# Omics Approaches for Crop Improvement

## Abstract

The term "omics" refers to a collection of high-throughput biological research methods that allow for a detailed examination of biological systems. Omics techniques have a significant influence on crop development. New crop varieties that are better suited to specific environmental circumstances and more resistant to pests and diseases may be developed using the knowledge gained from omics technology. Despite the wide variety of omics techniques, the focus of this chapter is to give readers an in-depth understanding of the genomic, transcriptomic, proteomic, and metabolomic approaches of crops that can be used to sequence and identify the genes, proteins, and metabolites that are involved in important traits that are highly necessary for the enhancement of crops.

***Keywords:*** *Omics, Genomics, Transcriptomics, Proteomics, Metabolomics*

## Authors

**Sourav Kumar Moharana**
Ph.D. Scholar,
Division of Genetics,
ICAR – Indian Agricultural Research Institute, New Delhi, Delhi - 110012

**Dr. Madhu Bala**
Assistant Professor,
Department of Genetics and Plant Breeding, Navsari Agricultural University, Navsari, Gujarat – 396 450

**Ankita Anil Chakote**
Ph.D. Scholar,
Department of Agricultural Biochemistry, Chandra Sekhar Azad University of Agriculture and Technology, Kanpur, UP - 208002

**Jayaluxmi Devi Khumukcham**
Ph.D. Scholar,
Division of Genetics,
ICAR – Indian Agricultural Research Institute, New Delhi, Delhi – 110012

# I. INTRODUCTION

Although the use of omics techniques for crop improvement is still in its infancy, it has the potential to completely alter how crops are developed. Omics approaches can be used to create novel cultivars of crops that are more productive, nutrient-rich, more adaptable to specific environments, more resistant to biotic/ abiotic stresses, and more sustainable by providing an extensive knowledge of the genetic, transcriptomic, proteomic, and metabolomic outlines of crops. The four major omics used in plant breeding are discussed below.

## Genomics

“Genomics” is an umbrella term for the sub-branch of genetics which deals with finding the genome sequences, mapping the genes, deciphering the mutagenic and epigenomic changes, establishing the functions associated and revealing the evolutionary relationship among the individuals. It was coined by Thomas Roderick in the year 1986. Broadly, it is studied under three headings- a) Structural genomics, b) Functional genomics and c) Comparative genomics.

### 1. Structural Genomics

As the name suggests, the task of identifying nucleotides order and structure of chromosomes thereof, in an entire genome of an organism falls under this section of genomics. When it targets sequencing and genotyping of all the individuals at a taxon level (e.g., of a species) we call it pan-genomics. Pan-genome encompasses core genome of a species along with dispensable and unique genomes. It has contributed in many crop improvement programs including enhancing grain quality, imparting biotic and abiotic resistance etc. (Campbell *et al.*, 2016; Ram *et al.*, 2007 and Zhang *et al.*, 2006). The basic idea of structural genomics is to sequence the genome using available sequencing platforms, look for polymorphism, establish marker-trait associations, construct a physical map or linkage map and genome-wide selections.

The progress in structural genomics was always dependent on the parallel development of sequencing platforms and the gradual discovery of markers. Earlier technologies of DNA microarray were quite efficient in understanding the genetic diversity, curating germplasm accessions and focuses mainly on genic regions of the genome. However, it was not suitable to study the repetitive non-genic regions and identifying de novo transcripts (Dai *et al.*, 2022).

In 1977, Sanger sequencing came which still stays the backbone of modern sequencing platforms. With relatively long read lengths (~1000 bp), high cost and low throughput, it dominated the world for long 30 years. With the advent of second-generation sequencing platforms, the revolution in the field of plant genomics was seen. It was when for the first time the genome sequence of model plant *Arabidopsis thaliana* was sequenced and analyzed in the year 2000 (*Arabidopsis* Genome, 2000). NGS platforms like SOLiD, 454 Roche and Illumina made significant contributions to the field of genomics. These were of low cost, high throughput but having shorter read lengths and called for third-generation sequencing platforms like PacBio and Nanopore technologies which possess low-cost, less error, high-throughput, much longer read lengths of hundreds of kilobases and can detect epigenetic modifications and have no problem with homologous repeats or low/high GC %.

With the gradual understanding of the importance of genome-wide association mapping and greater efforts to study larger populations, microarray technologies again have caught attention of scientists and now we can see the integration of DNA microarray technologies with NGS platforms. DArT-seq is one such novel technology which efficiently provides both dominant and co-dominant markers in one single platform (Toporowska *et al.*, 2021). Starting from Sanger's dideoxy chain termination technique to recent platforms of Single Molecule Real Time sequencing, it is indeed a great leap for the scientific community in understanding the genetic architecture of an organism.

## 2. Functional Genomics

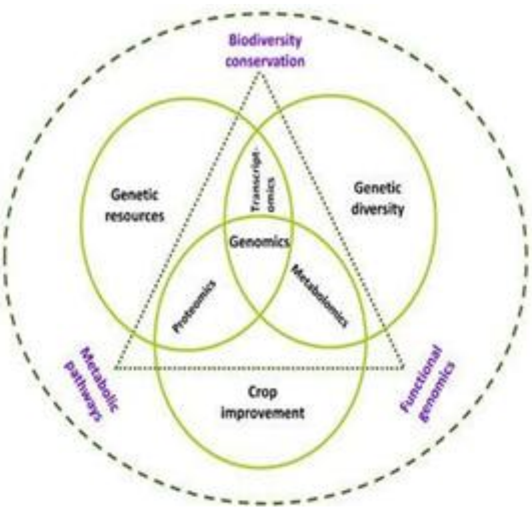


**Figure 1:** Crop improvement and understanding metabolic pathways through functional genomics (Maghuly & Laimer, 2013)

Functional genomics essentially adds meaningful information to the data generated by structural genomics approaches. It encompasses formulating genome-wide experiments to arrive at the functional aspects of the DNA sequences or more precisely to reveal function of genes (Hieter and Boguski, 1997). As contrast to the static components of the genomic information, such as DNA sequence or structures, functional genomics addresses the dynamic aspects of the genome, like gene transcription, translation, control of gene expression, and protein-protein interactions. Microarrays, ChIP-Seq, RNA-Seq, TILLING, CRISPR/Cas9, TALEN, RNAi, computational biology are some of the widely used tools and techniques in functional genomics. In broader sense, it uses transcriptomics and proteomics to identify and characterize genes, as well as to study their interactions with other genes and proteins. With the development in the field of metabolomics, it has also been depicted as a potential assist to functional genomics (Bino *et al.*, 2004).

## 3. Comparative Genomics

Comprehending the genomic distinctions and similarities between taxonomic groups is a key objective of evolutionary genetics and so as of the comparative genomics. It is crucial to understand the conserved genetic blocks among taxa to trace the common ancestry and to know how the information has passed down the lineage (Tang *et al.,* 2008).

Sequence alignment, phylogenetic analysis, GWAS, comparative transcriptomics and comparative proteomics are some of the techniques used in comparative genomics. BLAST, ClustalW, GeneWise, Ensembl Genome Browser, GeneID etc. are some of the basic bioinformatics tools used for comparing and aligning DNA sequences for finding homology between species. The application of nuclear rRNA genes' internal transcribed spacers (ITS), which are highly diverse and often used in molecular phylogeny at the species level but have significant drawbacks, as a DNA barcode is reviewed as a new attempt in comparative genomics by Shneyer, 2009. In order to compare various genomic architecture and organisation and to overcome the constraints of sequencing data alone (Lewin *et al.*, 2009), efforts are being put to consider expression data in the tools as well.

**Table 1:** Some of the Advanced Tools for Comparative Genomics

| Type | Tool | Website |
|---|---|---|
| sequence-based | Phytozome<br>Plaza PlantGenie | https://phytozome.jgi.doe.gov/<br>http://bioinformatics.psb.ugent.be/plaza/<br>http://plantgenie.org/ |

| | | |
|---|---|---|
| expression-based | Cornet Bar<br>Planex Atted-II<br>Podc<br>Aranet | https://bioinformatics.psb.ugent.be/cornet<br>http://bar.utoronto.ca/expressolog_treeviewer<br>http://planex.plantbioinformatics.org/<br>http://atted.jp/<br>http://plantomics.mind.meiji.ac.jp/podc<br>http://www.functionalnet.org/aranet |
| expression-network- based | Planet<br>Complex | http://aranet.mpimp-golm.mpg.de/<br>http://complex.plantgenie.org/ |

(Source: Singh *et al.,* 2021)

## II. TRANSCRIPTOMICS

Crop improvement has been reconstituted by transcriptomics, which is the analysis of all the RNA molecules in a cell or organism at a given time. Dai *et al.,* (2022) has mentioned that “traditional transcriptome refers to mRNA transcripts, but can also be generalized to include other types of transcripts such as microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNA)”.

Although there are several technologies used in transcriptomics, RNA microarray and RNA-Seq are the most widely utilized. RNA-Seq is a powerful technique based on NGS technology that has fundamentally altered how crops are developed by assisting in the discovery of important genes involved in crucial agricultural traits including yield, disease resistance, and nutrient absorption. ABioTrans (Zou *et al.*, 2019), is an additional tool that may be used for analyzing RNA-Seq data files. Researchers may precisely track gene expression during many biological processes using RNA-Seq (Wang *et al.,* 2009).

**Table 2:** RNA Sequencing Techniques

| Methods | Examples |
|---|---|
| Tag-based methods | DGE seq, 3’ end seq |
| Probe alternative splicing and gene fusion | SMRT, SLR-RNA-Seq |
| Targeted RNA sequencing | Target capture, Amplicon sequencing |
| Single cell RNA sequencing | CEL-seq2, Drop-seq |

(Source: Dai *et al.,* 2022)

Breeders may choose crops with desirable features and create more robust varieties by examining how the genes express under various environments. It may also be used to discover genes that are responsible for stress tolerance, in order to create crops that are more robust to tolerate stresses like drought, high temperatures, and other biotic/abiotic adverse conditions. Transcriptomics may also be utilized to create novel genetic engineering methods for crop enhancement and to investigate the processes of gene regulation.

## Proteomics

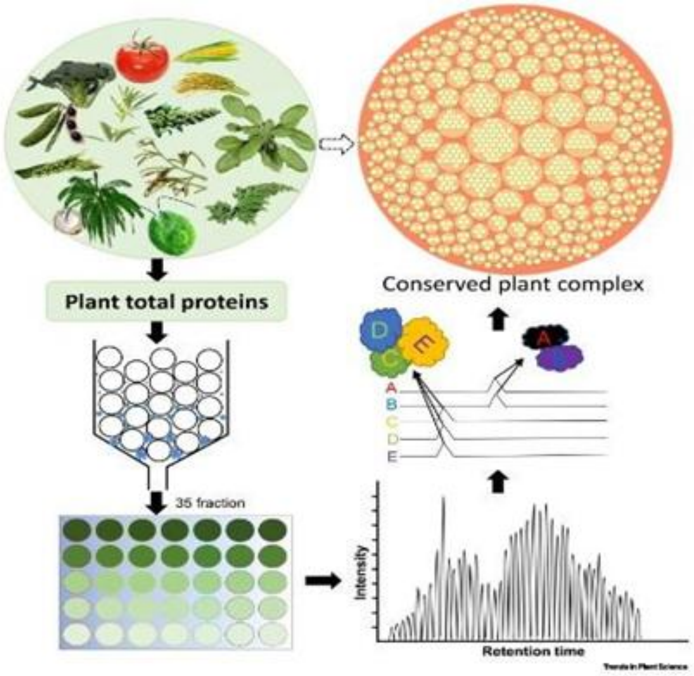


**Figure 2:** Co-Fractionation MS for Discovering the Conserved Plant Protein Complexes (Zhang *et al.*, 2020)

Proteomics is the study of all of the proteins that make up a cell or an organism, including their structural and functional characteristics. Post-translational modifications, subcellular localization, protein-protein and protein-metabolite interactions, interactions between proteins and other biomolecules, etc. are all covered under proteomics. Additionally, it makes an effort to compare and contrast the proteomes of various species at the taxon level.

Proteomics has helped to clarify the intricate connection between stress tolerance and crop yield, paving the way for the creation of better crops (Beyene & Haile, 2016; Kumar *et al.*, 2016). The identification of desirable genes and markers for screening in breeding programmes has made proteomics a significant tool in crop development. Protein-protein interaction studies have discovered more proteins that are differently expressed under adverse conditions, like heat shock proteins (HSPs) in stress tolerance (Eldakak *et al.*, 2013; Kumar *et al.*, 2023). In order to establish core assemblies and give a clearer interpretation of the "genotype-phenotype space" of plants, proteomics aids in the analysis of protein complexes from various species. This understanding of vital protein assemblies might help improve crops in the future (McWhite *et al.,* 2020; Zhang *et al.,* 2020).

Some of the bioinformatics tools frequently used in proteomics are InterProScan, Pfam, SwissProt, and others. SDS-PAGE, ELISA, Mass spectrometry (MS) etc. are some well-known techniques in proteomics. There are many recent technical developments which has made proteomics more efficient.

**Table 3:** Techniques widely used in Proteomics with their Purposes

| Purpose | Name of the Technique |
|---|---|
| Determining the amino acid sequence | HPLC |
| Identification of protein structure | X-ray diffraction, electron microscopy, NMR spectroscopy, crystallography |
| Protein identification | 2D-PAGE, LC–MS/MS, BIA–MS, MALDI– TOF–MS |
| Proteome comparison | iTRAQ, ICATs,MudPIT, TMTs |
| Protein biomarker validation | SRM, PRM, MRM |

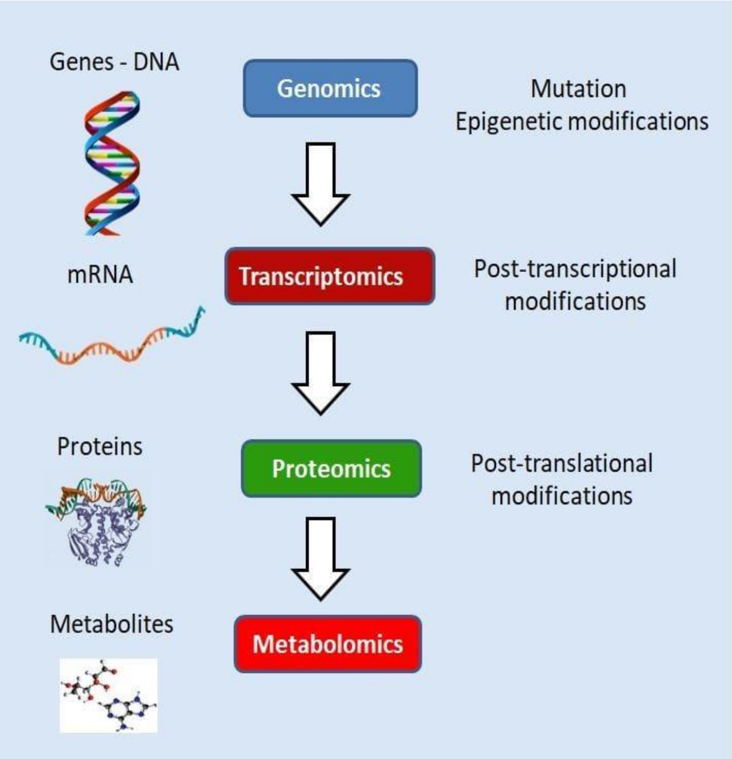
(Source: Singh *et al.,* 2021; Yang *et al.,* 2021)

There is an increase in efforts understanding the interaction between proteins and metabolites as it has been slowly revealed that metabolites also act as ligands/ cofactors and can affect structure and function of proteins by allosteric or competitive binding with the protein molecules (Dai *et al.,* 2020; Rinschen *et al.,* 2019).

## Metabolomics

The term "metabolome" refers to the collection of all metabolites found in a cell, which includes all biomolecules with the exception of the metals, transcriptome, genome, and proteome. The branch of omics dealing with metabolome is nothing but metabolomics. It may ascertain the metabolic expression of proteins and pinpoint the biochemical procedures essential to the operation of genes.

Due to the fact that plants create more metabolites than either animals or microorganisms, metabolomics is a crucial field in plant stress physiology. The secondary metabolites that plants generate are advantageous in coping with environmental stress. Plants respond to various abiotic stress by changing their metabolite composition. It was possible to find both general and particular changes in the basic metabolism of Arabidopsis to a variety of abiotic stresses assessed in separate trials over several years (Obata and Fernie, 2012; Obata *et al.*, 2016).



**Source:** (https://www.slideshare.net/Rudresh Rudri3/plant-metabolomics-157540410)

The research of metabolomics can be conducted using a variety of techniques like GC- MS, LC-MS, GC/EI-TOF-MS, NMR spectroscopy, Fourier transform infrared (FTIR) spectroscopy, direct flow injection (DFI) mass spectrometry, etc. Several of these strategies have been utilized regularly in response to biotic and abiotic stress conditions in crop species including rice, maize, soybean, tomato and several medicinal plants (Ghatak *et al.*, 2018; Kiran *et al.*, 2019). Additionally, metabolomics data may be analyzed to identify relevant metabolites using bioinformatics tools like MetaboAnalyst and MetFrag (Johnson *et al.*, 2015).

## III.SUMMARY

Crop enhancement depends on the contributions of all four scientific disciplines: genomics, transcriptomics, proteomics, and metabolomics. The genotype of crops' phenotypic characteristics may be identified using genomics.

While proteomics examines protein expression, transcriptomics explores gene expression at the transcript level. Both methods are employed in functional genomics to comprehend physiological and patho-physiological processes, and identify genes and proteins for inspection, diagnosing, or monitoring crop diseases. The study of metabolites generated by plants and their function in their growth and development is known as metabolomics. Researchers may gain insight into the genetic and metabolic frameworks of agricultural plants and create more effective methods for crop improvement by merging these four domains.

## REFERENCES

[1] Beyene, B. and Haile, G. (2016). Review on proteomics technologies and its application for crop improvement. Genome, 7(6).
[2] Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J. and Sumner, L. W. (2004). Potential of metabolomics as a functional genomics tool. Trends in plant science, 9(9): 418-425.
[3] Campbell, B. C., Gilding, E. K., Mace, E. S., Tai, S., Tao, Y., Prentis, P. J., ... and Godwin, I. D. (2016). Domestication and the storage starch biosynthesis pathway: signatures of selection from a whole sorghum genome sequencing strategy. Plant biotechnology journal, 14(12): 2240-2253.
[4] Dai, X., Shen, L. and Lu, P. (2022). Advances and trends in omics technology development. Frontiers in Medicine, 1546.
[5] Dai, Z., Ramesh, V. and Locasale, J. W. (2020). The evolving metabolic landscape of chromatin biology and epigenetics. Nature Reviews Genetics, 21(12): 737-753.
[6] Eldakak, M., Milad, S. I., Nawar, A. I. and Rohila, J. S. (2013). Proteomics: a biotechnology tool for crop improvement. Frontiers in Plant Science, 4: 35.
[7] Ghatak, A., Chaturvedi, P. and Weckwerth, W. (2018). Metabolomics in plant stress physiology. Plant genetics and molecular biology, 187-236.
[8] Hieter, P. and Boguski, M. (1997). Functional genomics: it’s all how you read it. Science (New York, NY). 278: 601–602. doi: 10.1126/science.278.5338.601
[9] Johnson, C. H., Ivanisevic, J., Benton, H. P. and Siuzdak, G. (2015). Bioinformatics: the next frontier of metabolomics. Analytical Chemistry, 87(1): 147-156.
[10] Kiran, Kumar P., Kirti S., and Daudi A.A. (2019). Phytochemical analysis and antioxidants activity of Silybummarianum (L.) Gaertn Annals of phytomedicine 8(1): 127-134. http://dx.doi.org/10.21276/ap.2019.8.1.16
[11] Kumar, P., Prasad, R., & Kumar, R. (2016). Underutilised Millets: a way to nutritional security. International Journal of Agticulture and Environment Research, 4, 726-734.
[12] Kumar, P., Kumar, P., Suniti, Kumar, U., Avni, & Mann, A. (2023). Transcriptional Regulatory Network Involved in Drought and Salt Stress Response in Rice. In Salinity and Drought Tolerance in Plants: Physiological Perspectives (pp. 237-274). Singapore: Springer Nature Singapore.
[13] Lewin, H. A., Larkin, D. M., Pontius, J. and O'Brien, S. J. (2009). Every genome sequence needs a good map. Genome research, 19(11): 1925-1928.

[14] Maghuly, F.and Laimer, M. (2013). Jatropha curcas, a biofuel crop: functional genomics for understanding metabolic pathways and genetic improvement. Biotechnology journal, 8(10): 1172-1182.
[15] McWhite, C. D., Papoulas, O., Drew, K., Cox, R. M., June, V., Dong, O. X., ... and Marcotte, E. M. (2020). A pan-plant protein complex map reveals deep conservation and novel assemblies. Cell, 181(2): 460-474.
[16] Obata, T. and Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. Cellular and Molecular Life Sciences, 69: 3225-3243.
[17] Obata, T., Florian, A., Timm, S., Bauwe, H. and Fernie, A. R. (2016). On the metabolic interactions of (photo) respiration. Journal of Experimental Botany, 67(10): 3003-3014.
[18] Pandey, M. K., Roorkiwal, M., Singh, V. K., Ramalingam, A., Kudapa, H., Thudi, M., ... and Varshney, R. K. (2016). Emerging genomic tools for legume breeding: current status and future prospects. Frontiers in Plant Science, 7: 455.
[19] Ram, T., Majumder, N. D., Krishnaveni, D. and Ansari, M. M. (2007). Rice variety Dhanrasi, an example of improving yield potential and disease resistance by introgressing gene (s) from wild species (Oryza rufipogon). Current Science, 987-992.
[20] Rinschen, M. M., Ivanisevic, J., Giera, M. and Siuzdak, G. (2019). Identification of bioactive metabolites using activity metabolomics. Nature reviews Molecular cell biology, 20(6): 353-367.
[21] Shneyer, V. S. (2009). DNA barcoding is a new approach in comparative genomics of plants. Russian Journal of Genetics, 45: 1267-1278.
[22] Singh, R. K., Sood, P., Prasad, A. and Prasad, M. (2021). Advances in omics technology for improving crop yield and stress resilience. Plant Breeding, 140(5): 719-731.
[23] Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M. and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. Science, 320(5875): 486-488.
[24] The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, 408(6814): 796–815.
[25] Toporowska, J., Sowa, S., Kilian, A., Koroluk, A. and Paczos-Grzęda, E. (2021). Discovery and chromosomal location a highly effective oat crown rust resistance gene Pc50- International Journal of Molecular Sciences, 22(20): 11183.
[26] Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics, 10(1): 57-63.
[27] Yang, Y., Saand, M. A., Huang, L., Abdelaal, W. B., Zhang, J., Wu, Y., ... and Wang, F. (2021). Applications of multi-omics technologies for crop improvement. Frontiers in Plant Science, 12: 563953.
[28] Zhang, X., Zhou, S., Fu, Y., Su, Z., Wang, X. and Sun, C. (2006). Identification of a drought tolerant introgression line derived from Dongxiang common wild rice (O. rufipogon Griff.). Plant molecular biology, 62: 247-259.
[29] Zhang, Y., Skirycz, A. and Fernie, A. R. (2020). An abundance and interaction encyclopedia of plant protein function. Trends in plant science, 25(7): 627-630.
[30] Zhang, Y., Skirycz, A. and Fernie, A. R. (2020). An abundance and interaction encyclopedia of plant protein function. Trends in plant science, 25(7): 627-630.
[31] Zou, Y., Bui, T. T. and Selvarajoo, K. (2019). ABioTrans: a biostatistical tool for transcriptomics analysis. Frontiers in Genetics, 499.

## Questions

1. What are the advantages of using NGS platforms in genomics?
2. What are some of the RNA sequencing techniques used in transcriptomics?
3. Compare and contrast the key objectives of structural genomics, functional genomics, and comparative genomics in the context of crop improvement. Provide examples of tools and techniques used in each of these genomics branches and their applications in understanding the genetic and evolutionary aspects of crops.
4. Examine the importance of metabolomics in understanding plant stress physiology, focusing on how metabolite composition changes in response to abiotic stress.

## Self Assesment

1. Which approach within the context of omics techniques primarily involves sequencing the genome, mapping genes, studying mutagenic and epigenomic changes among the individuals?
   a. Functional genomics
   **b. Structural genomics**
   c. Comparative genomics
   d. Transcriptomics

2. How does functional genomics complement structural genomics?
   a. By focusing solely on genic regions
   **b. By exploring the dynamics of gene expression and regulation**
   c. By ignoring the static aspects of genomic information
   d. By neglecting the role of mutations and epimutations

3. In what context do comparative genomics techniques aid in crop improvement?
   **a. To determine the conservation of genetic blocks across taxa**
   b. To ignore the differences between closely related species
   c. To overlook the significance of evolutionary history
   d. To disregard the value of synteny analysis

4. Which area of proteomics focuses on determining the amino acid sequence of proteins?
   **a. Peptide mass fingerprinting**
   b. Two-dimensional gel electrophoresis (2D-PAGE)
   c. High performance liquid chromatography (HPLC)
   d. Polymerase chain reaction (PCR)

5. What does genomics primarily focus on in crop improvement?
   a. Protein-protein interactions
   **b. Nucleotide order and chromosome structure**
   c. Metabolic pathways
   d. Epigenetic modifications

6. Which technology is considered the backbone of modern sequencing platforms?
   a. DNA microarray
   **b. Sanger sequencing**
   c. NGS platforms
   d. PacBio technology

7. Which bioinformatics tool is commonly used for sequence-based comparative genomics?
   a. Planet
   b. Phytozome
   c. Complex
   **d. VISTA**

8. What does transcriptomics primarily analyze in a cell or organism?
   a. Nucleotide order
   b. Protein structures
   **c. All RNA molecules at a given time**
   d. Metabolites

9. ABioTrans is a tool used for analyzing data related to which omics technique?
   a. Genomics
   b. Proteomics
   **c. Transcriptomics**
   d. Metabolomics

10. Proteomics aims to clarify the intricate connection between what factors in crops?
    a. Genes and metabolites
    **b. Proteins and metabolites**
    c. Genes and proteins
    d. Stress tolerance and crop yield

11. Which tool is NOT commonly used in functional genomics?
    a. ChIP-Seq
    b. TILLING
    **c. Mass spectrometry**
    d. CRISPR/Cas9

12. Which sequencing technique is tag-based and commonly used in transcriptomics?
    **a. DGE seq**
    b. X-ray diffraction
    c. MALDI–TOF–MS
    d. iTRAQ

13. Which tool is used for analyzing metabolomics data to identify relevant metabolites?
    a. InterProScan
    b. Pfam
    **c. MetaboAnalyst**
    d. ClustalW

14. Which sequencing platform dominated the field for a long 30 years before the advent of NGS platforms?
    a. Illumina
    b. SOLiD
    **c. Sanger sequencing**
    d. PacBio

15. Which technique is NOT used in metabolomics?
    a. GC-MS
    b. LC-MS
    **c. 2D-PAGE**
    d. NMR spectroscopy

16. Which tool is used for functional genomics to study gene interactions?
    a. Ensembl Genome Browser
    b. B. TALEN
    c. C. RNA-Seq
    **d. Aranet**