

Review Article

<https://doi.org/10.20546/ijcmas.2018.710.308>

Genomic Advances in Root Vegetable Crops

R. Selvakumar^{1*}, Gograj Singh Jat², Sajad Un Nabi¹, Geetika Malik¹,
K. Gangadhara³ and Chintha Pradeepika⁴

¹ICAR-Central Institute of Temperate Horticulture, KD Farm, Old Air Field, Rangreth,
Srinagar-190 007, Jammu and Kashmir, India

²ICAR-Indian Agricultural Research Institute, Pusa Campus, New Delhi-110 012, India

³ICAR-Central Horticultural Experiment Station, Vejalpur-389 340, Gujarat, India

⁴ICAR-Central Tuber Crops Research Institute, Sreekaryam, Trivandrum,
Kerala-695 017, India

*Corresponding author

ABSTRACT

The rapid development in genomic technology and its tools leads to the drastic change in vegetable research. Vegetable breeders can access and explore the enormous sequenced data to deep understanding of vegetable genetic variation at DNA sequenced level. Therefore, well understandings of genomic tools to access and analyze these data are most essential in current vegetable research. The number of available DNA sequenced data for root crops in public databases has constantly increasing recently. These databases offers to design and development of genomic SSR markers, DNA arrays, transposable elements, whole genome sequenced mitochondrial plast genomes and codominant markers and a high throughput SNP genotyping. Some of the vegetable genome databases have been established and continued expanding recently. Simultaneously, analytical methods of genomic bioinformatic tools are also well developed in many aspects of vegetable genomic research particularly functional and comparative genomic analysis, phylogenomics, evolutionary analysis and genome wide association study. However, regular updates of computational infrastructures such as performing analysis software is the most challenging in vegetable genome research. These genomic molecular tools and databases have been successfully used to gain new insights into root genetic diversity, chloroplast genome and mitochondrial genome evolution and to construct the first dense molecular based linkage maps which are opening a new challenging and exciting era for root crop scientific community. This review paper primarily focuses on recent advancement of genomic resources, genomic tools, challenges and opportunities which can forward the vegetable research in current vegetable genomics era as well as future for effective utilization of genetic and genomic resources tools to facilitate the advanced generation sequencing data to enhancement of vegetable productivity and nutritional security.

Keywords

Vegetable,
Transposable,
Genomic,
Phylogenomics,
Genotyping

Article Info

Accepted:
20 September 2018
Available Online:
10 October 2018

Introduction

In the present decade, an increasing amount of DNA sequencing experimental data and molecular marker, breeding techniques have generated vast resources of vegetable genomics, genetics and advanced breeding data. Complete sequenced data sets are become available online and open source, biology data resources have been emerged in vegetable crops as well as sequencing of other vegetable crops are ongoing. The rapid advancement of genomic data resources integrate DNA sequence data, RNA transcripts, genetic maps and phenotypical data associated with specific traits which encourage the breeders to get involved in assembling and annotation works (Wang *et al.*, 2013; Miller *et al.*, 2014 and Sundell *et al.*, 2015). These collective repositories of vegetable crop specific data can pathway to in depth understanding of vegetable genetics, molecular mechanism, taxonomical genera and species classification, vegetable pathology and vegetable pre/post-harvest physiology and are most important tools for lead the research initiatives focused on addressing problems confronted in Olericulture (Bhasi *et al.*, 2010; Mitsui *et al.*, 2015). Modern genome sequencing have forwarded the study of root biology to next level than before. The innovation of high-throughput sequencing methods pathway to vegetable breeders to the ability to explore the structure of genetic resources at molecular level which is known as 'genomics'. Omics research study has explored recently and advanced the main aim in vegetable research due to fast development of second generation sequenced genomes of root crops. It is most successfully impact of vegetable genome research on the improvement of economically important traits and knowledge of vegetable biology (Kim *et al.*, 2009; Tomcal *et al.*, 2013). The availability of open access web resources and constant updates of genomic information

offers to fruitful environment for root vegetable research to be forward. This requires global connection and cooperation among vegetable research community. The genomic database tools has been developed for lead the research to collect and organize genomic information useful for genetics, breeding, taxonomy, plant physiology and other data highly relevant to root vegetables. The current genomic database provides a holistic collection of DNA sequences, RNA transcripts, protein sequences, annotations of genes, splice signals, regulatory elements, transposable elements, and ESTs belonging to root crops. Further, it also has trait annotations, gene functions, molecular taxonomical data, genetic linkage maps, markers, predicted metabolic pathways as well as a huge collection of DNA splicing data (Kitashiba *et al.*, 2014). A variety of comparative and functional genomic analysis and visualization tools are available in the genomic web search tools for the search, download, submit, assemble, annotation and collection of these diverse data sets. In this paper, we review firstly the development of cytological, genetic, marker and genomic technologies and their applications in root vegetable genomic research. Then, we cataloged recent approaches of genomic database tools for assembling and annotation of genes. Most importantly, we summarize most successful genomic resources of root crops. In addition, we also provide basic knowledge of key methods of assembling and analysis of these genomic data such as comparative genomic analysis, functional genomic analysis, phylogenomics, evolutionary analysis and genome wide association study in root vegetable crops.

DNA sequencing in genome research

The successful establishment of DNA sequence of targeted species is to understand the function and structure of gene present in a

genome. In pre genomic era, DNA sequencing has restrictively been preceded with capillarybased, semi-automated applications of the Sanger biochemistry and its variations. Post sequencing of *Arabidopsis* plant in 2001, there in exponential increasing of DNA sequencing due to emergence of Next generation sequencing technology (NGS) in 2005. The recent development of high throughput sequencing technology pathway to explore the structure of genetic material at DNA level which is termed as “genomics”. In this sequencing era, NGS technologies such as pyrosequencing (Roche/454), sequencing-by-synthesis (Illumina/Solexa), sequencing-by-ligation (Illumina/Solexa) (SystemsP), ion semiconductor-based non-optical sequencing (Life Science), single molecule sequencing (Pacific Biosciences) and nanopore sequencing (Oxford Nanopore) have recently developed for ultrahigh-throughput platforms particularly for large sequence needs. These sequencing have been applied to many aspects in exome sequencing and studying genetic transmission of alleles/quantitative trait loci (QTLs) through whole genome sequencing. Recent emerging inexpensive and high throughput sequencing method, genotyping by sequencing (GBS) have been developed for enhancing genomic map construction, especially single nucleotide polymorphisms (SNPs) identification (Varshney *et al.*, 2009). In post genome sequencing era of recent years, several reviews and research on the model and economic crop sequencing project had been targeted. Draft sequenced genome data have been parsed, assembled and annotated through open access database. These databases contain novel genomic resources generated by plant genomic researchers (Table 1).

Genomic project database sources for genome research

The primary step of genomic project is to identify the existing and ongoing projects by

exploring the specific databases related to genomic databases. The targeted vegetable genomic databases can be done by key word search in the internet through NCBI genomes (NCBI, 2015), GOLD (Genomes Online Database) (Reddy *et al.*, 2015), GoGepedia and plaBi (Kitts *et al.*, 2016). The National Center for Biotechnology Information (NCBI) genomedatabase provides DNA sequences, maps, chromosomes, assemblies and annotation through Bioproject, Assembly, Nucleotide and Protein databases (<http://www.ncbi.nlm.nih.gov/genome/>).

Further, Assembly database access to download sequence and annotations for genome assemblies from FTP site (<http://www.ncbi.nlm.nih.gov/assembly/>). The GOLD database consist of nuclear and organelle genome projects, transcriptomic, methylation, exome and re-sequencing projects which offers information of genome and metagenome sequencing projects and their associated metadata (<https://gold.jgi.doe.gov/>). CoGepedia database performs to manage, analyze and visualize the whole generation sequenced data through comparative genomics research (<https://genomevolution.org/>). The plant genomic data base (plaBi) provides updated genetic and sequenced genomic information of plant species in phylogenomics perspective (<http://plabipd.de/>).

Comparative genomic databases for genome research

The comparative genomic analysis gave pathways to study the significant similarities and differences among species after description and definition of coding and non-coding sequences in the genome. Current genomic approaches evolved to new bioinformatic tools to manage and analyze the whole genome sequenced data. There are many comparative genomic databases which may provide extra information on plant genome such as Phytozome (Goodstein *et al.*,

2012), PLAZA (Proost *et al.*, 2015), GreenPhylDB (Rouard *et al.*, 2011), PlantsDB (Spannagl *et al.*, 2016), PlantOrDB (Li *et al.*, 2015), SALAD (Mihara *et al.*, 2010), PlantTribes (Wall *et al.*, 2008), PlantGenIE.org (Sundell *et al.*, 2015), POGs2 (Tomcal *et al.*, 2013), Genomicus Plants (Louis *et al.*, 2015), Piece (Wang *et al.*, 2013), PlantSEED (Seaver *et al.*, 2014) and PGDBj (Asamizu *et al.*, 2014). The world wide Phytozome data (<http://phytozome.jgi.doe.gov/>) provides information of gene, gene structure, gene family, evolutionary history, comparative level of sequences and genome organization which is accessible in world wide. Another famous online platform for published plant genome of functional and structural genomic tool PLAZA (<http://bioinformatics.psb.ugent.be/plaza/>) offers to comprehensive study of whole genome multiple genome alignments in the gene, genome evolution, phylogenetic trees, genomic co-linearity between species and genome evolution. The latest version of comparative and functional genomic open source database GreenPhylDB (<http://www.greenphyl.org/>) designed for relate to each gene, protein domains, orthologous gene prediction covering in the taxonomy of plants. The most common database PlantDB (<http://pgsb.helmholtz-muenchen.de/>) offers to access catalogs and classification of plants, and to visualize and explore the syntenic relationship between model plants to crops. The genome wide database PlantOrDB (<http://bioinfolab.miamioh.edu/plantordb/>) displays information on the evolutionary features of an individual gene and its homolog gene family and it offers to find orthologous genes clusters. Another important plant based proteome databases SALAD (<http://salad.dna.affrc.go.jp/>) and Plant Tribes (<http://fgp.bio.psu.edu/tribedb/>) have been used for construction of dendrograms for protein families and protein expressive pattern by using the information

derived from conserved motifs discovered using the MEME software. Transcriptomics web database PlantGenIE.org (<http://plantgenie.org/>) have been developed for in-depth exploration of standard features of genome browsers, gene list annotation, BLAST tools and gene information pages.

Bioinformatic analysis tools through web sources

Web based bioinformatic tools enables vegetable breeders without programming and systems administration expertise to analyze huge data sets through web computational analysis. Recent advances to make data intensive research of web browser are gave Galaxy (Afgan *et al.*, 2016 and Bornich *et al.*, 2016), Cyverse (Devisetty *et al.*, 2016), Taverna (Oinn *et al.*, 2004 and Wolstencroft *et al.*, 2013), GenePattern (Reich *et al.*, 2006), iobio (Miller *et al.*, 2014), Epiviz (Chelaru *et al.*, 2014), Genome Modeling System (Griffit *et al.*, 2015) and ZENBU (Severin *et al.*, 2014) cyber infrastructure platforms which have been developed for more accessible, transparent and reproducible computational genome analysis. The Galaxy web consists of Public Galaxy Server (<https://usegalaxy.org>), Galaxy software framework (<https://getgalaxy.org>), and Galaxy Tool Shed (<https://usegalaxy.org/toolshed>) which are offers to access data storage, dissemination and high throughput bioinformatics analysis and visualization in a Web browser (<https://usegalaxy.org/u/jeremy/h/hpac-exome-analysis>). Another online genomics platform, CoGe (<http://synteny.cnr.berkeley.edu/CoGe>) provides an integrative approaches for genomic datasets from any organism of gene interest in a number of formats (FASTA sequence, GenBank accession, LOC-number, gene name), find homologous features within the same genome or in different genomes, and compare multiple genomic regions by using CoGe's Genome Evolution tool (GEvo). CoGe

web browser integrates genomic databases from any organism, DNA alignment and assessment tools, and interactive graphic modules to enable researchers to compare any genomic DNA sequence, feature, or position with any other in any organism and further, it allows for comparison, visualization, and subsequent manipulation of multiple chromosomal DNA sequences from multiple organisms. Another recent high performance web tool iPlant (<https://de.iplantcollaborative.org/>) provides huge data storage, and DNA subway that predict and annotate genes, prospect entire plant genomes for related genes and sequences, determine sequence relationships and analyze RNA-Seq reads to measure differential expression

Carrot genome characteristics

Carrot (*Daucus carota* L.) is a cool weather crop grown in temperate and subtropical regions for its edible storage tap roots both for fresh as well as processed vegetable throughout the world and is most important of all the root crops (Iorizzo *et al.*, 2011). The estimated genome size of cultivated carrot ($2n=2x=18$) is 473 Mb with nine pairs of chromosomes (Arumuganathan and Earle, 1991) and 1.0 pg of DNA per 1C nucleus (Simon *et al.*, 2009). It is noticeably smaller genome size than other vegetable crops such as potato (844 Mb), tomato (900 Mb) and lettuce (2.7Gb). The cytological studies established that chromosome structure that heterochromatic domains are mainly confined to the pericentromeric regions of each chromosome. Four pairs of chromosomes are metacentric and five pairs are submetacentric. Based on a DNA association curves and thermal denaturation, the carrot genome consists of approximately 40% repetitive sequences, and the GC content is 37-38% (Iovene *et al.*, 2011). This basic information provides a foundation to initiate sequencing the whole genome.

Carrot bioinformatic and genomic tools

Carrot DB (Carrot Data Base) was developed based on Web-based genomic and transcriptomic database of *Daucus carota* sp. *sativus* (<http://apiaceae.njau.edu.cn/carrotdb>). It provides tools of Genome Map and Basic Local Alignment Search Tool (BLAST) for find and searching sequences of the scaffolds, target genes, simple sequence repeats (SSRs) markers, assembled transcriptomic sequences, expressed gene in the transcriptome, de novo assembled whole-genome sequences, putative gene sequences or gene fragments and putative protein sequences of carrot (DC-27) as well as checking of gene annotation, and submit information to the genome map group (Xu *et al.*, 2014). Further, germplasm part of CarrotDB are used for 45 carrot genotypes photos, accession numbers, names, countries of origin and colors of cortex, phloem and xylem regions of respective carrot genotypes. De novo assembling and analysis of DC-27 were done by whole genome sequencing. Sequence Read Archive (SRA) database were successfully used for downloading of transcriptomic sequences of DC-27 from National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/sra/>). The fourteen assembled transcriptomic sequences of carrot genotypes were mapped the whole genome sequence along with fragments per kilobase of transcript sequence per millions base pairs sequenced information (FPKM). The 2826 transcription factor (TF) genes classified into 57 families and identified in the whole genome sequences which are embedded into Carrot DB database as interface (Xu *et al.*, 2014). The RoBuST database offers effortlessly explore and analyze sequence annotations of 3663 genes 5959 RNAs, 22,723 ESTs and 11,438 regulatory sequence elements from Apiaceae and Alliaceae plant families (<http://robust.genome.com>), to

visualize and analysis of genomic sequence data by graphical tools and to search and download the traits, biosynthetic pathways, genetic linkage maps and molecular taxonomy data of root and bulb vegetable (RBV) crops. Moreover, RoBuST database helps for comparative genomic analysis of plant splicing pattern from 659,369 splice signals of 6015 plant species (Bhasi *et al.*, 2010).

Radish genome characteristics

Radish (*Raphanus sativus* L.) is a diploid dicot and an economically important annual or biennial root vegetable crop of the Brassicaceae family and it's grown and consumed all over the world. The genome size of radish ($2n=2x=18$) is 383 Mb which is consisting of 54,357 genes in the genome (Mitsui *et al.*, 2015). This genomic information provides extensive chromosome homology among *Raphanus* species, which would facilitate transfer of the genomic future molecular breeding programmes.

Radish genomic resources

De novo transcriptome sequencing and assembly study were successfully provided 53,642 unigenes in leaf transcriptome and 50,385 annotated unique sequences by Gene ontology (GO), Clusters of Orthologous Groups (COG) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Its further insights into plant circadian rhythmic pathway, bolting and flowering time regulatory networks such as total 24 candidate gene and 142 potential radish genes respectively by BLAST searching against the public protein databases. This RNA-seq technology further useful for dissecting the molecular mechanisms in Brassicaceae vegetable crops (Nie *et al.*, 2016). Draft genome of a Chinese contains 43240 protein-coding genes were functionally annotated in the *R. sativus* genome by using a combination

of *ab initio* gene prediction, protein-based homology searches, EST alignments and transcriptome sequencing data. Of these, 387.73 Mb of assembled scaffolds, 83.93% of the scaffolds were anchored onto nine pseudochromosomes. By comparative genome analysis of the radish genome against 10 other plant genomes, 2 275 genes in 780 gene families were found unique to *R. sativus*. This draft genome assemblies are most useful for traits evolution and genetic improvement of radish (Zhang *et al.*, 2015). As for organelle genomes of radish, the full-length of plastid and mitochondrial sequences have been reported for origin and phylogenetic relationships of radish (Yamane *et al.*, 2009; Kim *et al.*, 2009; Jeong *et al.*, 2014). The chloroplast genome of radish cv.WK10039 is 153,368 bp in length and contains a pair of IRs, each of which is 26, 217 bp in length. The *cp* genome contains 87 protein-coding genes, 8 rRNA genes, and 37 tRNA genes (Jeong *et al.*, 2014). Comparative genome of radish chloroplast genomes opening a new field of research on plant organelle genome evolution and demonstrating the utility of well-characterized cp genomes (Jeong *et al.*, 2014).

Radish transcriptome

Based on GO annotations, 40,705 of 65,457 *Raphanus* gene models were assigned with GO terms for tuberous root formation and development. In total, 142 KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathways including 13,795 genes were found in the *Raphanus* gene models. Annotation and analysis of the draft genome sequence of the radish (*Raphanus sativus* var. *hortensis*) showed the total size of 383 Mb (N50 scaffold: 138.17 kb) with 54,357 genes which indicated diversification between *Raphanus* and *Brassica* coincide with the time of whole genome triplication by syntenic and phylogenetic analyses. Transcriptome profiling analysis revealed that the gene

functions and pathways of sucrose metabolism were prominently activated in thickening roots, notably in cell proliferating tissues. It shows higher expression of sucrose transporter genes (SUTs and SUCs) which are involving in cell-to-cell and long-distance distribution of sucrose throughout the radish biological system during early seedling roots and leaves. Moreover, the higher expression levels of two homologous genes of sucrose synthase 1 (*SUS1*) were correlated with root thickening rates. The pungent determinant of glucosinolate (GLS) and myrosinase genes and its transcriptional regulating genes were profiled by transcriptome analysis in strongly expressed root tip and cortex region. The radish transcriptome contains majorly GLS biosynthesis-related genes (GLS genes) whereas no orthologue in the genome was observed for 10 GLS genes encoding methionine side-chain elongation genes (MAM1, MAM3, IPMI-SSU3 and IPMIDH3), core structure formation genes (CYP79F2), side-chain modification genes (CYP81F1, FMOGS-OX3, FMOGS-OX4, AOP2 and AOP3) and a transcription factor gene (MYB76).

Based on transcriptional analysis, expression of MYB28 and MYB29-like gene showed higher in pungent region of root correlated with GLS gene expression patterns which act as a key regulator of glucosinolate biosynthesis. Furthermore, transcriptome analysis revealed that eleven myrosinase encoding genes (myrosinase MA, MB and MC subfamilies) were distinctly large in the in *Raphanus* and *B. rapa* than Brassicae and these genes are highly expressed in pungent regions of peel, xylem and root tip and middle developmental stages, and were correlated with the presence of isothiocyanates. Additional RT-qPCR experiments of four genes, MYB28, BCAT4, CYP79F1 and

TGG1-3 (TGG1C), showed that these pungency-related genes were highly expressed in pungent tissues (Moghe *et al.*, 2014 and Nakatsuji *et al.*, 2011)

Radish mitogenome

Radish mitochondrial genome is the special target for development of cytoplasmic male sterility (CMS) based F₁ hybrids. Radish mitochondrial genome cv. WK10039 was sequenced by using NGS technologies, and the reads were assembled into a circular DNA (Moghe *et al.*, 2014). The total length of assembled radish mitochondrial genome is 244,054 bp consist of 82 protein-coding genes in which 48 ORF genes are encoded by hypothetical proteins and remaining 34 genes are encoded by know proteins (Nakatsuji *et al.*, 2011). The overall GC content is 45.3%, and was slightly lower in exons (44.2%).

The size of exons was 482 bp (105 of exons) with 20.7% of total mtDNA. Nine genes (*rps3*, *cox2*, *ccmFC*, *nad1*, *nad5*, *nad2*, *rpL2*, *nad4*, and *nad7*) contain more than one intron. Furthermore, the radish mitochondrial genome contains 17 tRNA genes coding for 14 amino acids, and rRNA genes including 5S rRNA, 18S rRNA, and 26S rRNA genes (Moghe *et al.*, 2014 and Nakatsuji *et al.*, 2011). Two 5529 bp repeats were located at 99.1 kb and 238.5 kb, thereby creating tripartite structure of the mtDNA. Sequence comparisons of the 5529 bp repeats with previously reported radish mtDNAs indicate that the repeats are completely identical to those found in Uchikigensuke mtDNA (AB694743) and are very similar to DCGMS (KC193578) and MS-gensuke (AB694744). Recombination between repeats is expected to generate at least two forms of subgenomes of the mtDNA, which results in 104.7 kb and 144.9 kb recombinant forms (Kitashiba *et al.*, 2014).

Table.1 Genomic tools for root vegetable crops

Database	Database link	References
Radish	http://www.nodai-genome-d.org/	Mitsui <i>et al.</i> , 2015
RadishBase	http://bioinfo.bti.cornell.edu/radish	Shen <i>et al.</i> , 2013
RGD	http://www.radish-genome.org/	Kitashiba <i>et al.</i> , 2014
<i>Raphanus sativus</i> genome data base	http://radish.kazusa.or.jp/	Shirasawa <i>et al.</i> , 2011
BRAD	http://brassicadb.org	Cheng <i>et al.</i> , 2011
CarrotDB	http://apiaceae.njau.edu.cn/carrotdb	Xu <i>et al.</i> , 2014
Carrot genome	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Dcarota	Iorizzo <i>et al.</i> , 2016
RoBuST	http://robust.genome.com	Bhasi <i>et al.</i> , 2010

Radish bioinformatic and genomic tools

RadishBase data base uses assembled and annotated radish mitochondrial genome sequences, expressed sequence tag (EST) transcripts and annotation, biochemical pathways from unigene sequences, predicted metabolic pathways, EST-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers, and genetic maps.

It is designed to store, query, analyze, integrate, retrieve and visualize the genetically important information of different radish resources through BLAST search and unigene annotation query interfaces, and tools to classify unigenes functionally, to identify enriched gene ontology (GO) terms and to visualize genetic maps (<http://bioinfo.bti.cornell.edu/radish>). Further, its generated genomics and genetics data sets including the genome sequences and RNA-seq data sets and the associated data mining and analysis tools (Shen *et al.*, 2013).

Turnip bioinformatic and genomic tools

The BRAD (Brassica database) had been developed to start a platform for collecting and organizing genomic information useful for Brassica researchers (Cheng *et al.*, 2011).

It contains the first whole-genome sequences of 'Chiifu-401' *B. rapa* which were de novo assembled and analyzed using second generation sequencing technologies such as Illumina GA II short reads and from BAC clone sequences. BRAD database (<http://brassicadb.org>) provides complete genome sequence of *B. rapa*, predicted genes, associated annotations (InterPro, KEGG2, SwissProt), non-coding RNAs, transposable elements, *B. rapa* genes orthologous to those in *Arabidopsis thaliana*, as well as genetic markers and linkage maps of *B. rapa* (Cheng *et al.*, 2011).

The current release of BRAD date base offers useful searching and data mining tools, including search across *Brassica* genome annotation datasets, search for syntenic or nonsyntenic orthologs, and to search the flanking regions of a certain target, as well as the tools of BLAST and Gbrowse. Further, it allows to submit and browse any kind of genetic information, such as gene families, genes, families of NBS genes, auxin genes, transcription factor of *B. rapa* or *A. thaliana* gene ID, haplotypes (derived from SNPs mapping) of the *B. rapa* germplasm collection, physical position or genetic markers, synteny browser of *B. rapa* to *B. oleracea*, levels of gene expression generated from transcriptome data in different organs of

B. rapa, synteny browser of *B. rapa* to *B. oleracea* and allele data and frequencies of genetic markers generated from genome resequences of different lines of *B. rapa* (Cheng *et al.*, 2011).

Genome wide association studies in root vegetable crops

Genome Wide Association studies is the most powerful application to study the association between plant traits and genetic variation and to explore the allelic variation in phenotypic diversity and higher resolution of QTL mapping for complex traits. Furthermore, it's used for to accelerate genetic mapping and gene discovery in plant after consequent high throughput phenotypic facility and DNA sequence variation from NGS technology. Two bioinformatic tools such as Heap and GnpIS-Asso have been developed for GWAS acceleration in which first one is used for detection SNPs whereas second one is used for get associations values between traits and markers obtained in association studies (Kitashiba *et al.*, 2014).

Advancement of bioinformatic tools beyond root genomic research

Right now, we are moving towards the post genomic era since whole genome sequencing technology continues reaching advanced innovation in sequencing level and high throughput. Past few decades many omics technologies have been emerged due to development of modern technology for exploring the knowledge of molecular system. The most recent added omics layers include interactomics, epigenomics, hormonomics, and metabolomics. Recent advancement of next generation sequencing provides feature for whole-genome sequencing/re-sequencing for various genomic analysis such as RNA sequencing is established for transcriptome and non-coding RNAome analysis,

quantitative detection of epigenomic dynamics and Chip-seq analysis for DNA–protein interactions. In addition, approaches in transcriptional regulatory networks research based on omics data have been published such as interactome analysis for networks formed by protein–protein interactions, transcriptome and metabolome analysis for metabolic systems (Lu *et al.*, 2008). Apparently, these rapidly growing omics databases widen the large-scale of genomic resources. Therefore, genomic tools has become more essential than ever for every aspect of omic-based research to be well managed and effectively analyze data (Varshney *et al.*, 2009).

In this review, we have attempted to catalog of different genomic data resources of root vegetable crops. With the increasing number of repositories of root vegetable crops, it is evident that assembled and annotated genomic sequenced data available on the web, associated with almost every aspects of genetics and genomic research. It can be successfully used for advance our understanding of horticultural traits, increase genetic gain by genetical selection and reduce the breeding cycles by incorporating genome wide selection strategies for breeding of root vegetable and its wild relatives.

This primary initiative would be change the understanding and nature of research in vegetable crops. The genomic data have been efficiently explored in the all possible genetic resources to search and analyze the date and to focus better visibility of extensive genetic screens for whole genome wide association analysis, functional and comparative genomics.

References

Arumuganathan K and Earle ED. 1991. Nuclear DNA content of some important plant

- species. *Plant Mol. Biol. Rep.*, 9: 208–218.
- Asamizu E, Ichihara H, Nakaya A, Nakamura Y, Hirakawa H, Ishii T, Tamura T, Fukami-Kobayashi K, Nakajima Y and Tabata S. 2014. Plant Genome Data Base Japan (PGDBj): a portal website for the integration of plant genome-related databases. *Plant Cell Physiol.* 55(1): e8.
- Bhasi A, Doug Senalik, Philipp W Simon, Brajendra Kumar, Vinu Manikandan, Philge Philip, Periannan Senapathy. 2010. RoBuST: an integrated genomics resource for the root and bulb crop families Apiaceae and Alliaceae. *BMC Plant Biology*, 10: 161
- Chelaru F, Smith L, Goldstein N and Bravo HC. 2014. Epiviz: interactive visual analytics for functional genomics data. *Nat. Methods* 11: 938–940.
- Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. 2011. BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biol.*, 11: 136.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N and Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40: D1178-86.
- Griffith M, Griffith OL, Smith SM, Ramu A, Callaway MB, Brummett AM, Kiwala MJ, Coffman AC, Regier AA and Oberkfell BJ. 2015. Genome Modeling System: A Knowledge Management Platform for Genomics. *PLoS Comput. Biol.*, 11: e1004274.
- Iorizzo M, Senalik DA, Grzebelus D, Bowman M, Cavagnaro PF, Matvienko M, Ashrafi H, Van Deynze A and Simon PW. 2011. De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genomics*, 12: 389.
- Iovene M, Cavagnaro PF, Senalik D, Buell CR, Jiang J and Simon PW. 2011 Comparative FISH mapping of *Daucus* species (Apiaceae family). *Chromosome Res*, 19: 493-506.
- Jeong Y, Won-Hyung Chung, Jeong-Hwan Munc, NamshinKim, Hee-Ju Yu. 2014. De novo assembly and characterization of the complete chloroplast genome of radish (*Raphanus sativus* L.). *Gene* 551 (4): 39–48.
- Kim S, Lee YP, Lim H, Ahn Y, Sung SK. 2009. Identification of highly variable chloroplast sequences and development of cpDNA-based molecular markers that distinguish four cytoplasm types in radish (*Raphanus sativus* L.). *Theor. Appl. Genet.* 119: 189–198.
- Kitashiba H, Li F, Hirakawa H, Kawanabe T, Zou Z, Hasegawa Y, Tonosaki K, Shirasawa S, Fukushima A, Yokoi S, Takahata Y, Kakizaki T, Ishida M, Okamoto S, Sakamoto K, Shirasawa K, Tabata S and Nishio T. 2014. Draft sequences of the radish (*Raphanus sativus* L.) genome. *DNA Res*, 21: 481–490.
- Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem, V, Sapojnikov V, Smith RG, Tatusova T, Xiang C, Zherikov A, DiCuccio M, Murphy TD, Pruitt KD and Kimchi A. 2016. Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.*, 44: D73-80.
- Li L, Ji G, Ye C, Shu C, Zhang J and Liang C. 2015. PlantOrDB: a genome-wide ortholog database for land plants and green algae. *BMC Plant Biol.* 15: 161.
- Louis A, Murat F, Salse J and Crollius HR. 2015. Genomicus Plants: a web resource to study genome evolution in flowering plants. *Plant Cell Physiol.* 56(1): e4.
- Lü N, Yamane K and Ohnishi O. 2008. Genetic diversity of cultivated and wild radish and phylogenetic relationships among *Raphanus* and Brassica species revealed by the analysis of trnK/matK sequence. *Breeding Sci*, 58: 15–22.
- Mihara M, Itoh T and Izawa T. 2010. SALAD database: a motif-based database of protein annotations for plant comparative genomics. *Nucleic Acids Res.* 38: D835-42. 7.

- Miller CA, Qiao Y, Di Sera T, Dastous B and Marth GT. 2014. Bam. Iobio: a web-based, real-time, sequence alignment file inspector. *Nat. Methods*. 11: 1189–1189.
- Mitsui Y, Shimomura M, Komatsu K, Namiki N, Shibata-Hatta M, Imai M, Katayose Y, Mukai Y, Kanamori H, Kurita K, Kagami T, Wakatsuki A, Ohyanagi H, Ikawa H, Minaka N, Nakagawa K, Shiwa Y and Sasaki T. 2015. The radish genome and comprehensive gene expression profile of tuberous root formation and development. *Sci Rep.*, 5: 10835.
- Moghe GD, Hufnagel DE, Tang H, Xiao Y, Dworkin I, Town CD, Conner JK and Shiu SH. 2014. Consequences of whole-genome triplication as revealed by comparative genomic analyses of the wild radish *Raphanus raphanistrum* and three other Brassicaceae species. *Plant Cell*, 26: 1925–1937.
- Nakatsuji R, Hashida T, Matsumoto N, Tsuru M, Kubo N and Hirai M. 2011. Development of genomic and EST-SSR markers in radish (*Raphanus sativus* L.). *Breeding Sci*, 61: 413–419.
- NCBI Resource Coordinators. 2015. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 43: D6-17.
- Nie S, Chao Li, Liang Xu, Yan Wang, Danqiong Huang, Everlyne M. Mulekel, Xiaochuan Sun, Yang Xie and Liwang Liu. 2016. De novo transcriptome analysis in radish (*Raphanus sativus* L.) and identification of critical genes involved in bolting and flowering. *BMC Genomics* 17: 389
- Oinn T, Addis M, Ferris J, Marvin D, Senger M, Greenwood M, Carver T, Glover K, Pocock, MR, Wipat A. 2004. Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics*, 20: 3045–3054.
- Proost S, Van Bel M, Vanechoutte D, Van de Peer Y, Inzé D, Mueller-Roeber B and Vandepoele K. 2015. PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* 43: D974-81.
- Reddy TB; Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA and Kyrpidis NC. 2015. The Genomes on Line Database (GOLD) v.5: a metadata management system based on a four level (meta) genome project classification. *Nucleic Acids Res.*, 43: D1099-106.
- Reich M, Liefeld T, Gould J, Lerner J, Tamayo P and Mesirov JP. 2006. GenePattern 2.0. *Nature Genet*, 38, 500–501.
- Rouard M, Guignon V, Aluome C, Laporte MA, Droc G, Walde C, Zmasek CM, Périn C and Conte MG. 2011. GreenPhylDB v2.0: comparative and functional genomics in plants. *Nucleic Acids Res.* 39: D1095-102.
- Seaver SM, Gerdes S, Frelin O, Lerma-Ortiz C, Bradbury LM, Zallot R, Hasnain G, Niehaus, TD, El Yacoubi B, Pasternak S, Olson R, Pusch G, Overbeek R, Stevens R, de Crécy-Lagard V, Ware D, Hanson AD and Henry CS. 2014. Highthroughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. *Proc. Natl. Acad. Sci.* 111(26): 9645-50.
- Severin J, Lizio M, Harshbarger J, Kawaji H, Daub CO, Hayashizaki Y, Bertin N and Forrest ARR. 2014. Interactive visualization and analysis of large-scale sequencing datasets using ZENBU. *Nat. Biotechnol.*, 32: 217–219.
- Shen D, Honghe Sun, Mingyun Huang, Yi Zheng, Xixiang Li and Zhangjun Fei. RadishBase: A Database for Genomics and Genetics of Radish. *Plant Cell Physiol.* 54(2): e3(1–6) (2013)
- Shirasawa K, Oyama M, Hirakawa H, Sato S, Tabata S, Fujioka T, Kimizuka-Takagi C, Sasamoto S, Watanabe A, Kato M, Kishida Y, Kohara M, Takahashi C, Tsuruoka H, Wada T, Sakai T, Isobe S. 2011. An EST-SSR linkage map of *Raphanus sativus* and comparative genomics of the *Brassicaceae*. *DNA Res.* 18: 221-232.

- Simon P, Pollak L, Clevidence B, Holden J and Haytowitz B. 2009. Plant breeding for human nutritional quality. *Plant Breed. Rev.*, 31: 325-392.
- Spannagl M, Nussbaumer T, Bader KC, Martis MM, Seidel M, Kugler KG, Gundlach H and Mayer KF. 2016. PGSB PlantsDB: updates to the database framework for comparative plant genome research. *Nucleic Acids Res.* 44: D1141-7.
- Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin YC, Sjödin A, Van dePeer Y, Jansson S, Hvidsten TR and Street NR. 2015. The Plant Genome Integrative Explorer Resource: PlantGenIE.org. *New Phytol.*, 208(4): 1149-56.
- Tomcal M, Stiffler N and Barkan A. 2013. POGs2: a web portal to facilitate cross-species inferences about protein architecture and function in plants. *PLoS One* 8(12): e82569.
- Varshney RK, Nayak SN, May GD and Jackson SA. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* 27: 522–530.
- Wall PK, Leebens-Mack J, Müller KF, Field D, Altman NS and dePamphilis CW. 2008. PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic Acids Res.* 36: D970-6.
- Wang Y, You FM, Lazo GR, Luo MC, Thilmony R, Gordon S, Kianian SF, Gu YQ. 2013. PIECE: a database for plant gene structure comparison and evolution. *Nucleic Acids Res.* 41: D1159-66.
- Wolstencroft K, Haines R, Fellows D, Williams A, Withers D, Owen S, Soiland-Reyes S, Dunlop I, Nenadic A and Fisher P. 2013. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Res.*, 41: W557–W561.
- Xu Z, Hua-Wei Tan, FengWang, Xi-Lin Hou, Ai-Sheng Xiong. 2014. CarrotDB: a genomic and transcriptomic database for carrot. *Database*, 2014, 1–8.
- Yamane K, Lü N and Ohnishi O. 2009. Multiple origins and high genetic diversity of cultivated radish inferred from polymorphism in chloroplast simple sequence repeats. *Breeding Sci.* 59: 55–65.
- Zhang X, Yue Z, Mei S, Qiu Y, Yang X, Chen X, Cheng F, Wu Z, Sun Y, Jing Y, Liu B, Shen D, Wang H, Cui N, Duan Y, Wu J, Wang J, Gan C, Wang J, Wang X and Li Xixianga, 2015. A de novo Genome of a Chinese Radish Cultivar. *Horticultural Plant Journal*, 1 (3): 155–164.

How to cite this article:

Selvakumar, R., Gograj Singh Jat, Sajad Un Nabi, Geetika Malik, K. Gangadhara and Chinthia Pradeepika. 2018. Genomic Advances in Root Vegetable Crops. *Int.J.Curr.Microbiol.App.Sci.* 7(10): 2649-2660. doi: <https://doi.org/10.20546/ijcmas.2018.710.308>