

Review Article

<https://doi.org/10.20546/ijcmas.2020.907.223>

## An Overview on Systems Biology Driven Plant Breeding

Manoranjan Senapati<sup>1\*</sup>, Sai S. Dash<sup>2</sup>, Biswajit Lenka<sup>3</sup> and Debadatta Panda<sup>4</sup>

<sup>1</sup>Division of Genetics, ICAR- IARI, New Delhi-110012, India

<sup>2</sup>Department of Plant Breeding and Genetics, College of Agriculture,  
OUAT, Bhubaneswar- 751003, India

<sup>3</sup>Department of Genetics and Plant Breeding, College of Agriculture,  
JAU, Junagadh-362001, India

<sup>4</sup>Department of Genetics and Plant Breeding, TNAU, Coimbatore-641003

\*Corresponding author

### ABSTRACT

#### Keywords

Plant systems  
biology, Omics  
data, modelling and  
simulation

#### Article Info

Accepted:  
17 June 2020  
Available Online:  
10 July 2020

In biological science, we mainly focus to comprehend life in its wholeness. Development in the field of genetics and molecular biology has immensely contributed to the advancement in plant science. This gradual progress in these fields leads to accumulate large amount of data at different levels including genomics, transcriptomics, proteomics and metabolomics. This vast information of data can be studied in together with the help of suitable mathematical models to give a final idea about how plant systems responses to different factors. In this review we provide an idea about what is systems biology and how it operates in plant systems.

### Introduction

In biological science that deals with the study of all living organisms, we mainly want to comprehend life in its entirety. We study all organism at different levels starting from its phenotype to its corresponding genotype. We are more focused in examining the structural and functional details of living organisms including their origin, evolution and classification. For actually comprehending life, we have to study not only the individual components, that are complex enough, but

also needs to examine their interaction with other components at various level of biological function (De Costa *et al.*, 2018).The science of plant breeding is believed to be originated with the process of domestication of wild types followed by long-term selection. Mendel's law and the gradual progress in understanding genetics helps in identifying the number of genes governing a trait. Development in plant molecular biology and biotechnology allowed us to categorize, map and clone individual genes. Again, specific allele of a gene can be detected using

molecular markers and high throughput sequencing technologies. These advanced sequencing technologies generates vast amount of data with respect to Genomics (for genotypic data), Transcriptomics (for transcript data), Proteomics (for protein data) and Metabolomics for metabolite data (Yuan *et al.*, 2008). Although we are more attentive towards studying one of these components at a time, we need to study all components together and try to identify their inter-relationship among them. Hence it is a necessary to shift our strategies from reductionist paradigm to holistic paradigm. In holistic paradigm, individual units are not only collectively studied together but also integrated in such a way that they express the dynamic property of biological systems.

Comparison between holism and reductionism is not a novel concept, rather it's an age-old practice. Reductionism enlightens the individual key component is answerable for the systems behaviour. On the opposite site, holism follows the property of the systems in its entirety. The knowledge of holism can be taken out from Aristotle's *Metaphysic* that explains "The whole is superior than the summation of its parts".

Reductionism describes single major component, and believe that whole property can be fully understood, if you examine its components and the property of their integration (Yin and Struik, 2007). In contrary, holism examines the property of the whole systems in entirety and hence it is an integrative approach. Holism believes that the whole systems contain some properties that cannot be discovered by studying its component parts. The reductionist technique has identified most of the components and various relations with others in a biological system but, it does not provide any significant perceptions to realize how system properties arise (Sheth and Thaker, 2014). The different

types of cause and effect in a biological network system are better understood by studying quantitative procedures of several components simultaneously and integration of these huge data by using good mathematical models.

### **Plant Systems biology**

As a novel emerging branch of biological science, plant systems biology deals with systematic learning of complex interactions in biological systems. Systems biology is nothing but taking all components together instead of considering individually, it practices integration strategies rather than reduction approaches (Yuan *et al.*, 2008). Mathematical models that can forecast or describe complex interaction processes of biological systems are considered here. A comprehensive view on how systems biology operates in plants is discussed in fig 1.

Changing from reductionistic to holistic approaches can only be possible if the complete informational data regarding individual components are accessible. Advancement in high-throughput techniques like next generation sequencing tools, yeast two-hybrid screens, microarrays, protein chips etc., support in examining cell at any given stage and time period. Based on this individual component information and by using appropriate mathematical modelling strategies, we can guess several types of relations and interaction networks among these components at different level of strata for example, nucleotide-nucleotide interaction, nucleotide – protein interaction, interaction at transcriptional, translational and metabolite level, signalling networks and gene regulatory networks. Systems biology can be used to find genetic diversity with respect to different omics data (Kliebenstein, 2010).

Numerous agriculturally important traits are complex in nature which are controlled by quantitative genes with several interactive genes. These traits are again influenced by developmental stage and environmental conditions. There is a need of comprehensive understanding about these traits (Libault *et al.*, 2017). Genetic information flows linearly from DNA level to transcript level, protein level, metabolite level and finally to the phenotypic level. Based on these flows of information there are mainly three approaches for comprehensive understanding of biological systems, starting from end (i.e. from phenotypic level called top-down approaches); starting from initial level (i.e. from genomic level called bottom-up approaches) and starting in the middle known as middle-out approaches (Yin and Struik, 2007). Among these strategies bottom-up and top-down approaches are mostly followed to describe biological systems. A schematic representation of approaches in systems biology is given in fig 2.

### **Components in systems biology**

Plant systems biology helps in examining the plants for various abiotic and biotic responses by monitoring its corresponding factors like genes, RNAs, proteins, metabolites and integrating these by framing suitable mathematical model to clarify the structure and function of biological systems in a holistic way (Coruzzi *et al.*, 2018). In general plant systems biology can be developed by including the following things-

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics
- Interactomics
- Phenomics
- Integration of multiple 'omics' data
- Modelling and simulation

### **Genomics**

The term 'genome' means complete set of genetic material of an organism and genomics is the study of whole genome of an organism. Genomics as a scientific branch in its own principle, lies at the basal level of understanding for any complex traits and it gives an idea about organisms' behaviour (Oliver *et al.*, 1998). It is necessary to understand genomics and their role in the development of an organism's system. Several novel next Generation Sequencing (NGS) strategies were developed to sequence the complete genome of an organism in a high throughput and cost-effective way (Falter-Braun *et al.*, 2019). Complex traits that were governed by quantitative trait loci (QTLs) can be identified and mapped using suitable mapping population, molecular markers and appropriate statistical tools. Again, Association mapping have capable enough to identify smaller effect QTLs with high resolution. Development of single nucleotide polymorphisms (SNPs) markers by sequencing the DNA help in accelerate these processes. Size of the genome and the complexity of organism are the major challenges in complete genome sequencing.

Annotation of function and deposition of this genomic sequence data into the public databases is the ultimate step in genomics. This process leads to development and progress of a new branch of bioscience called bioinformatics, where large amount of complexed data can be mined and utilized further. Currently available public data base like national center for biotechnology information (NCBI Genbank), European Molecular Biology Laboratory (EMBL) and DNA Databank of Japan (DDBJ) contains more than 7.4 million plant DNA sequences including both organellar and nuclear genome.

## **Transcriptomics**

Transcriptome means complete set of all the RNAs present in cell. Term transcriptomics denotes the study of transcriptome. It considers all the genomic counterparts that were either expressed as mRNAs or noncoding RNAs like tRNAs, rRNAs and short RNAs at a particular time in a cell for any given environment (Wang *et al.*, 2010). Conventionally transcriptional studies include northern blots analysis but with the advancement of several high throughput and cost-effective technologies it is possible to elucidate the complete transcriptome of any model and non-model plant at a particular time. The up-to-date methods of transcriptome analysis are microarray. Microarray analysis has been used since last many years and it provides a good way of understanding towards gene regulatory networks in varying environmental condition. An Arabidopsis based technology known as ATH1 Genome Array that is developed by Affymetrix™ is mostly used for transcriptional studies.

## **Proteomics**

The term Proteome is coined by Wilkins *et al.*, (1996) is the whole set of protein present in the system at a particular time and environmental conditions. The systemic study of proteome is known as proteomics. The availability of large amount of post translational modifications makes the proteome more difficult to analyse (Wilkins *et al.*, 1996). Hence, it is more complex and dynamic in nature when compared with transcriptomics. The over-all workflow of proteomics studies includes separation of proteins from other impurities, treatment with enzymes like proteases and subsequently polypeptide mass fingerprinting. It helps in determining the mass of the fragmented peptides and it is sequenced with the help of

MS/MS procedure followed by searching in databases. Now a days, A wide range of options are available to elucidate proteome like mass spectrometric methods, Edman sequencing and classical 2D Gel electrophoresis. Among this Mass spectrometry technique is more popularized and several variants of this like MS-MS, MALDITOF MS (Matrix-Assisted Laser Desorption Ionization Mass Spectrometry) and ESI-MS (Electron spray ionization) have contributed more.

## **Metabolomics**

Metabolite means the small molecules that are produced inside the body. The complete set of metabolites is called metabolome and the term coined by Oliver *et al.*, (1998). The complete analysis of all metabolome in an organism under a given environmental condition is called metabolomics. There is larger heterogeneity found in metabolites with respect to their physical and chemical properties including their stability, quantity, size and polarity (Fiehn 2002). Hence these are more complex in nature as compared to genome and transcriptome.

A vast range of analytical platforms are available to study metabolome and among this mass spectrometry and its variants are frequently used (Keurentjes *et al.*, 2006). Analytical platforms like NMR (nuclear magnetic resonance), CE-MS (capillary electrophoresis mass spectrometry), GC-MS (gas chromatography mass spectrometry), LC-EC-MS (liquid chromatography electrochemistry mass spectrometry), FT-MS (Fourier transform ion cyclotron mass spectrometry), DI-MS (direct infusion mass spectrometry), IR (infrared spectroscopy) and HPLC (high-performance liquid chromatograph) equipped with UV or fluorescent detectors are popularly used in this context.

## **Interactomics**

Active and systemic interaction is required for an effective function of a cell or any system (Sauer *et al.*, 2007). Several types of interactions are found among different types of macromolecules (like DNA or RNA, Protein and lipid) and small molecules (like Metabolites). Among these interactions, protein-protein and nucleotide – protein interactions are most frequent to be found. Formation of chromatin by binding of Histone to DNA is a common example of nucleotide-protein interaction. Several *in vitro*, *in vivo* and *in silico* methods are available to study these interactions. Among these technologies split-ubiquitin system (SUS), Split-luciferase system, yeast two-hybrid (Y2H), bimolecular fluorescence complementation (BFC), bioluminescence resonance energy transfer (BRET), fluorescence resonance energy transfer (FRET) are more popularly used.

## **Multipleomics data integration**

The advancement in technologies has enabled us to accumulate high throughput and accurate data at about individual omics level. However, as already discussed individual omics data is not sufficient to recover our target of getting the full idea about the plant systems. Hence, individual data need to be summed up using suitable bioinformatics tools (Aggarwal and Lee, 2003). Prior to integration, individual omics data are need to be pre -processed. Building comprehensive predictive models that should sound statistically good is required for integration of several data.

## **Modelling and simulation of data**

The key step in system biology is successful modelling that helps in comprehensive understanding. Here in systems biology, cells or further higher organisations like tissue or plant parts are considered as systems of

networking elements. For a comprehensive explanation, it is necessary to know the identity of individual constituents, their behaviour and dynamic interactions among the components, of the living systems, under consideration. This data information can finally be united to form a model, which is both reliable enough and can be able to forecast the system behaviour for an unknown perturbation. Modelling and simulation will help in joining the gaps between theoretical concepts and experiments (Lavarenne *et al.*, 2018).

A biological network has basically two parts: the components of the system and the interaction among them. These components are represented as vertices or nodes and the interaction part are represented as connecting link between nodes called edges. A node can be a DNA or RNA or protein or metabolite that is present inside the cell. An edge can be any type of relation (physical, chemical, metabolite reactions and regulatory connections) found among these components. Different types of edges are seen in biological network e.g. non directed edges (indicating mutual interaction between nodes but unknown flow of information), directed edges (starting from source to sink and it may unidirectional or bidirectional). A large number of nodes that make in contact with other nodes creating a very complex network is referred as hub. Generally, the number of hubs in a biological system is much lesser than that of nodes with less edges. A biological model can be static or dynamic based on their behaviours with respect to time. The common types of biological networks are classified based on the components studied at a time (transcriptional regulatory, protein–protein interaction, gene to metabolite and gene regulatory networks), Amongst these, gene regulatory network is dynamic and rest all three are static in nature. Gene regulatory network is discussed in details here.

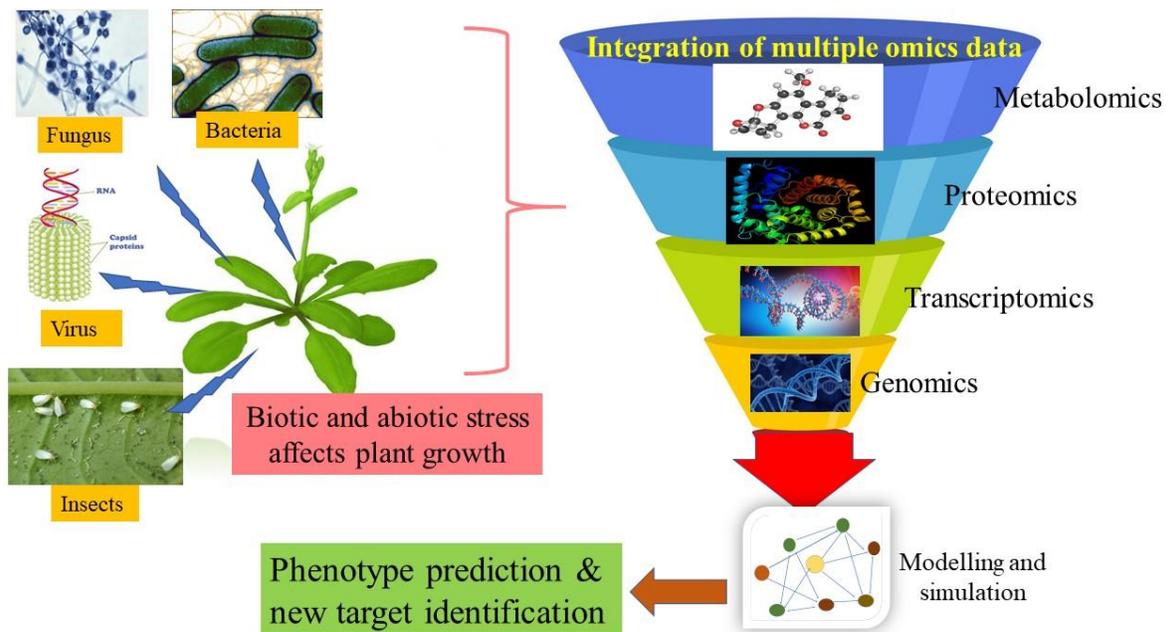
## Gene regulatory network (GRN)

Here nodes represent either a gene, RNA or protein and the edges corresponds to interactions like activation, repression or inhibitory interactions of a biological network. Complex GRN can also consider metabolites and other signalling elements as nodes. GRN interpretation includes the use of data explaining gene regulation to guess the regulatory actions that is present in those nodes. GRN can be non-integrative or integrative based on the different data and methods used. Non integrative methods focuses at gathering information, from an individual data set obtained under a particular experimental condition, statistical links between nodes can be understood as regulation or co-expression relationships (Lavarenne *et al.*, 2018).

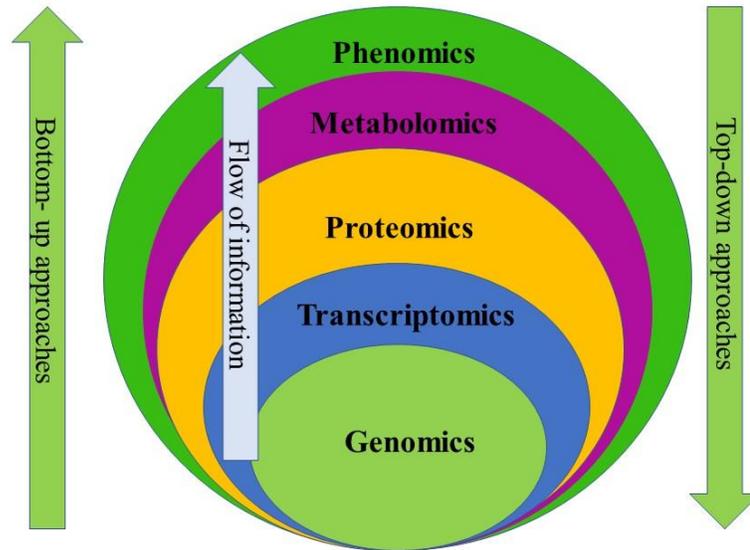
The first approaches in GRN analysis after collection of individual data is to visualize the

biological network by using suitable computational tools that describes positions of nodes with reference to their regulatory node. This is done by using software like Cytoscape. A cluster or clique in GRN means a highly interconnected set of nodes forming a network. These cliques are generally function together in any particular biological process or subprocess. Identification of cliques in biological networks is a difficult job but, software like OSLOM, InfoMap and CONCLUS can easily detect based on clique size and network topology. Qualitative and quantitative measures are possibly redundant markers of special topological explanations indicating the key role of certain nodes in the GRN. Hence these nodes can be our first level of target to investigate. But there is a limitation here in topological studies that only examine the network structure without considering progressive temporal variations. Hence the network structure needs to be analysed at dynamic level.

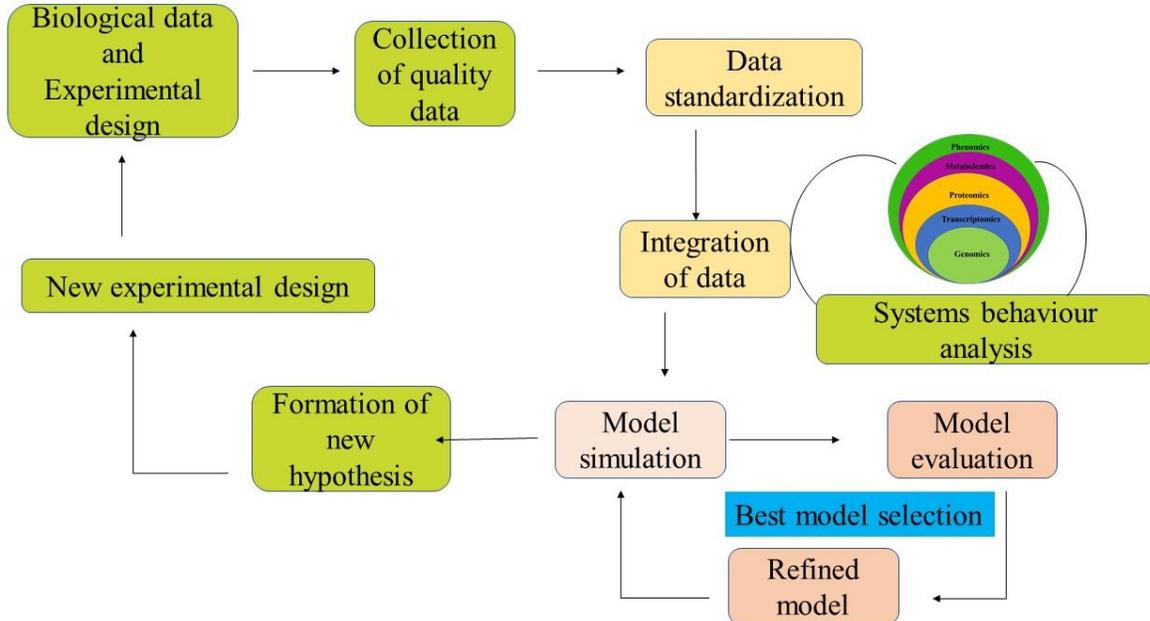
**Fig.1** A comprehensive view of plant systems biology



**Fig.2** Schematic representation of approaches in systems biology



**Fig.3** Steps in plant systems biology



The GRN model need to be examined over time period and progressive plant growth, this temporal study indicates the dynamic level of GRN. Each member nodes of GRN need to be studied for their activity level with respect to

time. Hence this level of studies can help in identification of temporal influential genes or nodes. Based on these influential nodes we can predict the modifier nodes that is indirectly our next level target candidate

node. Regulators and interactions in an GRN can be assessed through systemic testing of models. For identification of targets in GRN engineering techniques like virtual mutation can be applied. At last a step wise understanding of how systems biology operates is given in figure 3.

In conclusions, multi-disciplinary effort across entire crop science is possible the greatest path towards systems biology. It offers a complete sense of plant systems. Holistic systems biology approaches provide a complete sense of the plant systems. It allows plant biologist to comprehend the structural consistency of plants, their management (Hong *et al.*, 2019). This is how our trait of interest can be examined by developing suitable models in systems-biology. They offer hints to identify the biological factors that are employing major impact on these biological systems, even in cases where feedback loops or gene redundancies rejects conventional gene search approaches. Again, it helps in identifying new targets for plant breeding with the help of virtual mutations to these biological networks. Besides having many favours in this omics era, it faces some challenges like it required quality pure data that is standardised too. It is necessary to develop more temporal and dynamic models that can fit throughout the growth stages. More analytical platforms and more studies are required on network topology to completely understand this. Although these biological models could not accurately mimic the biological systems, still they are capable enough to explain the biological complexity and can provide new key targets from these models. Crop systems biology appears to be a more confident approach to realize real-world targets.

## References

Aggarwal, K. and Lee, H. K. (2003). Functional genomics and proteomics

as a foundation for systems biology. Briefings in Functional Genomics, 2, 175-184.

Coruzzi, G. M., Burga, A. R., Katari, M. S. and Gutiérrez, R. A. (2018). Systems biology: principles and applications in plant research. Annual Plant Reviews online, 3-40.

De Costa, W. A. J. M. (2018). Raising crop yields: The missing links from molecular biology to plant breeding. Ceylon Journal of Science, 47, 213-220.

Falter-Braun, P., Brady, S., Gutiérrez, R. A., Coruzzi, G. M. and Krouk, G. (2019). iPlant Systems Biology (iPSB): An International Network Hub in the Plant Community. Molecular plant, 12, 727-730.

Fiehn, O. (2002). Metabolomics—the link between genotypes and phenotypes. In Functional genomics. Springer, 155-177.

Hong, W. J., Kim, Y. J., Chandran, A. K. N., and Jung, K. H. (2019). Infrastructures of systems biology that facilitate functional genomic study in rice. Rice, 12, 15.

Keurentjes, J. J., Fu, J., De Vos, C. R., Lommen, A., Hall, R. D., Bino, R. J. and Koornneef, M. (2006). The genetics of plant metabolism. Nature genetics, 38, 842-849.

Kliebenstein, D. J. (2010). Systems biology uncovers the foundation of natural genetic diversity. Plant Physiology, 152, 480-486.

Lavarenne, J., Guyomarc'h, S., Sallaud, C., Gantet, P. and Lucas, M. (2018). The spring of systems biology-driven breeding. Trends in plant science, 23(8), 706-720.

Libault, M., Pingault, L., Zogli, P. and Schiefelbein, J. (2017). Plant systems biology at the single-cell level. Trends in Plant Science, 22, 949-960.

- Oliver, S. G., Winson, M. K., Kell, D. B. and Baganz, F. (1998). Systematic functional analysis of the yeast genome. *Trends in biotechnology*, 16, 373-378.
- Sauer, U., Heinemann, M. and Zamboni, N. (2007). Getting closer to the whole picture. *Science*, 316, 550-551.
- Sheth, B. P. and Thaker, V. S. (2014). Plant systems biology: insights, advances and challenges. *Planta*, 240, 33-54.
- Wang, L., Li, P. and Brutnell, T. P. (2010). Exploring plant transcriptomes using ultra high-throughput sequencing. *Briefings in functional genomics*, 9, 118-128.
- Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F. and Williams, K. L. (1996). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and genetic engineering reviews*, 13, 19-50.
- Yin, X. and Struik, P. C. (2007). Crop systems biology. Scale and complexity in plant systems research: gene-plant-crop relations. Springer, 63-73.
- Yuan, J. S., Galbraith, D. W., Dai, S. Y., Griffin, P. and Stewart Jr, C. N. (2008). Plant systems biology comes of age. *Trends in plant science*, 13, 165-171.

**How to cite this article:**

Manoranjan Senapati, Sai S. Dash, Biswajit Lenka and Debadatta Panda. 2020. An Overview on Systems Biology Driven Plant Breeding. *Int.J.Curr.Microbiol.App.Sci*. 9(07): 1956-1964. doi: <https://doi.org/10.20546/ijcmas.2020.907.223>