnatisectum]. A total of 5810 statistically significant genes ($p \leq 0.05$), of which 2101 genes ($\geq$ 2-fold) were up-regulated and 3709 genes were down-regulated. It was observed that defence responsive genes played a key role in late blight resistance mechanism in potato somatic hybrid clone P-7.

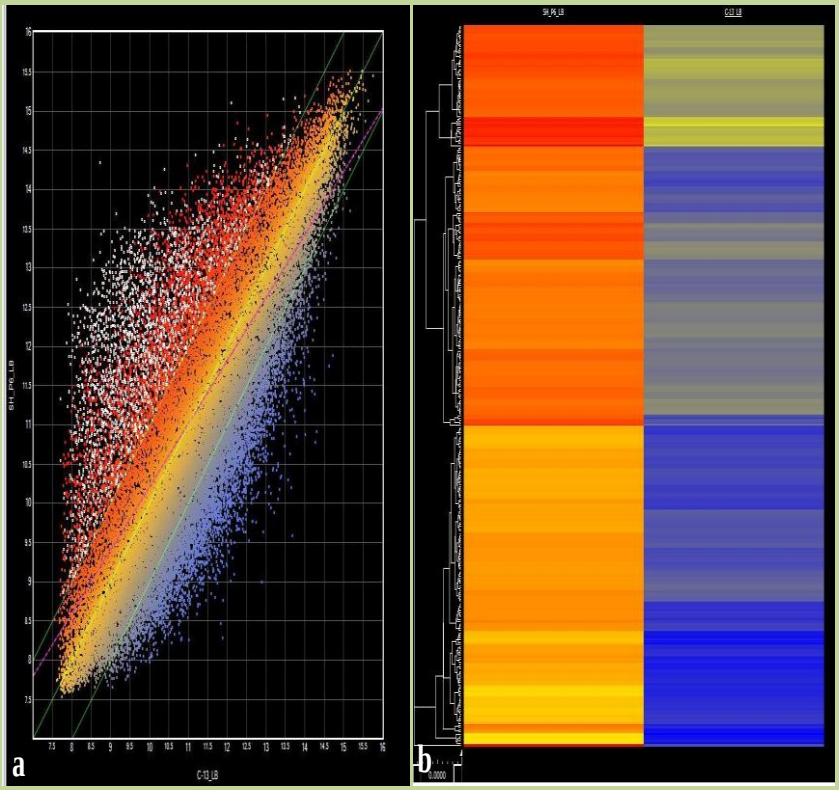


Fig. 11. Microarray analysis showing distribution of genes in somatic hybrid

2.1.2 MicroRNAs: Late blight is the most devastating disease of potato caused by Phytophthora infestans. MicroRNAs have been shown to play a significant role in local defense, but their association with SAR is unknown. We investigated the role of miR160 in local and SAR responses to P. infestans infection in potato. MiR160 is associated with local defense and systemic acquired resistance of potato against Phytophthora infestans infection. Study demonstrates that miR160 plays a crucial role in local defense and SAR responses during the interaction between potato and P. infestans. In addition, identified miRNAs (miR395, 821, 1030, 1510, 2673, 3979, 5021 and 5213) and their targets in somatic hybrid 'C-13' (+) S. pinnatisectum for late blight resistance by in silico approach. Majority of the predicted target genes of these miRNAs are involved in different biological functions, including disease resistance proteins (NBS-LRR domains) and transcription factors families.

2.1.3 Carbohydrate binding module 1 protein (CBD1) based antiserum: Early detection of Phytophthora infestans in potato using carbohydrate binding module 1 protein (CBD1) based antiserum is a novel tool for late blight management. Desirable and successful management strategies against P. infestans depend on understanding of its biology, disease cycle and more importantly rapid and accurate detection at early stage of infection. The CBD1 based antiserum can be effectively used for early detection of the disease in the standing crop and for taking decision for scheduling fungicide spray. Recombinant DNA technology was used for easier production of this specific antigen from bacterial clones carrying the CBD genes and higher sensitivity of pathogen detection was achieved.

2.1.4 Formulation of dsRNA: DsRNA formulation is as an alternative and novel protection strategy for potato late blight. Phytophthora infestans evolves rapidly and adapts quickly to the host background and new fungicide molecules because of predominance of transposable elements in its genome. RNAi technology offers an innovative approach for development of dsRNA solution as a novel protection strategy. Spraying of dsRNA based formulation provides an environmentally safer strategy for late blight management in potato.

2.1.5 Identification of novel genes by RNA sequencing: Identified genes for late blight resistance in potato somatic hybrids by transcriptome (RNA sequencing) analysis. Potato somatic hybrids (C-13 + S. pinnatisectum; C-13 + S. cardiophyllum) and parent C-13 were analyzed for late blight resistance by total RNA sequencing. Plants were challenge inoculated with the pathogen in control chamber and genes were analyzed at before and after inoculation (96 h). A total of 110 potato genes were found statistically significant ($p < 0.01$) and differentially regulated in various samples for late blight resistance.

2.1.6 Identified genes for late blight resistance in wild Solanum species by allele mining: Late blight, the most devastating disease of potato caused by Phytophthora infestans, can be managed by introduction of resistance (R) genes from the wild Solanum species into the cultivated potato. Identified 17 RB-homologous gene fragments in 11 wild species (S. chacoense, S. pinnatisectum, S. polyadenium, S. trifidum, S. cardiophyllum, S. lesteri, S.

huancabambense, S. verrucosum, S. jamesii, S. polytrichon and S. stoloniferum) of total 44 wild potato species. The RB gene was isolated from the wild Solanum bulbocastanum. The isolation and characterization of 17 RB-homologous gene fragments (NCBI sequence No. KJ472305-KJ472320 and KJ610025) from wild potato species may serve as an important genomic resource for the novel gene discovery in late blight resistance breeding program.

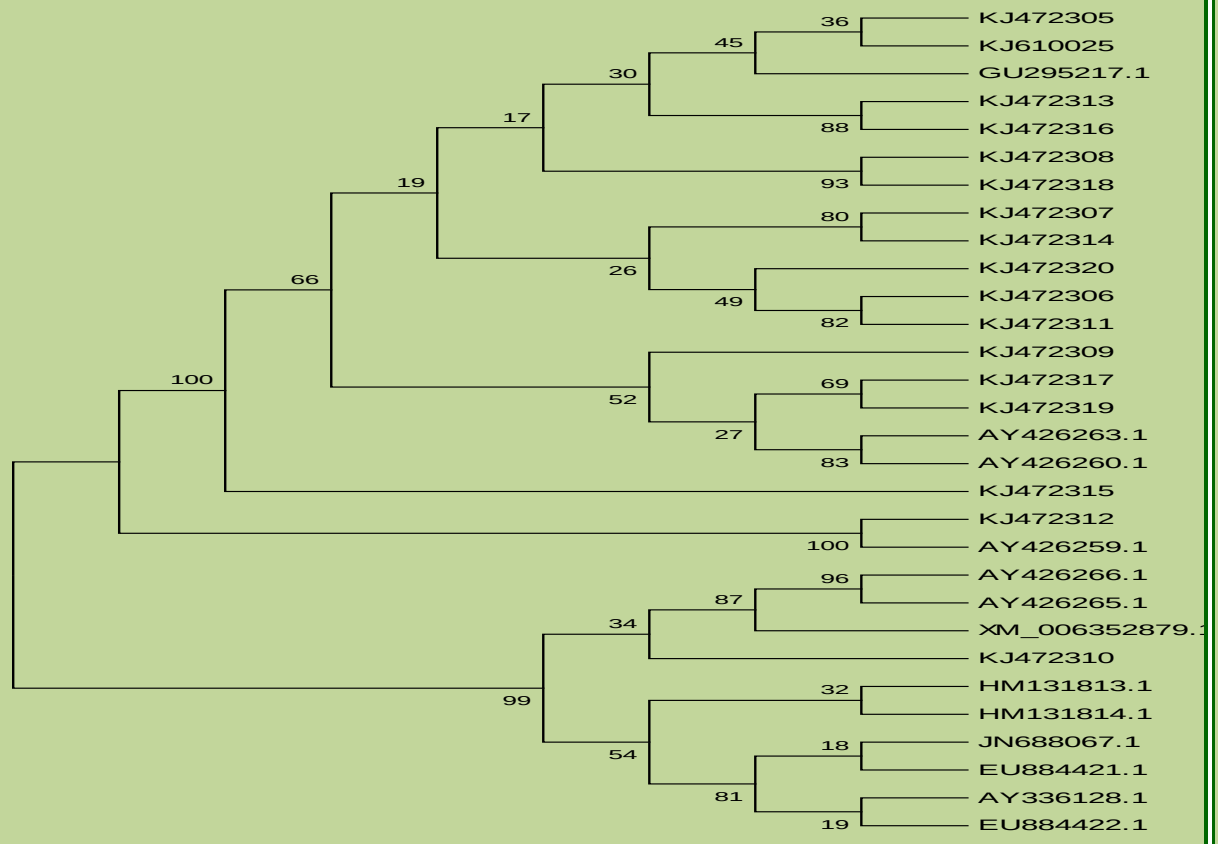


Fig. 12. A cluster analysis based on the Neighbor-Joining coefficient showing relationship between the nucleotide sequences of 17 *RB*-homologues and the reference genes.

2.2 Virus resistance

Microarray analysis for differentially regulated genes in response to Apical leaf curl virus resistance. Apical leaf curl disease, caused by tomato leaf curl New Delhi virus-[potato] (ToLCNDV-[potato]), is one of the most important viral diseases of potato in India. Genetic resistance source for ToLCNDV in potato is not identified so far. Microarray analysis showed that a total of 1111 genes and 2588 genes were differentially regulated in Kufri Bahar (resistant) and Kufri Pukhraj (susceptible), respectively. We identified a total of selected 680 genes in kufri Bahar, in response to ToLCNDV-potato infection. These altered transcripts were involved in stress responses, signal transduction pathways, protein binding, cellular transport and metabolic process. These genes will be used for future resistance breeding programme for development of varieties resistance to viral diseases

2.3 Potato tuberization

2.3.1 Identification of novel genes by microarray: Identified genes controlling potato tuberization in somatic hybrids ['C-13' (+) *S. etuberosum*] by microarray. A total of 468 genes (94 up-regulated and 374 down-regulated) was identified that was statistically significant and differentially expressed in tuber-bearing potato somatic hybrid (E1-3) versus control non-tuberous wild species *S. etuberosum* (Etb). Overall, findings showed that candidate genes induced in leaves of E1-3 were implicated to the tuberization process such as transport, carbohydrate metabolism, phytohormones and transcription/ translation/ binding functions.

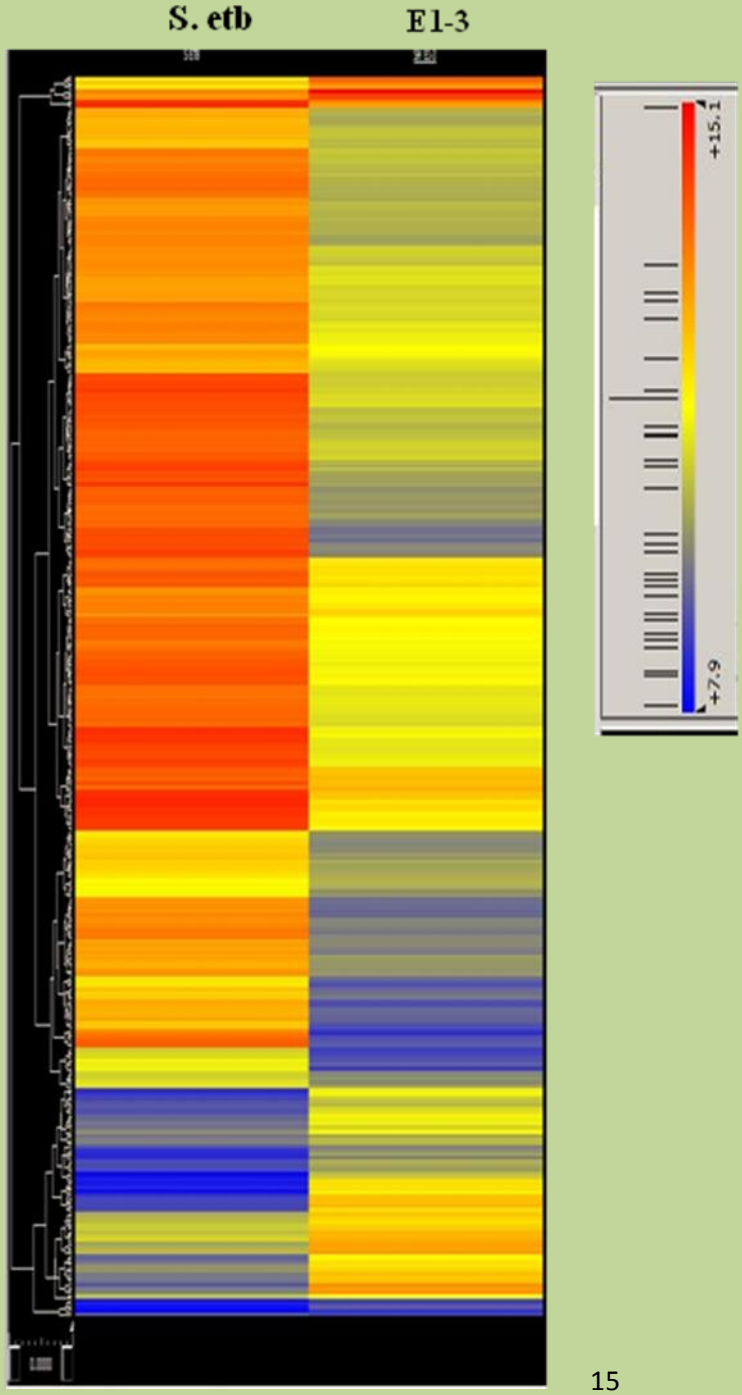
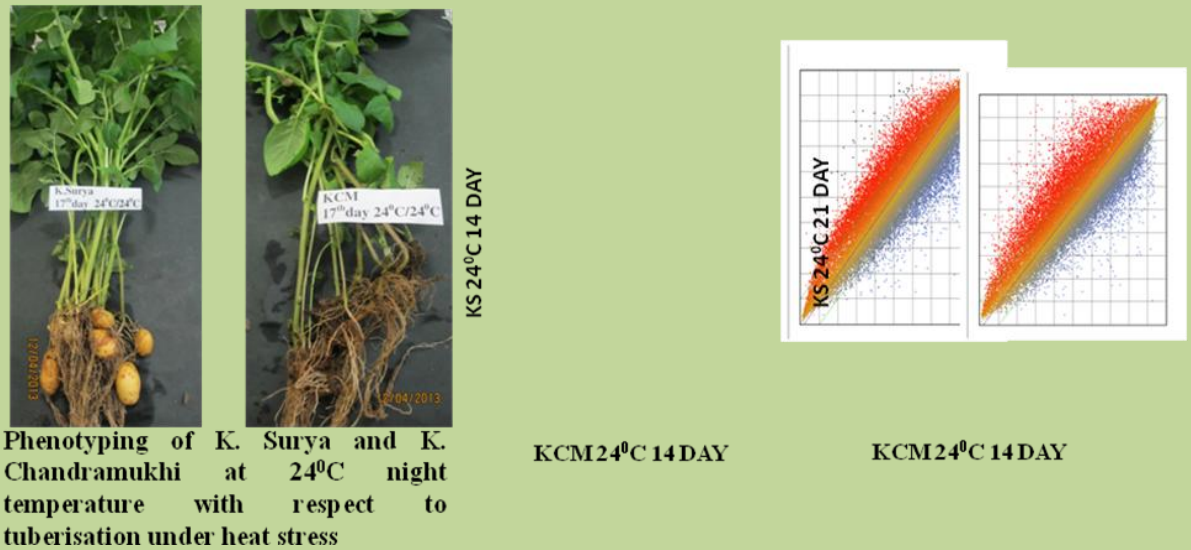


Fig. 13. Heat map profile shows differentially expressed genes ($p \leq 0.05$) in somatic hybrid E1-3 versus control *Solanum etuberosum* (S. etb).

2.3.2 Identification of heat tolerant genes by microarray: Identification of genes and pathways affected by high temperature is crucial for developing thermo tolerant cultivars. Identified heat tolerant genes in potato in response to high night temperature for tuberization. Night temperature beyond 20 ^{0}C drastically reduces tuber formation, constraining potato cultivation in tropics and subtropics (heat tolerant: Kufri Surya vs. heat sensitive: Kufri Chandramukhi). Microarray gene expression analysis showed a total of 2500 genes were differentially expressed on 21 days and 4096 genes on 14 days after stress. This study provided useful information on potato tuberization at elevated night temperatures (24^{0}C) and made available a framework for further investigations into heat stress in potato.

Two cultivars with contrasting tuberization behaviour at night temperatures (24^{0}C) were selected for gene expression analysis and identified a total of 2500 genes were differentially expressed. The selected few genes (StUGT+StBHP+StSP6A+ StSSH2+StWTF) were validated through VIGS technology and proved their role in tuberisation at elevated night temperature. This study provided useful information on potato tuberization at elevated temperature and makes available a framework for further investigations into heat stress in potato for developing heat tolerant varieties

2.3.3 Identification of drought tolerant genes: Transcriptomes were analysed by RNA sequencing for effect of drought stress on potato plants during prestolonization stage in cv. Kufri Jyoti. Drought stress (both 50 % and 25 % field capacity) increased root system on potato plants during pre-stolonization stage. A large number of differentially expressed genes were identified in response to drought tolerance in potato.



Phenotyping of K. Surya and K. Chandramukhi at 24^{0}C night temperature with respect to tuberisation under heat stress

Fig. 14. Transcriptional expression analysis of potato tuberization in response to high (24^0 C) night temperature by microarray analysis

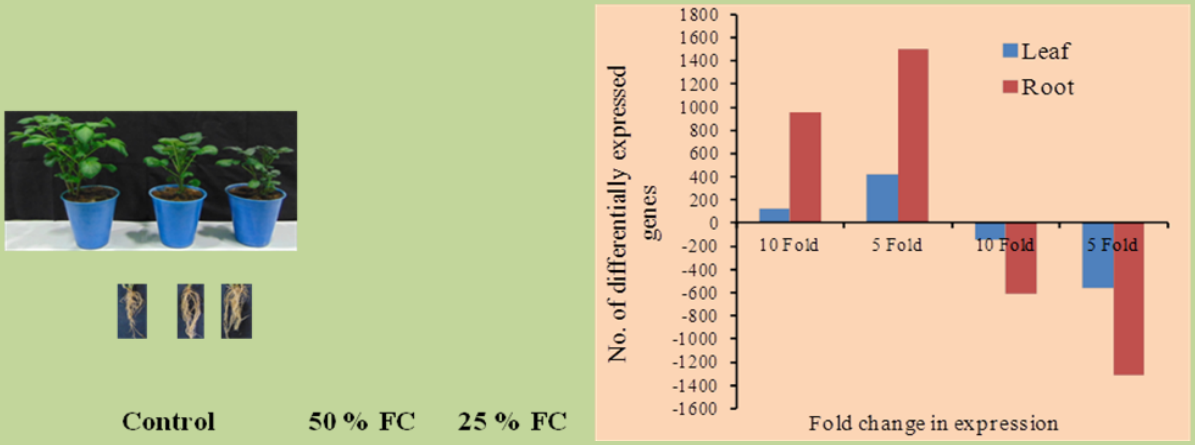


Fig. 15. Effect of drought stress on potato plants during pre-stolonization stage (cv. Kufri Jyoti) and gene expression pattern

2.4 Improving nitrogen use efficiency (NUE) of plant

2.4.1 Genomics for improving NUE in potato: Potato is an N fertilizer responsive crop to produce high tuber yield. The excessive use of N can results in environmental damage and high cost of production, hence improving NUE of potato plant is one of the sustainable options to address these issues and increase yield. Advanced efforts have been undertaken to improve NUE in other plants like *Arabidopsis*, rice, wheat and maize. Conversely, in potato NUE studies have predominantly focused on agronomy or soil management. Focus of this concept is to adapt knowledge gained from other plants to carry out investigation of N metabolism and associated traits in potato with the aim to improve NUE in plants grown under aeroponics applying integrated genomics, physiology and breeding methods.

2.4.2 Identified novel genes RNA sequencing: Potato varieties Kufri Jyoti (N inefficient) and Kufri Gaurav (N efficient) were grown in aeroponics with our solution (patent applied), without N (starvation), low N and high N. Transcriptomes (total RNA sequencing) were analyzed in leaf, root and stolon tissues to identify genes and regulatory elements controlling NUE in potato. Under N starvation condition, 233 genes were up regulated, whereas 1188 genes were down regulated in leaf tissues of Kufri Jyoti compared to N sufficient condition. In root tissues of Kufri Jyoti, 645 genes were down regulated and 250 genes were up-regulated under N starvation than N sufficient conditions.

2.4.3 Identified microRNAs for improving NUE in potato by small RNA sequencing: Potato variety Kufri Jyoti was grown in aeroponics with low nitrogen (1 mM) and high nitrogen (25 mM). Root and shoot tissues were analyzed by NextSeq 500 (Illumina) for small/micro RNA involved in N metabolism. We identified unique 52, 47, 52, 47 known miRNAs from high N Root, low N root, high N shoot and low N shoot samples, respectively. Using psRNATarget finder we could able to predict 34,135 unique targets from 723 unique microRNAs.

Fig. 16. Plant growth in aeroponics to investigate genes associated with N metabolism and to improve NUE in potato

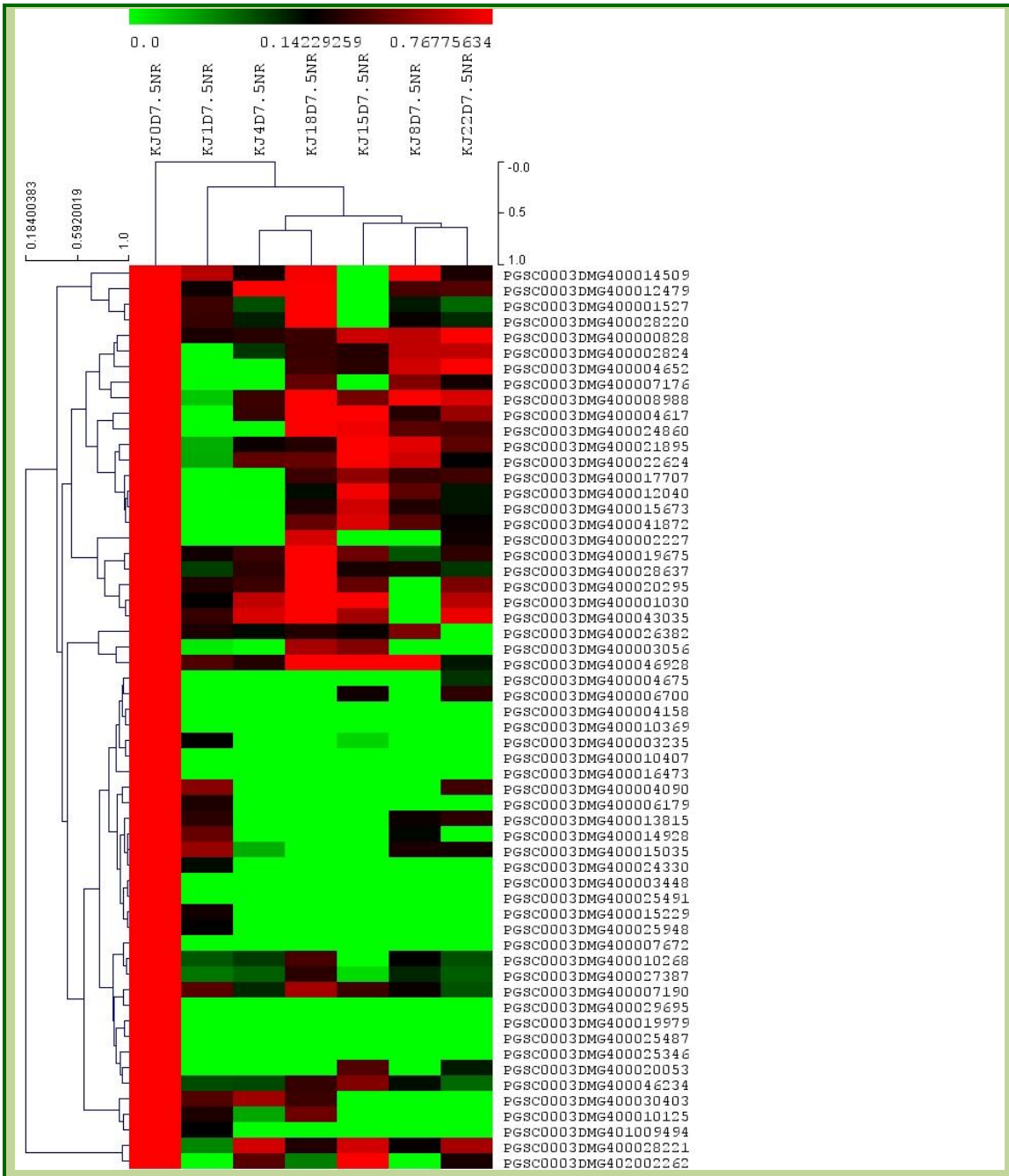


Fig. 17. Heat map showing up-regulated and down-regulated genes for high and low N condition in potato grown in aeroponics

2.4.4 Cloning and sequence variation analysis of genes involved in nitrogen metabolism in potato: Sequence variation was analyzed for genes involved in nitrogen (N) metabolism in potato for future breeding purpose. Two contrasting potato varieties Kufri Jyoti (N inefficient) and Kufri gaurav (N efficient) were used in the study. Of the total 17 PCR primers tested for N metabolism genes in the both varieties, only single, distinct and un-fractioned 12 fragments amplified by six primers representing five genes (nitrate transporter-NRT, ammonium transporter-AMT, nitrate reductase-NR, nitrite reductase-NIR and asparagines synthetase-AS) involved in N metabolism were cloned and sequenced. Following

sequence analysis, non-redundant sequences with uninterrupted open reading frames of 12 '*N*-homologous genes' were identified to the known N metabolic pathways genes. Thus, the identified 12 *N*-homologous genes may serve as an important genomic resource for novel gene/marker discovery and would be useful for MAS in potato with better nitrogen use efficiency (NUE).

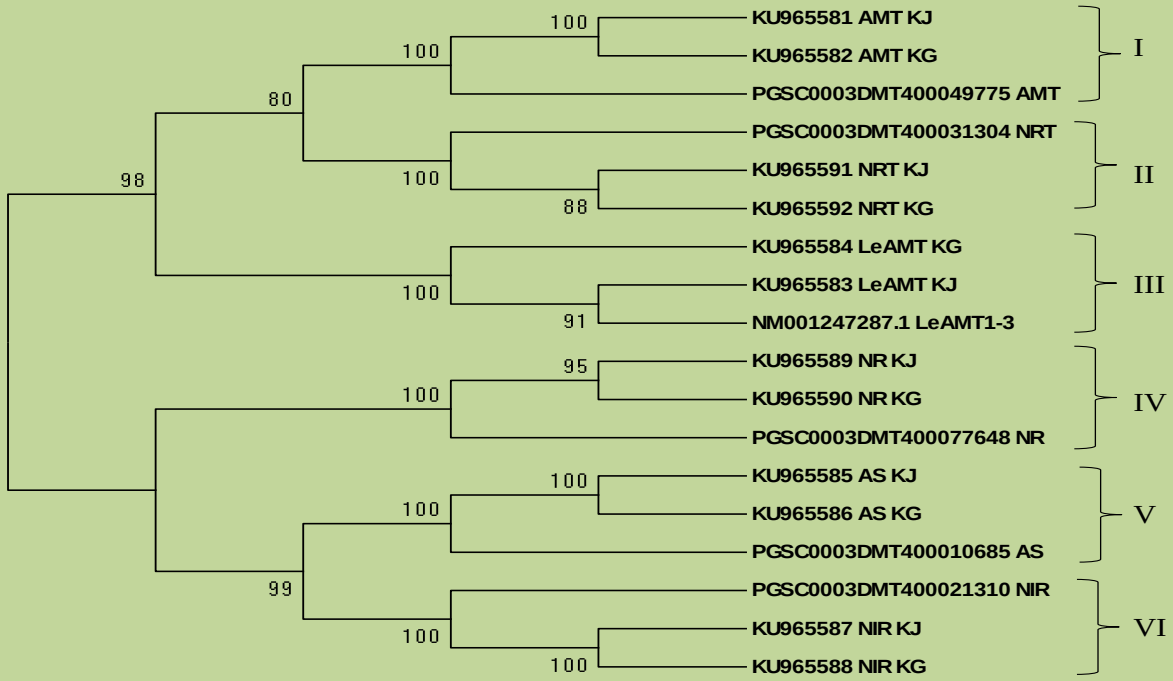


Fig. 18. A cluster analysis showing relationship in N metabolism genes.

3. Virus induced gene silencing (VIGS)

VIGS is a versatile too for understanding the functionality of biotic and abiotic stress genes in potato genome. VIGS is a post-transcriptional gene silencing (PTGS) method used by plants as a defense mecha¬nism by targeting the integrity of invading viruses. It involves cloning a short cDNA sequence from gene of interest into a viral delivery vector and transfecting the plant using Agrobacterium. A double-stranded RNA (dsRNA) is synthesized which is further degraded by plant Dicer-like enzymes into small interfering RNA (siRNA) molecules resulting in activation of PTGS thus leading to generation of siRNA homologous to the target gene which finally result into silencing of the endogenous plant gene. In potato, PVX (Potato Virus X) and TRV (Tobacco Rattle Virus) vectors have been found suitable for VIGS based silencing. The main advantages of VIGS include its low cost and rapid performance by identifying a loss of function phenotype for a particular gene within a single generation. Since, its expression is transient in nature; therefore, it does not require the laborious transformation procedures for the development of transgenic plants. Therefore, it is extensively used as powerful tool for decoding the functional relevance of the genes. However, this technique has certain limitations as the phenotypes obtained are not heritable, hence, cannot be used for genetic engineering. Also, VIGS cannot eliminate the involvement

of a gene for a particular function if a phenotype is not apparent and can miss phenotypes that are masked by functional redundancy between gene family members. In addition, the levels of silencing can also vary between plants and experiments depending on the construct and the growth conditions. At the institute, Scientists have applied VIGS system for functional analysis of genes in Tomato leaf curl New Delhi virus (ToLCNDV) susceptible potato cultivar Kufri Pukhraj by silencing three genes, viz. TMV induced protein 1-2 gene, peripheral type benzodiazephine receptor and conserved gene of unknown function. So, VIGS has been proved as a valuable tool in identification of plant genes involved in infection and in resistance to begomoviruses.

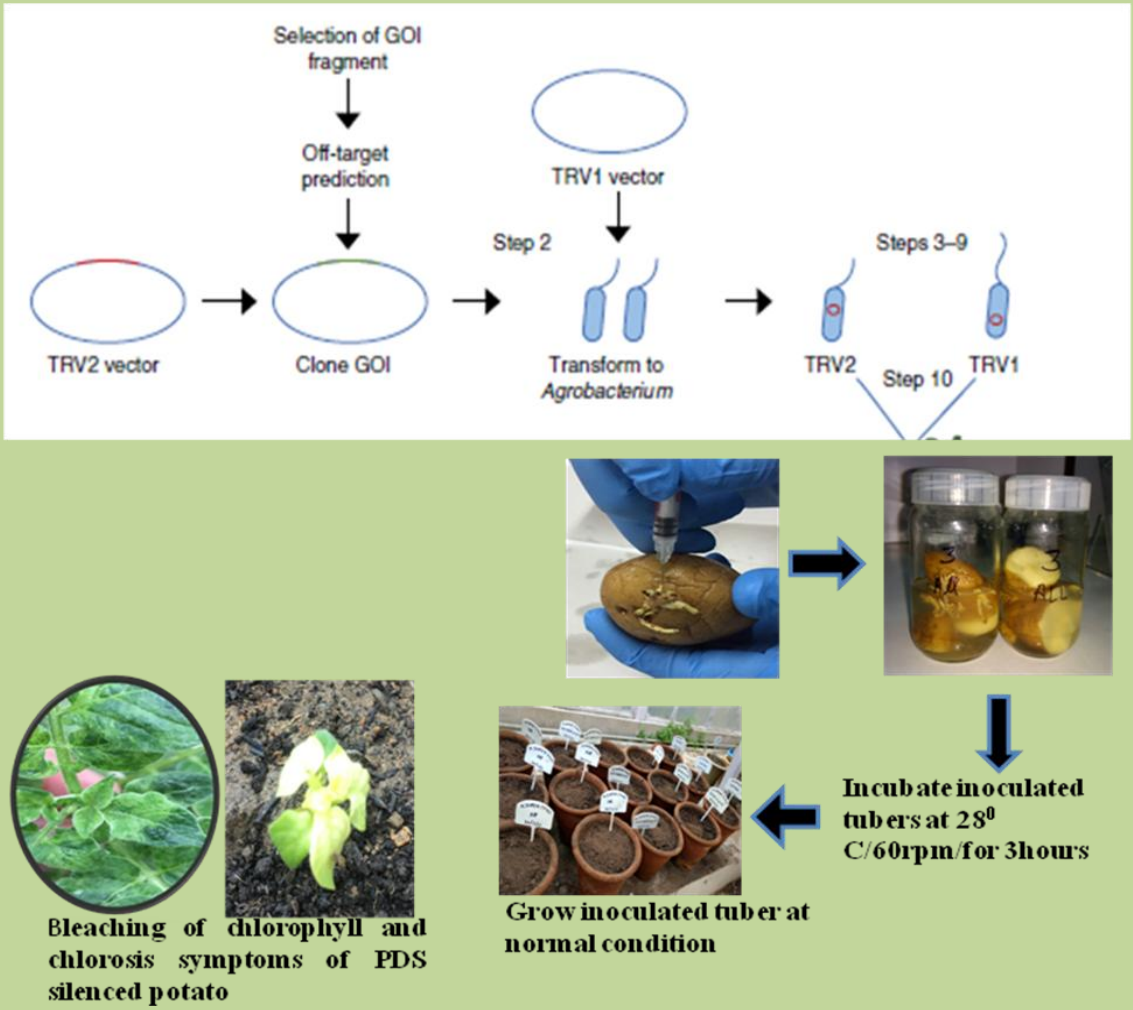


Fig. 19. Overview of VIGS experiment in potato PDS gene used for understanding the functionality of PDS gene in chlorophyll synthesis

Fig. 20. VIGS protocols developed for identification of heat tolerant genes, late blight and bacterial wilt in potato

4. Quantitative Real Time-PCR Technology

The introduction of real-time PCR technology has significantly improved and simplified the quantification of nucleic acids, and this technology has become an invaluable tool for many scientists working in different disciplines. Especially in the field of molecular diagnostics, gene expression and copy number detection, this real-time PCR-based assays have gained importance.

The technologies are based on the measurement of fluorescence during the PCR. The amount of emitted fluorescence is proportional to the amount of PCR product and enables the monitoring of the PCR reaction. The resulting PCR curve is used to define the exponential phase of the reaction, which is a prerequisite for accurate calculation of the initial copy number as well as for expression of the gene at the beginning of the reaction. The simplest and cheapest principle is based on intercalation of double-stranded DNA-binding dyes. This technology can be easily applied to already established PCR assays and does not need any additional fluorescence-labelled oligonucleotide. However, specific and nonspecific PCR

products are both detected. Therefore, these assays require careful optimization of the PCR conditions and a clear differentiation between specific and nonspecific PCR products using melting-curve analysis.

Table 1. Overview of TaqMan- and SYBR-Green based detection

Parameters	TaqMan®-Based Detection	SYBR®-Green Based Detection
Chemistry overview	Uses a fluorogenic probe to enable the detection of a specific PCR product as it accumulates during PCR cycles.	Uses SYBR Green I dye, a highly specific, double-stranded DNA binding dye, to detect PCR product as it accumulates during PCR cycles.
Specificity	Detects specific amplification products only.	Detects all amplified double-stranded DNA, including non-specific reaction products.
Applications	• One-step RT-PCR for RNA quantification • Two-step RT-PCR for RNA quantification • DNA/cDNA quantification -Allelic discrimination -Plus/ Minus assays using an internal positive control	• One-step RT-PCR for RNA quantification • Two-step RT-PCR for RNA quantification • DNA/cDNA quantification
Advantages	Specific hybridization between probe and target is required to generate fluorescent signal.	Enables you to monitor the amplification of any double-stranded DNA sequence.
Disadvantages	A different probe has to be synthesized for each unique target sequence.	Because SYBR Green I dye binds to any double-stranded DNA—including nonspecific double-stranded DNA sequences—it may generate false positive signals.

5. Genome editing

Genome editing is a versatile advance breeding tool which can be employed for crop improvement programme by removing unwanted sequences and make it desirable character of interest by gene knockout, knock in and overexpression studies. Sequence editing either by deleting or modifying the genes individually and then studying the subsequent mutant phenotypes can address the challenges of understanding the function of genes. For precise DNA manipulations, new genome editing systems which are able to induce double-stranded breaks (DSBs) at specific sites in the genome and repaired naturally using DNA repair mechanism (non-homologous end-joining or homologous recombination mechanism) thereby, ensuring the gene mutation at target site. This system is facilitated by protein-guided nucleases, such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), or special RNA/DNA-guided nucleases, including RNA-dependent

DNA cleavage system like clustered regularly interspaced short palindromic repeats (CRISPR) – associated protein (Cas) system.

5.1 CRISPR

CRISPR is an advanced and improved molecular breeding technique in terms of creation of mutagenesis for better improvement programme. CRISPR (clustered regularly interspaced short palindromic repeats)/ Cas (CRISPR-associated) system, has been developed as an easy, specific targeted genome modifications and has emerged as the most powerful method due to ease in designing and construction of gene-specific single guide RNA (sgRNA) vectors. These sgRNA vectors are easily reprogrammable to direct *Streptococcus pyogenes* Cas9 (SpCas9) to generate double stranded breaks (DSBs) in the target genomes that are then repaired by the cell via the error-prone non-homologous end-joining (NHEJ) pathway or by precise homologous recombination (HR) pathway.

Gene editing technologies have been applied for many traits in different crops including potato. CRISPR/Cas9 technology has been successfully implemented in potato (MSU group USA) for targeted editing to generate mutations (by means of NHEJ) as well as gene targeting to edit an endogenous gene (by HR). In this bulletin, we describe procedures for designing sgRNAs, protocols to clone sgRNAs for CRISPR/Cas9 constructs to generate knockouts, design of donor repair templates. Accordingly, CPRI also initiated CRISPR based targeted editing of potato genome for generating variety specific potato seeds and late blight resistance lines.

5.2 Steps for Designing CRISPR/Cas9 cassette for genome editing

- **Define target sequence/gene or promoter:** Selection of 20-bp Target (Gene of interest target region – named as spacer) sgRNA Sequence 5'-NNNNN(20 bases)-NGG-3' (targeting template strand) can be selected using web-based tools such as CRISPR design tools (http://www.rgenome.net/) or CRISPR-P 2.0 (http://crispr.hzau.edu.cn/CRISPR2/) or CCTOP (https://crispr.cos.uni-heidelberg.de/) which have the *Solanum tuberosum* Group Phureja (PGSC v4.03) potato genome sequence.

- **For manual selection, the sgRNA target site:**
  - Must be specific/unique to the gene of interest.
  - Must immediately precede the 5'-NGG PAM.
  - Must be in 50 exonic regions encoding a functional domain so as to disrupt gene function.
  - Must take off-targets into account and be minimized, which can be done by BLAST search and using Cas-off-finder (http://www.rgenome.net/cas-offinder/).
  - For tetraploid potato, allele information for the target gene must be considered. For this purpose, the gene must be cloned and sequenced to determine the allelic composition.
  - To knockout the gene, sgRNA from a conserved region targeting all alleles must be selected.

➢ **The major steps in this protocol include the following:**

- Create vector that contain Cas9 gene function as a scissors and at least two gRNAs based on the flanking sequences of target genomic region of interest for deletion, whoch is driven by U6 promoter.
- Transform the CRISPR construct into potato genome using intermodal stem cuttings as explant either Agrobacterium-mediated or biolistic particle-mediated DNA delivery method to produce transgenic callus lines and potato plants.
- The CAS 9 gene will be integrated randomly and produce protein that will be directed by the single guide RNA to the target site in the genome and cut repair by natural DNA repair mechanism (Non Homologous and Homologous end joining)
- Selecting plants with mutation or replacement or insertion in the target region.
- Screening the clonally propagated plants for Cas9, random locus containing the T-DNA single guide RNA.
- Analyse inheritance of the deletions and select plants with only variants.

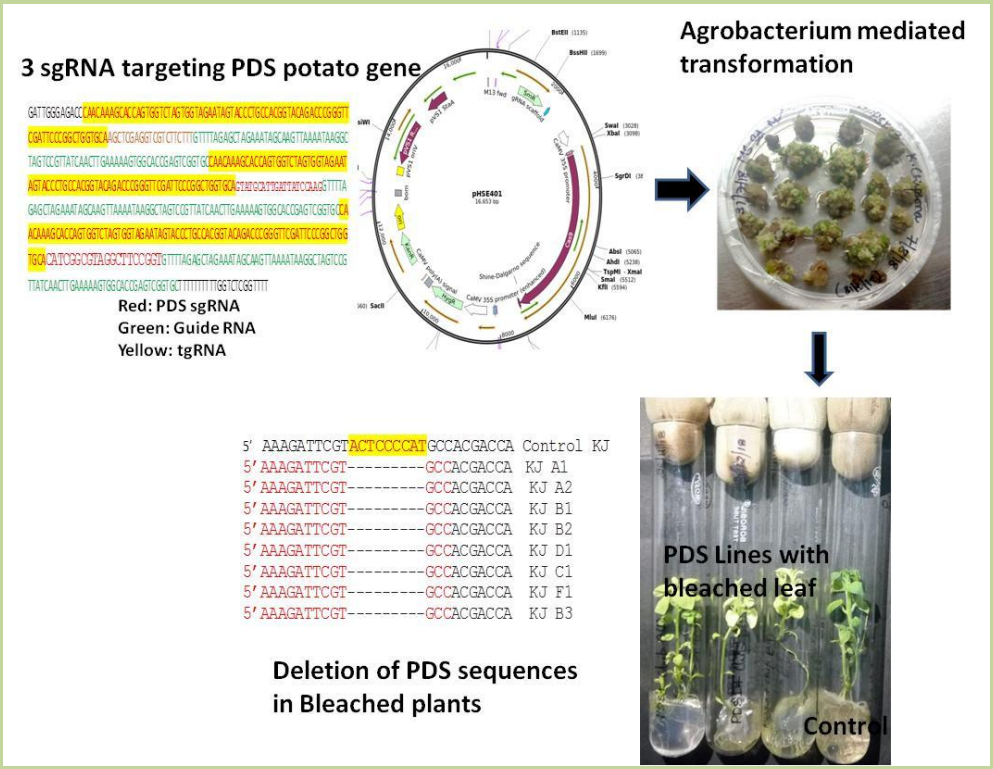


Fig. 21. Overview of gnome editing of PDS gene in potato