ABSTRACT

The 2020 world population data sheet estimates that 30 years from now, the population is intended to outstretch approximately up to 9.9 billion. The exponential rise in population has led to the agitation of land and the environment. Abrupt changes in climatic conditions and increasing global population emphasize the propagation of novel crops and acclimatization of the available plants and crops to obtain sufficient food in the extended future. In addition to ‘Oomics’, bioinformatics has key role in understanding the underlying mechanism of molecular functional systems in various plants. Bioinformatics contributes towards multi-disciplinary interactions and helps to reshape agricultural tradition and production, providing knowledge for enhanced plant quality, and it also provides a plan of action for protection against adverse environmental conditions. This review examines recent approaches in systems biology for predicting the functionality of genes and their networks (GRNs) along with its influential effect in current research involving various disciplines which will be very useful to researchers in plant sciences.

Keywords: computational biology; bioinformatics; genome sequencing; proteomics; transcriptomics; metabolomics
INTRODUCTION

Almost 80% of the food obtained, and plants make up 98% of the oxygen consumed, thus being the core basis of life (Mann et al., 2002). With the increase in population, it is of utmost importance to obtain expeditious and imperishable development of plants and crops (Tillman et al., 2011). But there are various complications like unfavorable climatic conditions, compact land spaces, water scarcity, elevated carbon dioxide concentration etc., that lead to declination in the yield of plants (Godfray et al., 2010). Considering the changes and their repercussions, we must be prepared for good quality crops to secure food productivity later.

Analytical understanding of these developmental features contributes to an increase in plant flexibility and durable agriculture, ultimately providing food security (Hu et al., 2018). The functions of different genes in plants and metagenomes affect plant growth; thus, an appreciating step has already been taken to improve crop quality and metagenomes explore the functionality of different genes. The advanced high throughput next-generation sequencing technology (NGS) enables sequencing of plant genomes, thus providing assemblage and analysis of data at the systems level, further developing a gene regulatory network (GRN).

Bioinformatics has been a great aid to reveal the data analysis generated by these technologies. Previously, researchers referred to textbooks and scientific journals and articles to proceed with their work, but the development of technologies has made it easier. Scientists rely on the web and databases to obtain data and information on different aspects of their investigations. Furthermore, that provide opportunities to store the data systematically and perform comparative analysis respectively (Pop et al., 2004). Thus, the convergence of omics and bioinformatics may help in crop-quality improvement, which may also lead to the introduction of new varieties.

Some of the databases provide onsite analytical tools to help nascent and skilled professionals. The three most widely used sequence-related databases are Genbank NCBI USA (Benson et al. 2006), ENA in Europe (Cochrane et al., 2006) and DDBJ (Sugawara et al., 2008). They all are collaboratively interconnected by the INSD with continuous upgradation. There are also various databases specifically targeting the Plant species, comprising all the sequenced crops. This review aims to summarise recent bioinformatics applications tools with currently available databases in the study of various interdisciplinary “Oomics” (genomics, proteomics, transcriptomics, phenomics, interactomics and others) as shown in Figure 1.

GENOME SEQUENCING AND RNA SEQUENCING

The genome sequences of varieties are obtained using high throughput sequencing technologies (Ni et al., 2016; Redekar et al., 2017). NGS sequencing has exponentially extended in plant genomics. Databases like NCBI and DDBJ store raw sequencing data obtained by NGS sequencing for further use. Researchers collect the gene expression data with specifications like temporal, spatial, perturbation, or a combination, considering the area of interest in plant science.

Spatial specific genome data are obtained from different cells, tissues or organs (Liseron-Monfils et al., 2018). Temporal data are collected to study the gene expression during the different plant growth and developmental stages (Gupta et al., 2018; Van den Nishiyama et al., 2018). Perturbation data emphasizes the change in expression due to environmental change or mutation (Pal et al., 2017; Jiang et al., 2018).
Figure 1. Inter relational application of Bioinformatics with Functional genomics in recent trends.

Impact of NGS sequencing in plant science research

The sequencing of plant genomes has been found beneficial in generating the width of genes functions involved in attaining agronomic or high yield traits (Edwards et al., 2013). Although different sequencing techniques were used to sequence the genomes of *Arabidopsis* and rice, the NGS (Next Genome Sequencing) technology has been adopted since around 2005 (Goodwin et al., 2016), (Table 1). NGS technology produces an impactful effect in the advancement of plant research. To explore the functional effect of different genes, the whole genome sequencing of various crops is performed, thus accommodating the agricultural practice (Weigel et al., 2009). Knowing the functionality of all genes in plants helps to further modify and interrogate genomes, thus producing high yield quality plants. Plant yield and quality is also determined by plant and soil microbiome, thus playing an important role in agriculture (Babu et al., 2015; Deusch et al., 2015). To decipher the role of metagenomes in plant nutrition, the NGS sequencing of environmental samples is also done (Lavecchia et al., 2015; Pii Y et al., 2016).

Table 1. NGS technologies used in genome sequencing

<table>
<thead>
<tr>
<th>Technology</th>
<th>Read length</th>
<th>Reads per run</th>
<th>Run Time</th>
<th>Link</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roche454 pyrosequencing</td>
<td>700</td>
<td>~700 thousand</td>
<td>23 hr</td>
<td>(<a href="http://www.my454.com">http://www.my454.com</a>)</td>
<td>Lai et al., 2012</td>
</tr>
<tr>
<td>Illumina</td>
<td>300</td>
<td>~300 billion</td>
<td>8.5 d</td>
<td>(<a href="http://www.illumina.com">http://www.illumina.com</a>)</td>
<td>Buggs et al., 2012</td>
</tr>
<tr>
<td>Ion Torrent</td>
<td>200</td>
<td>~60 billion</td>
<td>2h</td>
<td>(<a href="http://www.iontorrent.com">http://www.iontorrent.com</a>)</td>
<td>Howden et al., 2011</td>
</tr>
<tr>
<td>SOLiD</td>
<td>100</td>
<td>~200</td>
<td>7d</td>
<td>(<a href="http://www.appliedbi">http://www.appliedbi</a>)</td>
<td>Ashelford et al., 2011</td>
</tr>
</tbody>
</table>
**GENOME ASSEMBLY**

The short reads generated by NGS technologies are aligned and merged to construct a whole-genome sequence (Schatz et al., 2010). Several computer programs (Table 2) are available that perform genome assembly. These assemblers use the single or paired reads data to assemble the genome (Baker, 2012). OLC and de-bruijn-graph (DBG) algorithms are used for genome assembly (Li et al., 2012). CAP3, Newbler, ALLPATHS-LG, Velvet, AbySS, SOAPdenovo2 are few assemblers utilized for plant genome assembly.

**Table 2. Assemblers based on OLC and DBG algorithms for plant genome assembly**

<table>
<thead>
<tr>
<th>Assemblers</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAP3</td>
<td>OLC</td>
</tr>
<tr>
<td>Newbler</td>
<td>OLC</td>
</tr>
<tr>
<td>ALLPATHS-LG</td>
<td>DBG</td>
</tr>
<tr>
<td>Velvet</td>
<td>DBG</td>
</tr>
<tr>
<td>ABySS</td>
<td>DBG</td>
</tr>
<tr>
<td>SOAPdenovo2</td>
<td>DBG</td>
</tr>
</tbody>
</table>

**TRANSCRIPTOME ASSEMBLY**

RNA-Seq is crucial in generating transcripts and estimating its abundance. Approaches used in assembly of transcriptomes are De novo assembly (Table 3).

**OMICS APPROACHES IN PLANT SCIENCE**

The diagrammatic representation of omics approaches in plant sciences were presented (Figure 2).

![Figure 2. Diagrammatic representation of Omics Approaches in plant sciences](image)

**Table 3. Transcriptome Assemblers**

<table>
<thead>
<tr>
<th>Assembler</th>
<th>Type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>QPALMA</td>
<td>Reference genome-based assembler</td>
<td>De Bona et al. (2008)</td>
</tr>
<tr>
<td>TopHat</td>
<td>Reference genome-based assembler</td>
<td>Trapnell et al. (2009)</td>
</tr>
<tr>
<td>Trinity</td>
<td>De novo assembler</td>
<td>Grabherr, (2011)</td>
</tr>
<tr>
<td>Oases</td>
<td>De novo assembler</td>
<td>Schulz et al. (2012)</td>
</tr>
</tbody>
</table>

**1. Epigenomics**

NGS sequencing has aided the development of epigenome as a new omics field (Schmitz and Zhang, 2011). Plant epigenetic changes are mediated by small RNAs (Matzke et al., 2009) (Table 4). Epigenomes regulate chromatin and genome stability to predict gene information.
2. Interactomics
Interactome analysis helps to elucidate plant cellular functions by Protein–protein interactions (Morsy et al., 2008).

3. Analysis of plant hormones
The study of these hormones’ structures aids in determining the relationship between receptors and hormones and their signalling processes in plants (Tan et al., 2007; Shimada et al., 2008). Several report results revealed how different hormones interact at different stages of development.

4. Plant Metabolomics
Plant metabolomics plays a crucial role in computational approaches to functional analysis of plants. Various reviews have been published that discuss major applications of metabolomes (Sumner, 2010). Few resources of plant metabolomes are PRIMe, Metabolome Express etc.

5. Plant Phenomics
Phenomics is the summation of phenotypes considered at different levels, different time periods and varying environmental conditions. Phenotypic data obtained from public databases are essential for bioinformatics users. The ontologies obtained by the crops under various stress conditions provide a common language for both researchers and computational algorithms.

6. Ionomics
The study of minerals, nutrients, and trace elements in the inorganic pool of cellular and organism systems is known as ionome. Initially, ionomics research focused on the characterization of Arabidopsis and other model organisms (Lahner et al., 2003). The ionomics technique has recently been used for Lotus japonicus and other plant species (Chen Z. et al., 2009). Currently, laboratories are using an ionomics technique to study elements in rice, maize, and soybean (Chen Z. et al., 2009; Baxter et al., 2007; Gandhilyan et al., 2009). The "Arabidopsis ionomics project" at the Purdue Ionomics laboratory, which has examined the leaf ionome of over 125 000 plants, is the largest known ionomics project. PiMIS was created in response to the want to retain track all the data generated by this research (Baxter et al., 2007).

COMPUTATIONAL APPROACHES ADMINISTERED TO DIFFERENT TYPES OF GENE EXPRESSION DATA
1. Gene Regulatory Network Inference
Plant growth and development are better understood using network inference (Gunasekara et al., 2018; Nishiyama et al., 2018; Liesecke et al., 2018), biotic and abiotic stress (Pal et al., 2017; Jiang et al., 2018), and other factors. To conclude, links between various genes, correlation are commonly used. Liesecke et al., 2018 employed correlation coefficients or built an A. thaliana gene network to find genes engaged in the same metabolic pathways (Liesecke et al., 2018). Montes et al. used the MI Inference algorithm from ARACNE (Margolin et al., 2006) to build GRNs in A. thaliana (Chávez Montes et al., 2014). The TGMI approach was utilized to discover regulatory TFs in lignin production pathways in A. thaliana using perturbation data (Gunasekara et al., 2018). The algorithms ExRangeS (Desai et al., 2017), DDGni (Yalamanchili et al., 2014), and CDAA (Koryachko et al., 2015) create a network that shows the link between regulators and possible targets. A. thaliana root stem cell regulators found by GENIST are employed in probabilistic graphical approaches. JRmGRN was used to infer networks in various tissues and under various environmental circumstances (Deng et al., 2018). GENIE3 (Shibata et al., 2018; Huynh-Thu et al., 2010), BTNET (Park et al., 2018), and fused LASSO are three regression tree inference methods. In addition to this, GRNs are inferred using a variety of
inference methods. It actively encourages the creation of plugin tools to enhance the basic functionality. The Biological Networks Gene Ontology tool (BiNGO) was created as a Cytoscape plugin. BiNGO's key advantage over similar programmes and transcriptional coregulation networks, which can be shown in Cytoscape. BiNGO also provides a lot of freedom when it comes to using ontologies and annotations. Aside from the standard GO ontologies, BiNGO's Cytoscape graphs can be visualized, modified, and saved in a variety of ways.

2. Tools used for inference of GRNs

Many user-friendly interfaces, such as TuxNet and Seidr, have been developed to make GRN inference methods accessible and understandable. TuxNet helps you infer a network by plugging in gene expression data and choosing an analysis approach for GRN inference (Luis Balaguer de et al., 2018). Seidr allows you to get an aggregate GRN from various techniques, which you can then evaluate for more inference (Schifftalher et al., 2018). These technologies aid the scientist's familiarity with a large amount of supporting data to understand plant regulatory mechanisms better.

ROLE OF BIOINFORMATICS

The unprecedented amount of biological data obtained from sequencing technologies required computational tools and databases for storage. This led to the exposure of Bioinformatics in the field of science. Bioinformatics was initially functional in the arrangement and governance of biological data, but with time it also emerged as a beneficial field comprising enormous software to perform data analysis, interpretation, modeling and various other applications (Moore et al., 2010).

MODELLING APPROACHES IN UNDERSTANDING PLANT BIOLOGY

The increasing demand for understanding plant functioning (development, growth and functions) for agronomical and horticultural applications led to the emergence of Modelling Approaches (Prusinkiewicz and Runions, 2012). The modelling studies help to analyze different aspects like the effects of stresses on plants (Chen et al., 2018), factors determining the quality of plants and fruits.

1. Biological databases

As indicated, most databases are public repositories that can retain data for lengthy periods and are managed by government and international organizations (Table 7). NCBI established a comprehensive sequential-based taxonomy with the help of external experts and curators EMBL and DDBJ (Federhen, 2003).

Cross-referencing of entries became a significant feature for most databases as bioinformatics grew in popularity. Many additional databases, such as Protein sequence databases (Interprot and Uniprot), Protein structure databases (PDB), gene ontology databases, Pubmed literature databases, and cross-reference DNA entries, allow hosts to correlate information. TAIR was one of the Model Plant Databases that illuminated the information about their role as a model system plant research. Various databases can house the extensive "-omics" data represented in the report (Table 7).

Table 7. Different Biological databases comprising annotated collection of data

<table>
<thead>
<tr>
<th>Databases</th>
<th>Annotated data</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenBank</td>
<td>Gene sequences</td>
</tr>
<tr>
<td>UniProt</td>
<td>Protein informations</td>
</tr>
</tbody>
</table>
The Map Viewer can build plots from the NCBI toolkit's extensive resources (UniGene, LocusLink, and RefSeq), as well as GenBank's large collections of sequence data and annotations. Sequence similarity search capabilities are also accessible in BLAST, which is analogous to Ensembl and the UCSC browser. Tools in the BioPerl package provide less invasive access to NCBI's data (Stajich et al., 2002).

Apollo (Lewis et al., 2002), the Integrated Genome Browser, Artemis (Berriman and Rutherford, 2003), NCBI Genome Workbench, SynView, SynBrowse (Stajich et al., 2002), Sybil, which runs on a GMOD Chado database (Crabtree et al., 2007), are some commonly specified user examples of standalone web-based browsers.

2. Plant genome databases
The generation of tremendous amounts of data in plant science and genome repositories (Table 8). These repositories are both general as well as species-specific. The European plant genome database-PlaNet [http://www.eu-plant-genome.net] comprises plant genomic data.

<table>
<thead>
<tr>
<th>Databases</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>PlantGDB</td>
<td>Dong et al. (2004)</td>
</tr>
<tr>
<td>Wheatgenome.info</td>
<td>Lai et al. (2012)</td>
</tr>
<tr>
<td>Phytozome</td>
<td>Goodstein et al. (2012), Zhang et al. (2015)</td>
</tr>
<tr>
<td>EuroPineDB</td>
<td>Fernández-Pozo et al. (2011)</td>
</tr>
<tr>
<td>TreeGenes</td>
<td>Ware et al. (2002), Wegrzyn et al. (2008)</td>
</tr>
<tr>
<td>Gramene</td>
<td>Tello-Ruiz et al. (2016)</td>
</tr>
<tr>
<td>Sol Genomics</td>
<td>Bombarely et al. (2011)</td>
</tr>
<tr>
<td>Legume information system</td>
<td>Gonzales et al. (2005)</td>
</tr>
<tr>
<td>TAIR</td>
<td>Swarbreck et al. (2008)</td>
</tr>
</tbody>
</table>

3. Sequence similarity and comparative genomic analysis tools
The end of whole-genome sequencing marks the beginning of defining different genes, made possible through the comparative analysis of genomes. Various tools have been constructed to manage these datasets. These tools help to explore the best alignment based on dynamic programming algorithms like the Smith-Waterman algorithm etc. FASTA and BLAST perform sequence similarity search very rapidly (Altschul et al.,1990). GoloMTranscriptome DB (Kopka et al., (2005) and ATTED-II (Obyasachi et al.,2007) are the tools mostly used for plant genome analysis. Comparative
analysis of genomes is also performed via different online platforms like PLAZA (Van Bel et al., 2011) (http://plaza.psb.ugent.be/), GreenPhyl DB (Rouard et al., 2010), (http://south.greencird.fr/), (Table 8).

4. Phylogenomic analysis tools

Phylogenomic studies are performed to analyze evolutionary history and hypothesize the functions of different genes. Similarity-based methods with the reference genome are the systematic approaches used in Bioinformatics. It is based on the principle that similar sequences or structures might perform similar functions, and this is helpful in evolutionary studies (Mathé et al., 2002).

Various affordable open-source software (Table 9) has been continuously evolving. For example, Hyb-Seq, and ExaML.v.3 (Exascale Maximum Likelihood), performs large scale phylogenetic analyses on Intel hardware platforms, is equipped with better phylogenetic analysis algorithms and analyzes data sets up to 55,000 taxa (Kozlov et al., 2015).

Table 9. List of tools and software used for phylogenomic analysis in plants

<table>
<thead>
<tr>
<th>Tools &amp; Softwares</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHYLIP</td>
<td>Guo et al., 2008; Völgyi et al., 2009.</td>
</tr>
<tr>
<td>PAUP</td>
<td>Swofford (2002)</td>
</tr>
<tr>
<td>MEGA</td>
<td>Kumar et al. (2016)</td>
</tr>
</tbody>
</table>

5. Bioinformatics approaches on plant breeding

Genetic maps of varieties of plants and crops have been resolved, the functions of genes can be elucidated from these genetic maps, thus allowing incorporation of candidate genes explored in model plant species into the similar loci of plants or crops to obtain high yield. Relational databases comprising gene sequence, position and function may help breeders predict the hypothesis for producing phenotypes of interest (Deckers et al., 2002). The analysis of genomes in plants helps determine biodiversity based on quantitative traits of the phenotype of interest. QTLs (chromosomal regions that affect a trait) and MAS (markers) are used to select genes (Morgante et al., 2003). French BioMercator analyses the functions and position of QTLs. GWAS tool is mostly used to study the allelic variations for phenotypic diversity and QTL mappings. Some of the Bioinformatics Tools for GWAS are Heap (Kobayashi, 2015), GnPS-Asso (Steinbach, 2015), BIOGPU (Huang, 2015), BHIT (Bayesian high-order interaction toolkit) (Wang, 2015). These computing tools effectively analyze large data sets with high performance.

CONCLUSION

We attempted to outline recent Systems Biology techniques in Plant Research in this study. In genomics and transcriptomics research, sequencing throughput is no longer an issue. The NGS technology, the most widely used sequencing approach, is featured, as is the freshly developed NGS platform. In one or more iterations, it generates enough depth findings for genome assembly. However, complex genomic sequence assembly and annotation are still tricky. Differences hamper the cross-platform analysis of diverse datasets. Therefore, it might be not easy to draw useful conclusions from large amounts of data. Life
processes now digitized, and saved in a computer's hard disc thanks to advances in computer technology and high-throughput analytic systems. Many new bioinformatics tools and software packages are in erg of development, resulting in new dimensional prospects added continuously to functional genomics and sub-branches. The scientists may obtain the data from these databases and analyze the genomes of different plant species to continue with their research work.

Bioinformatics contributes to multi-disciplinary interactions and helps to reshape agricultural practices and productions, providing knowledge for improved plant quality, and it also provides a plan of action for protection against adverse environmental conditions. Various tools have been developed containing relational databases comprising gene sequences, genetic maps, functions of model plants. The bioinformatics approach is a vital part of plant science research, and if plant researchers start incorporating bioinformatics approaches in their investigations, it may help the breeders to create plants of desired phenotypes.

REFERENCES


