Journal of AgriSearch **5**(2):83-89
An Open Access International Peer Reviewed Quarterly

ISSN : 2348-8808 (Print), 2348-8867 (Online)
https://doi.org/10.21921/jas.5.2.1

Balanced gene expression: network of genes in legumes

JESHIMA KHAN YASIN*, SAKSHI CHAUDHARY, BHARAT KUMAR MISHRA[†], NIDHI VERMA[2], ANIL KUMAR SINGH[1] AND NEETA SINGH

Division of Genomic Resources, ICAR-NBPGR, PUSA Campus, New Delhi, India

ARTICLE INFO

Received on	:	15-11-2017
Accepted on	:	04-06-2018
Published online	:	07-06-2018

ABSTRACT

Handful of theories explained modes of gene fixation and drift by contradicting one after the other hypothesis. Detailed analyses of these hypothetical or established theories illustrate that, all these theories holds good at one or the other circumstances existing in a living cell and whole organism as well, at a functional balance of gene expression. This theory explains that, the earlier reported neutral theories based on selection acquiesce as it tries to conquer an expression balance of genes. The meagre existence of a deleterious gene is being even-handed by suppression of negative gene expression into a functional protein. At an exacting time there is a planned balance attained by the individual cell to survive; otherwise abscission, apoptosis or necrosis occurs, leading to death of a cell or organ or the organism. This balance could be of ions or cytosolic pH or the impact of total proteins imparting cellular quiescence. Mutations at active domains impede cellular processes by blocking a network of genes and finally protein pool of any functional cell. Balancing pH of any cell at a time is a final outcome of its acidic or alkaline components being regulated by gene expression at a particular time. Whether to make the cell survive or to initiate cell death is decided by pH flux, either to protect the adjacent cells from spreading pathogen invasion or any other physiological stress. Variations in such traits are obtained through evolutionary phenomenon which is an underlying, powerful, fundamental and important concept of life.

Speculations in genomics and proteomics ponder upon network of functional proteins encoding genes in manifesting phenotype of an intact organism through proteome. Though it is counterintuitive, many tried a direct way of creating a mutant that has acquired changes or deletions in their nucleotide sequences or knockouts to find its missing functions during its absence. To establish the fact that alterations in a single gene cannot be stable and it needs to balance another set of genes the present investigation was carried out.

Keywords: Fitness; gene expression; genetic drift; legumes; pH balance; survival

INTRODUCTION

The neutral theory of evolution claims that the great majority of evolutionary changes at the molecular (DNA) level are not caused by Darwinian selection but by random fixation of selectively neutral or nearly neutral mutants, and we propose this to be the neutrality attained within a single cell[5]. The Neutral theory of evolution states that the silent or synonymous mutations occur when the change of a single DNA nucleotide (A, T, C, G) within a gene does not affect the sequence of amino acids that make up the gene's protein. Here we emphasise that, such a change is warranted by the organism to maintain its intracellular ionic or pH balance.

We already reported that the survival of any cell is based on its ability to maintain its pH balance (Yasin, 2013, Yasin, 2015, Yasin *et al*., 2016). Balancing pH of any cell at a time is a final outcome of its acidic or alkaline components being regulated by gene expression at a particular time. Whether to make the cell survive or to initiate cell death is decided by pH flux, either to protect the adjacent cells from spreading pathogen invasion or any other physiological stress (Kurkdjian and Guern, 1989)

The internal pH of any cell at a particular time is being maintained by the flow of ions impacted by many transporter genes as well as hormones and its network with intra-cellular transport (Haynes,1990) Conditions of higher or lower cytosolic concentration of Ca^{2+} ions were found to be toxic in many organisms (Jones and Lunt. 1967, Rengel,1992) . Such toxic conditions arise due to imbalance in pH flux to an extreme condition, where the cell is crippled to regain its balance in expelling the excess ions. Many plants growing in high salinity condition have larger vacuoles with higher level of salts. This could be due to a mechanism of excluding the salts to maintain the pH balance of living cytoplasm (Yasin, 2015).

When an organism is under stress (biotic/abiotic) condition then the consolidated pH of an individual affected cell changes and the organism tries to regain its balance by intracellular transport. If the cell is incapacitated in grappling intracellular neutrality, then the cell starts signalling apoptosis or individual cell death (Foyer and Noctor, 2005).

[1]ICAR-Research Complex for Eastern Region, Patna- 800 014, India
[2]Indian Council of Agricultural Research, New Delhi-110 012, India
[†]Present address: Department of Biology, University of Alabama at Birmingham, AL, USA AL 35294-1170
*Corresponding Author Email : Yasin.Jeshima@icar.gov.in

Hence, the impairment does not spread to the adjacent cell and the organism is protected. Accordingly, the number of cells per unit area and size gets reduced under extreme stress to accelerate a structural compaction as a tolerance mechanism for adaptation (Yasin *et al.*, 2012).

The earlier theories of natural selection (Ospovat, 1995) (1838-1859); neo-darwinian selectionist theory (1918-1932) (Nei, 2005); Neutral theory (non-darwinian- 1969-1983) (Kimura, 1983) and nearly neutral theory (1972-2002) (Ohta, 1992) assert that majority of protein and DNA polymorphisms are selectively neutral and we propose that this selectivity is based on its ability to bring back its intracellular balance. If a mutation is selected and is partially fixed in few generations, it may get drifted over generations due to its inability to adjust under changing cellular environment or stress condition. Drift is programmed at proteins level.

Proteins form enormously sophisticated chemical devices with dependent functions on detailed chemical properties of their surfaces. Binding sites for ligands are formed as surface cavities in which precisely positioned amino acid side chains are brought together by protein folding. Proteins reversibly alter their shape when ligands bind to their surface. In the same way, unreactive amino acid side chains can be activated to make and break covalent bonds. Allosteric changes affect the binding of ligand and linkage between two ligand-binding sites provides a crucial mechanism for regulating cell processes (Meyer and Peters, 2003).

Highly efficient protein machines are formed by incorporating mutations into larger assemblies in which the allosteric modifications are coordinated. Feedback regulations of metabolic pathways affected by small molecules played by either inhibiting or activating enzymes of a pathway. Enzymes controlled in this way generally form symmetric assemblies, allowing cooperative conformational changes to create a steep response to changes in the concentrations of regulatory ligands. Three dimensional changes in protein structure can be determined in unidirectional manner by the expenditure of chemical energy. By coupling allosteric modifications with ATP hydrolysis, proteins generate mechanical force in intracellular locomotion.

The three-dimensional structure of proteins revealed impact of small local changes caused by nucleoside triphosphate hydrolysis is amplified to create major changes elsewhere by the proteins. By such means, these proteins can serve as input–output devices that transmit information, as assembly factors, as motors or as membrane-bound pumps and enzymes. Enzymes are catalytic proteins that greatly speed up reaction rates by binding the high-energy transition states for a specific reaction path; they also perform acid and base catalysis simultaneously. The rates of enzyme reactions are often so fast that they are limited only by diffusion; rates can be further increased if enzymes that act sequentially on a substrate are coupled into a single multienzyme complex, or if the enzymes and their substrates are confined to the same compartment of the cell. Such altered molecular machines if adapted will perform important reactions in cells otherwise drifted.

Through million years of evolutionary time, the amino acid sequence of each protein has been selected not only for the conformation that it adopts but also for an ability to fold rapidly, as its polypeptide chain spins out of the ribosome starting from the N-terminal end. Experiments have demonstrated that a protein domain in a multi-domain protein emerges from the ribosome; it forms a compact structure within few seconds that contains most of the final secondary structure with α helices and β sheets aligned in roughly the right way (Hartl and Hartl, 2002). For protein domains, that unusually opens a flexible structure, which is called a molten globule, is the starting point in which many side-chain adjustments occur that eventually form functional tertiary structure.

Nevertheless, because it takes several minutes to synthesize a protein of average size, a great deal of the folding process is complete by the time the ribosome releases the C-terminal end of a protein. Such kind of conserved sequences without changes in active sites of Cytochrome-c were exchangeable among different mammalian species (Margoliash, 1963). Thus, amino acid substitutions outside the active sites are mostly neutral or nearly neutral as they are not impacting changes at biochemical processes.

MATERIALS AND METHODS

Set of 879 genes were identified from reported *Arabidopsis thaliana* genome as either directly or indirectly contributing to ion transport and pH balance of intracellular cytoplasm and their co-expression details were mined from databases (http://stitch.embl.de database) and evidence based network is drawn using co-expression data.

Sets of contrasting expression values were generated from three different legumes from RANseq data available from public databases. Each set differ in their source sample from root and shoot of redgram, from different varieties of azhudki beans and control and diseased sample from soybean.

Heat map is generated from the data generated by genomic microarray sets from array of genes. The Co-expression data was converted into a matrix for heat map generation and fed to Heatmapper which is a freely available web server for interactive visualization of data in the form of heat maps through graphical interface. Heatmap shows results of hierarchical clustering analysis (Euclidean Distance, Complete Linkage) which clusters the similarity of contigs and samples by expression level (normalized value) from the significant list.

RESULTS AND DISCUSSION

Comparative expression analyses data revealed that the major set of pH balance genes are unchanged except the diseased sample where the altered pH balance indicated its susceptibility. We are representing the graphical

interpretation of expression data in which individual values contained in the matrix are represented as colour codes in columns of a heat map might indicate different samples while the rows represent different genes. For heat map we have taken three different samples in which one is root and shoot sample another is disease and control and last one is from two different varieties. Up regulated genes are represented by green colour and down regulated genes by red colour. In these samples expression of genes are balanced. Levels of over expressed genes are almost equivalent to down regulated genes. This evidence is suggesting the existence of a balance in relationship between up regulated and down regulated genes. Such relationships will be important to gain a better understanding of the role of regulatory balance in evolution.

Plants has its own pH identity, with some plants being more acidic and others more alkaline, the cytosol and vacuole of different intra cellular regions has a pH identity which is determined by a variety of factors altered by gene expression. The growth of plant cells and organs are affected by light and humidity, which along with pH and temperature are the most important factors that influence growth. pH can be expressed in terms of the more elementary quantity like chemical potential of protons mainly affected by transporters. The highly informative network of genes as depicted in fig. 2 explains our results in terms of balance in co-expression of pH contributing genes.

According to the neutral theory, mutations appear at rate μ in each of the 2N copies of a gene, and fix with probability 1/(2N). If all mutations were neutral, the rate at which fixed differences accumulate between divergent populations is predicted to be equal to the per-individual mutation rate, e.g. during errors in DNA replication; both are equal to μ. When the proportion of mutations that are neutral is constant, so is the divergence rate between populations. This provides a rationale for the molecular clock, although the discovery of a molecular clock predated neutral theory (Zuckerkandl and Pauling, 1962 and Zuckerkandl, 1987). The neutral theory posits that the great majority of evolutionary change at the molecular level is caused not by Darwinian selection but by the random fixation of selectively neutral mutations.

Another hypothesis of neutral theory is that most evolutionary changes are the effect of genetic drift acting on neutral alleles (Slatkin, 1987). After mutation, a neutral allele may become more common within the population. Usually, it will be lost, or in rare cases it may become fixed, making the new allele a standard one in the population. This stochastic process is assumed to comply with equations describing random genetic drift by means of accidents of sampling. Many including Darwin, have thought independently of the same idea but it was not until Kimura's forcible advocacy. The proposal of the neutral theory and subsequent pros and cons of this hypothesis were indispensable fuel in the propulsion of population genetics in the 1970s and promoted the establishment of molecular evolution and molecular taxonomy.

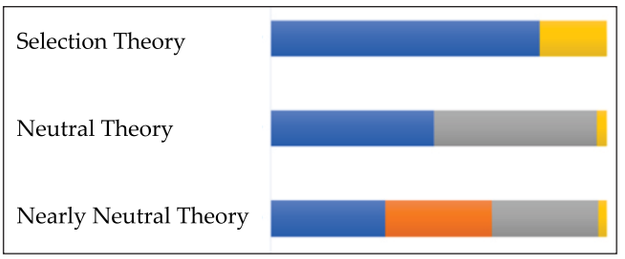
Kimura (Takahata, 1996) favoured the method of diffusion approximations; a mathematical means of describing gene frequency changes caused by stochastic evolutionary forces, such as random genetic drift and applied it to significant problems concerning gene frequency and amino acid sequence data. The nature of both methodology and data in such analyses has recently changed dramatically; the neutral theory has been reformulated and many statistical tests have been devised by the method of gene genealogy, the family relationships among genes. In addition, a wealth of DNA sequence data, sometimes together with detailed information about: the function and structure of molecules, has become available. These advances have permitted us to look at the neutral theory from much deeper view points than ever before.

Population geneticists have often defined neutral mutations using mathematical theory (Ewens, 2012 and Wade and Goodnight, 1998). Fisher (1930) and Wright (1931) showed that if the relative fitnesses (Wij) of genotypes A1A1, A1A2, and A2A2 are given by

W11 = 1, W12 = 1 + s, and W22 = 1 + 2s , respectively, the probability of fixation (u) of a new mutant allele (A2) in the population is

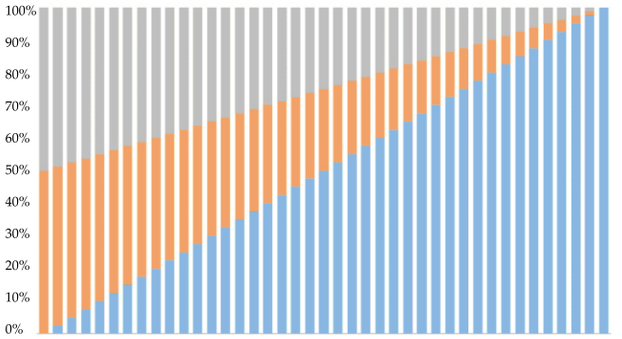
$$u = 2s / (1 - e^{-4N s}) \quad (1)$$

Where, N is the effective population size. Noting that u for N s = 1 is approximately 50 times higher than that for N s = −1, Fisher concluded that natural selection is very effective, because he believed N is of the order of 109. (Fisher did not have the concept of effective population size, which is often much smaller than the actual population size). In other words, even selection coefficients (s) as small as ± 10−9 have significant effects on u. For this reason, he became a pan-selectionist, and there was no need for him to examine the possibility of neutral mutations. Interestingly, Kimura (Kimura, 1967) (supplementary information 1b) used essentially the same definition of neutrality ($|2N s| \leq 1$). However, because he knew the effective population size, he believed this definition of neutrality would be sufficient (Nei, 1987). Furthermore, the above definition was a mathematical formality for Kimura, and he was actually interested in neutral or nearly neutral mutations in the biological sense (Nei *et al.*, 2010).



Supplementary information 1: Earlier theories as explained before elsewhere

Deleterious Nearly Neutral
Neutral Advantageous



Supplementary information 2: Balancing expression of intracellular neutrality contributing genes

Neutral genes Deleterious genes Favorable genes

The neutral theory of molecular evolution has been controversial for a long time, but the general pattern of molecular or genomic evolution is broadly consistent with the expectation from the neutral theory.

We tried to explain the neutrality within each cell as a balance in expression of genes. If the expression of disadvantageous genes increases in a tightly linked network of genes (Fig. 1, fig. 2 and supplementary information 1c, 2a) the expression of favourable genes neutralises the condition enabling the cell to survive.

Because of its simplicity and value as a basis for prediction, Kimura's neutral theory has been tested extensively. Whether the theory is accepted or rejected by particular tests, it has

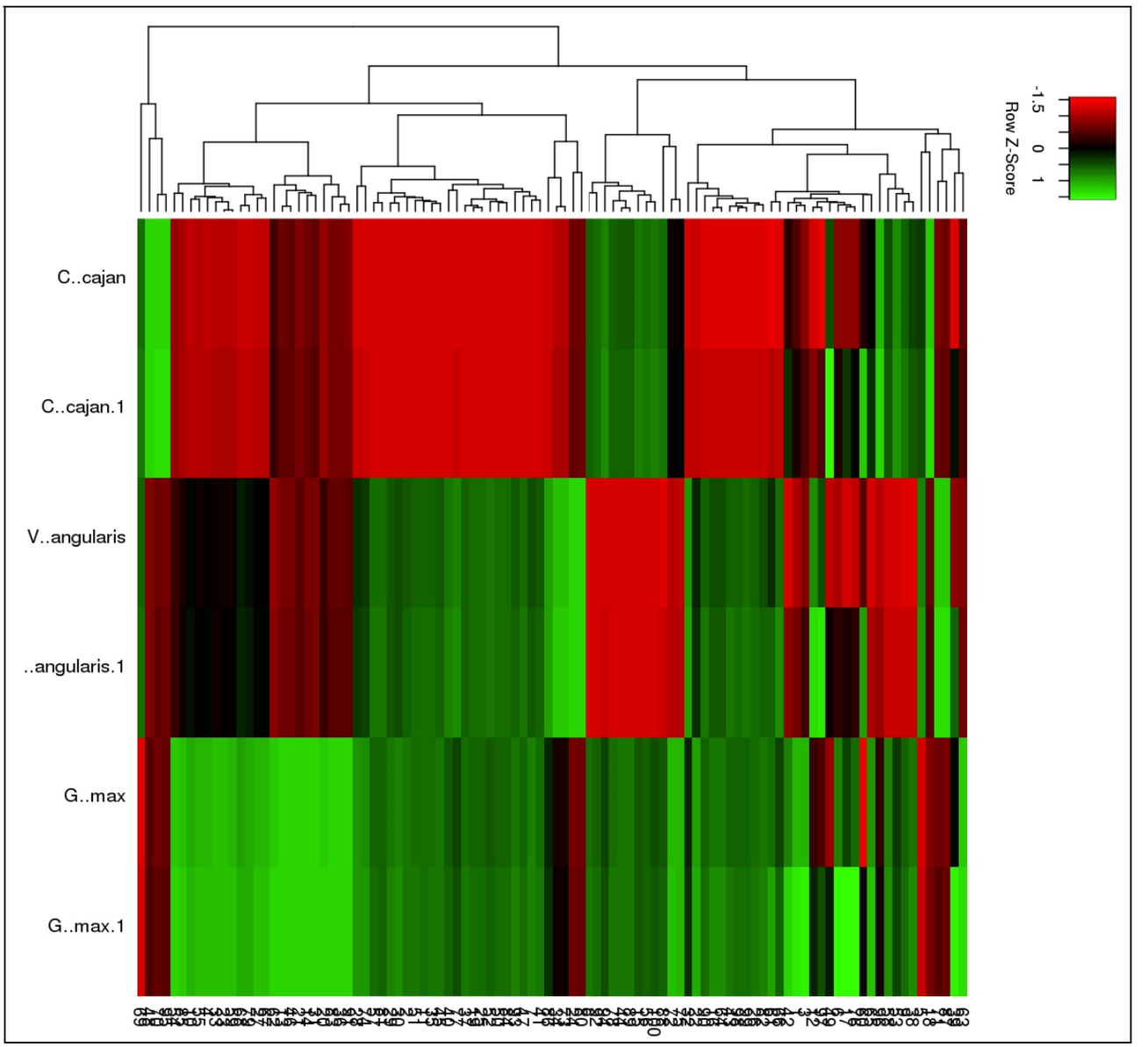


Fig.1: Comparative expression of selected intracellular neutrality contributing genes from legumes

without question served as a fundamental guiding principle in the study of molecular evolution. Although there is agreement about the importance of random genetic drift in molecular evolution, there is obvious disagreement about the relative roles of other evolutionary forces, especially natural selection. We explained here that, the natural selection starts at individual cell and programmed at protein level. Convincing arguments for the theory of S and MHC molecule evolution have been possible only by knowledge, albeit imperfect, of their function and structure which could be related directly to the causes of natural selection.

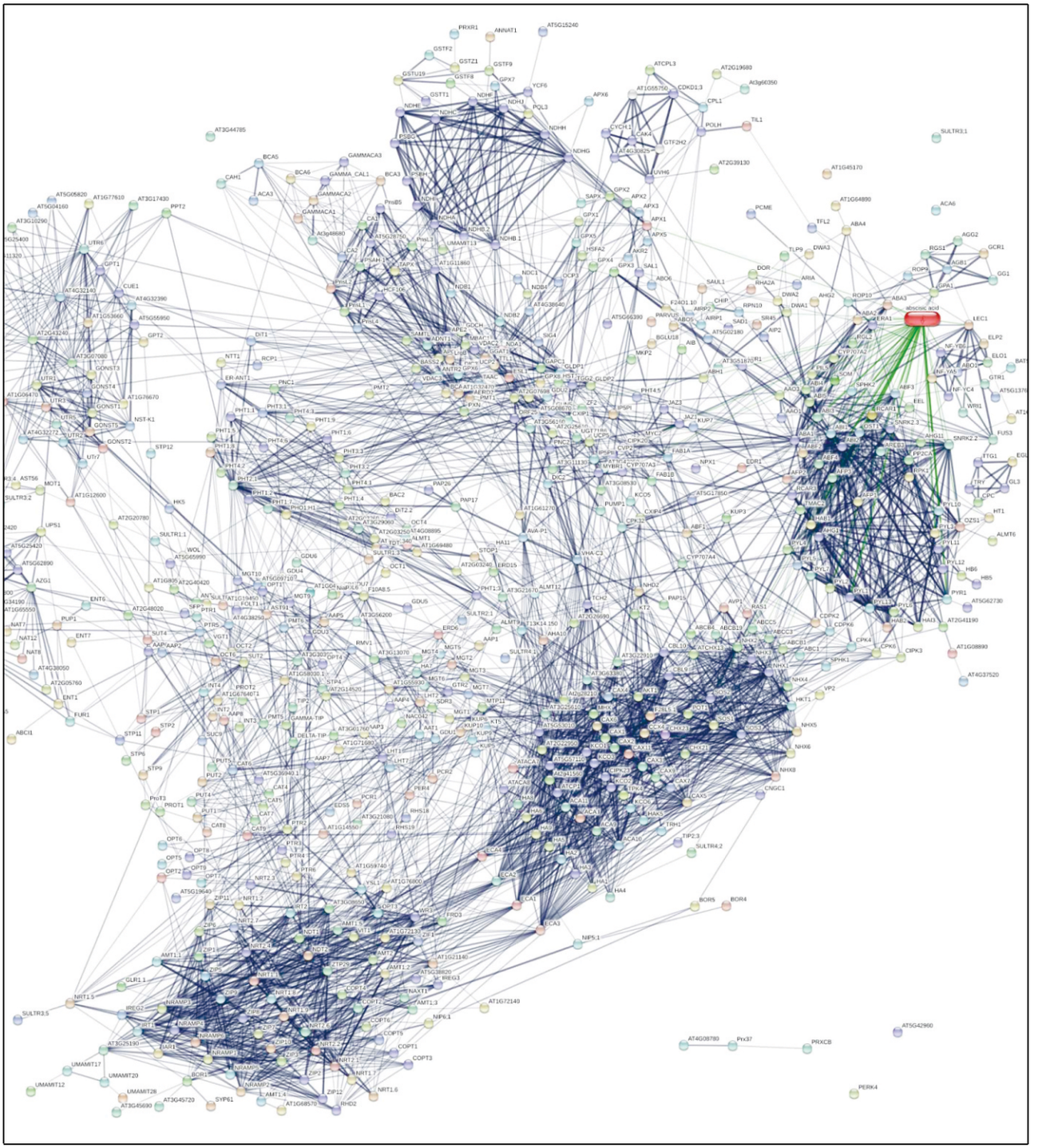


Fig.2: Predicted network of genes contributing to intracellular neutrality in *Arabidopsis thaliana*

Clues to gene function can be obtained by gene expression analyses of a cell or organism. Determining the pattern and timing of gene expression can be accomplished by replacing the coding portion of the gene under study with a reporter gene. In most cases, the expression of the reporter gene is then monitored by tracking the fluorescence or enzymatic activity of its product (Elowitz *et al.*, 2002). From the earlier reports, if we make substitutions;

The rate of mutation or substitution is

$$H_e = \frac{4N_e\mu}{(1+4N_e\mu)} = k\Delta \quad \text{.....(1)}$$

Since $\theta = 4N_e\mu$, by substitution of

$$\frac{\theta}{1+\theta} = k\Delta \quad \text{.....(2)}$$

$$\theta = k\Delta + k\Delta\theta$$

$$\theta(1 - k\Delta) = k\Delta$$

By solving the above equations, equation (3) can be arrived

$$\theta = \frac{k\Delta}{1-k\Delta} \quad \text{.....(3)}$$

Since, $\theta = \frac{S}{a}$, where S is selection co-efficient,

Equation (3) can be re written as

$$\frac{S}{a} = \frac{k\Delta}{1-k\Delta} \quad \text{.....(4)}$$

Sequence of amino acids in a protein determines its functional three-dimensional structure.

Following that, it has been repeatedly observed that protein structure dictates where and how it will act. Nowhere has this been more obvious than in function of enzymes. The shape and structure of proteins is a crucial aspect of gene expression biology and links our understanding of gene expression to the biology of the cell. Interestingly, proteins are not only gene regulators. Regulatory molecules come in the form of RNA also and act on other nucleic acids by changing or disrupting them. One example is the family of riboswitches, ribonucleic acid molecules that form three-dimensional structures that halt or interfere with transcription, given the proper external signal. While primarily concerned with protein molecules that act on DNA and RNA sequences, such as transcription factors and histones, the study of gene expression also focuses on where the cell expression is modulated (Thomas *et al.*, 2002). In fact, the modulation of gene expression can occur in the nucleus, the cytoplasm, or even in the cell membrane due to the impact of proteins on RNA in those cellular sub regions. In a network of genes, the expression of genes is being regulated to achieve a balance within intracellular level.

CONCLUSIONS

The altered proteins and intermediates formed along the way would aggregate and propelled off-pathway dead ends without intervention of chaperone resulting in degradation. The supplementary information 1 and 2 illustrates more about its gene expression balance in explaining earlier explained. Hence, introduction of a single gene or over expression of single gene cannot result in better phenotype. It's because of balance in a network of gene expression or its co-expression plays a major role.

ACKNOWLEDGEMENTS

We are grateful to Director, ICAR-NBPGR for providing facilities to conduct this research.

Funding

This work was supported Indian Council of Agricultural Research.

Availability of data

The datasets processed in this study are bundled and made available with the manuscript as supplementary information.

Author contributions

YJK designed and planned the research, BKM, YJK and SC carried out *in-silico* analyses, and developed graphical representations; YJK and AKS compiled the results, wrote the manuscript, and communicated.

Declaration

Additional Information

Supplementary information accompanies this paper will be available from online version.

Competing financial interests

The authors declare no competing financial interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

REFERENCES

Elowitz MB, Levine AJ, Siggia ED and Swain PS. 2002. Stochastic gene expression in a single cell. *Science* **297**(5584):1183- 1186.

Ewens WJ.2012. Mathematical population genetics 1: theoretical introduction. *Springer Science and Business Media* **27**.

Fisher RA.1930. The Genetical Theory of Natural Selection. *Oxford: Clarendon Press*.

Foyer CH and Noctor G.2005. Oxidant and antioxidant signalling in plants: a re-evaluation of the concept of oxidative stress in a physiological context. *Plant Cell & Environment* **28**(8):1056-1071.

Hartl FU and Hartl M. 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295**(5561):**1852-1858.

Haynes RJ.1990. Active ion uptake and maintenance of cation-anion balance: A critical examination of their role in regulating rhizosphere pH. *Plant and Soil* **126**(2)**:**247-264

Jones RW and Lunt OR.1967. The function of calcium in plants. *The Botanical Review* **33**(4)**:**407-426.

Kimura M.1967. Molecular evolutionary clock and the neutral theory. *JMO/ Evol.*, **26**:24.

Kimura M.1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**(5608): 275-276.

Kimura M.1983. The neutral theory of molecular evolution. *Cambridge University Press.*

Kurkdjian A and Guern J.1989. Intracellular pH: measurement and importance in cell activity. *Annual review of plant biology* **40**(1): 271-303.

Margoliash E.1963. Primary structure and evolution of cytochrome c. *Proceedings of the National Academy of Sciences* **50**: 672-679.

Meyer B and Peters T. 2003. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie International Edition* **42**(8): 864-890.

Nei M. 1987. Molecular evolutionary genetics. *Columbia university press*

Nei M. 2005. Selectionism and neutralism in molecular evolution. *Molecular biology and evolution* **22**(12): 2318-2342.

Nei M, Suzuki Y and Nozawa M. 2010. The neutral theory of molecular evolution in the genomic era. *Annual review of genomics and human genetics* **11**: 265-289.

Ohta T.1992. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* **23**(1): 263-286.

Ospovat D.1995. The development of Darwin's theory: Natural history, natural theology, and natural selection. *Cambridge University Press*, 1838-1859. (1995)

Rengel Z. 1992. The role of calcium in salt toxicity. *Plant, Cell & Environment* **15**(6)**:** 625-632.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236**(4803)**:787-793.

Takahata N. 1996. Neutral theory of molecular evolution. *Current opinion in genetics and development* **6**:767-772.

Thomas C.H. *et al*. Engineering gene expression and protein synthesis by modulation of nuclear shape. *Proceedings of the National Academy of Sciences*, **99**(4),1972-1977 (2002) Thomas CH, Collier JH, Sfeir CS and Healy KE.2002. Engineering gene expression and protein synthesis by modulation of nuclear shape. *Proceedings of the National Academy of Sciences.* **99**(4):1972-7.

Wade MJ and Goodnight CJ.1998. Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. *Evolution* **52**(6):1537-1553.

Wright S. 1931. Evolution in mendelian populations. *Genetics* **16**(2):97-159.

Yasin JK, Singh N, Bhogal I and Mishra BK. 2016.Carbonic anhydrase genes network: Key role players in pH flux and abiotic stress tolerance. *Journal of AgriSearch* **3**(4):195-198.

Yasin JK.2013. Cellular transport and cell singnaling in Diplotaxis and Arabidopsis. *Science*. http://www.sciencemag.org/content/342/6155/175.2.full

Yasin JK.2015. Intra cellular pH flux and cyclosis in plant cells under abiotic stress. *Journal of AgriSearch* **2**(2):150-151.

Yasin JK.*et al*.2012. Structural compaction: Mechanism of acid tolerance in moisture stress responsive accessions of horse gram. The 8TH International symposium on "Plant soil interactions at low pH". Mathematical population genetics 1: theoretical introduction. *Springer Science & Business Media* **27**.

Zuckerkandl E and Pauling LB. 1962. Molecular disease, evolution, and genetic heterogeneity". *Horizons in Biochemistry*. 189**–**225.

Zuckerkandl E. 1987. On the molecular evolutionary clock. *Journal of Molecular Evolution* **26**:34-46.

Citation:

Yasin JK, Chaudhary S, Mishra BK, Verma N, Singh NAK and Singh N.2018. Balanced gene expression: network of genes in legumes. *Journal of AgriSearch* **5** (2): 83-89