O-SF-01

Dissecting the Transcriptional Regulatory Network to Reveal Beneficial Network Topology for Cassava Brown Streak Virus Resistance

Thanakorn Jaemthaworn[1], Treenut Saithong[1,2] and Saowalak Kalapanuluk[1,2*]

[1]*Bioinformatics and Systems biology program, School of Bioresources and Technology, School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand*

[2]*Systems biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand*

**Corresponding author. E-mail: saowalak.kal@kmutt.ac.th*
DOI:

ABSTRACT

Cassava Brown Streak Virus (CBSV) represents a major threat to cassava production, especially in endemic areas, and can lead to substantial storage root yield and quality losses. The developed resistant cultivars have often performed inconsistently, especially in differently environments largely due to limited knowledge of the underlying resistance mechanisms, and the roles of gene cooperation and gene-environment interaction. This study aimed to propose a robust structural transcriptional regulatory network (TRN) of CBSV resistance in cassava. We performed a topological analysis of reconstructed conditional TRNs, namely TRN of CBSV resistance in control (TRN-RC) and treatment (TRN-RT) conditions and that of susceptibility in control (TRN-SC) and in treatment (TRN-ST) conditions. The reconstruction was performed via integrative systems biology approach, combining time-series gene expression datasets of resistant and susceptible cultivars in naïve and infection conditions with the TRN retrieved from PlantRegMap. The results demonstrated that the global network properties, network dimension, were not different among traits, but the local clustering coefficient of TRN-RC was significantly higher than that of TRN-SC (p-value $=1.84\times10^{-9}$). The higher local clustering coefficient found in TRN-RC could be related to its robustness with respect to the regulation of CBSV resistance in cassava.

Moreover, the reconstructed TRNs of CBSV resistance in control and treatment conditions contained more clustering coefficient in heat shock proteins (HSPs) regulatory system, which are important for plant protection against viral infections, than those for susceptible traits in similar conditions. Our findings would contribute to precision CBSV resistance-breeding in cassava, through the use of bioinformatics screening tools for cultivar improvement. Moreover, the knowledge gained could be applied for other cassava viral diseases such as the cassava mosaic disease (CMD), which is currently affecting cassava production in Thailand.

Keywords: Cassava, Cassava brown streak virus, Transcriptional regulatory network, Topological analysis

INTRODUCTION

Cassava (*Manihot esculenta* Crantz) is an important staple crop and a major source of carbohydrates for a billion people around the world, especially in Africa and South East Asia. It is relatively tolerant to drought and poor soil conditions (Hillocks and Maruthi, 2015), and even grows in extreme environments were most crops would fail, which makes it an important crop for attaining food security. Among the several challenges facing cassava production and yield improvement efforts are the threats posed by pathogens such as bacteria, fungi, and viruses (Hillocks and Maruthi, 2015; Tomlinson et al., 2018). Cassava brown streak virus (CBSV) is one of the major pathogens that affects cassava production. Cassava brown streak virus, a member of the family *Potyviridae*, is transmitted by the use of infected stem cuttings and by vectors such as white fly (*Bemisia tabaci*). Characterized by symptoms such as chlorosis of leaves and the appearance of brown streaks on stems and brown necrotic lesions in storage roots (Kawuki et al., 2016), CBSV can lead to substantial storage root yield and quality losses. After the outbreak of CBSV in Africa, approximately 20-60 percent of the produce was un-marketable due to necrosis (Mangana, 2003; MtundaI et al., 2003). Despite the successes recorded in developing resistant cultivars, such as 'Namikonga' (Tomlinson et al., 2018), that demonstrate different levels of tolerance, the consistency of these in different environmental conditions has been questioned. Moreover, mechanisms of CBSV resistance have not fully been understood.

Nowadays, the emergence of high throughput technologies allows us to study resistance mechanisms of plants within a short period of time.

Analysis of genome-wide gene expression data of CBSV resistant (Namikonga) and susceptible (Albert) cultivars under viral infection demonstrated that nucleotide binding-leucine rich repeat genes (NB-LRR) and heat shock protein genes (HSP) were differentially expressed at the early stage of infection, compared with time zero (before infection), only in the resistant cultivar. The NB-LRR has been proposed to be involved in the regulation of HSPs in response to CBSV infection (Amuge et al., 2017). Moreover, another study on pea seed-borne mosaic virus, which is one of the potyviruses, found that down regulation of HSP genes was related to the viral replication and virulence (Aranda et al., 2002). Thus, in-depth understanding of the regulation mechanisms and cooperation among genes that are involved in CBSV resistance is crucial for unlocking the mystery of plant-pathogen interaction and for developing new resistant cultivars.

To fill this knowledge gap, study of a robust transcriptional regulatory network (TRN) of CBSV resistance, via systems biology approach, is crucial. Robustness is a property of biological systems to withstand environmental perturbations (Carlson and Doyle, 2002; Kitano, 2004). One way to investigate the robustness of a biological system is by topological analysis of the transcriptional regulatory network. For example, a scale-free network (whose node degree distribution follows a power law) containing a few numbers of high degree nodes—hub genes, and a large number of low degree nodes—peripheral genes (Wu et al., 2017), can maintain their main functions using remaining hub genes when perturbation occurs (Barábsi and Albert, 1999; Jeong et al., 2001; Barabási and Oltvai, 2004). To gain insights into resistance mechanisms in cassava against CBSV, we constructed four TRNs of susceptible and resistant traits under control and treatment (CBSV) conditions, namely TRN-RC, TRN-RT, TRN-SC, and TRN-ST and then investigated their topological properties, including number of nodes, number of edges, network diameter, clustering coefficient and characteristic path length. This study enhances our understanding of how cassava copes with CBSV and would facilitate precision breeding for resistant cultivars.

MATERIALS AND METHODS

Materials

Transcriptome data. Cassava transcriptome datasets for 'Albert'—a susceptible cultivar and 'Namikonga'—a resistant cultivar were retrieved from Amuge et al. (2017). These contained expression patterns of genes in leaf at control and treatment (CBSV) conditions, following the grafting of

CBSV-infected and healthy (mock) plants onto 2-month-old plants of each cultivar. Each condition was sampled using 3 biological replicates at 8 time points: at time zero (sampled only in control condition before grafting), at 6 and 24 hours after grafting (HAG), and at 2, 5, 8, 45, and 54 days after grafting (DAG).

Transcriptional regulatory network (TRN). The PlantRegMap cassava transcriptional regulatory network (TRN-PlantRegMap), which was used for reconstructing the conditional TRNs for CBSV resistance under control and treatment conditions (TRN-RC and TRN-RT) and susceptibility under control and treatment conditions (TRN-SC and TRN-ST) respectively, was retrieved from PlantRegMap in PlantTFDB 4.0 constructed by computational method; motif scanning (Jin et al., 2017). It consists of 31,930 genes, of the total 33,033 genes in the cassava genome (Goodstein et al., 2012), encoding 2,062 transcriptional factors (TF) and 31,924 target genes (TG).

Transcriptome analysis. Raw reads were assessed for quality using FastQC (Andrews, 2010), and low-quality reads were trimmed using Trimmomatic (Bolger et al., 2014). Then, the reads were aligned with the cassava reference genome version 6.1 obtained from the Phytozome database (Goodstein et al., 2012) using STAR aligner (Dobin et al., 2013). Unambiguously aligned reads, with more than 75% alignment (Graaf and Franke, 2018), were counted by HTseq (Anders et al., 2015). Next, counted reads were normalized by the relative log expression method (Love et al., 2014), and the obtained normalized values were divided by the corresponding mean expression values of the genes.

Significant genes identification. To reconstruct a transcriptional regulatory network that is representative of the conditions (TRN-RC, TRN-RT, TRN-SC, and TRN-ST), genes that are active in each condition were identified based on their fluctuation, i.e., the standard deviation (SD) of each gene across the time series (Saithong et al., 2015). Genes with less than 3 expression time points for each condition were excluded from our analysis. Specifically, the SD of each gene was computed for each of the conditions, after which the distribution was plotted. Genes, whose SD were higher than a given threshold (77th – 80th percentile) and were associated with other genes (at least one gene), based on an absolute Pearson correlation coefficient of more than 0.95 at p-value < 0.05, were considered active for that specific condition.

Conditional TRN reconstruction. After the significant genes for each condition were identified, they were mapped to the TRN-PlantRegMap and then visualized using Cytoscape version 3.6.0 (Shannon et al., 2003).

Network topology analysis

The topology analysis of the four reconstructed TRNs of cassava: TRN-RC, TRN-RT, TRN-SC, and TRN-ST, was performed using NetworkAnalyzer (Assenov et al., 2008). Topological parameters such as number of nodes, number of edges, network diameter, clustering coefficient and characteristic path length of the four TRNs were globally compared, using the average properties. The local clustering coefficient (C_i) of the conditional TRNs was computed based on the formula by Watts and Strogatz (1998):

$$C_i = \frac{2\left|e_{jk}: v_j, v_k \in N_i, e_{jk} \in E\right|}{k_i(k_i - 1)}$$

where (e_{jk}) refers to the actual connected degree of first neighbor of node i (N_i) in a graph; $G = (V, E)$ and normalized by all possible connected degree; $k_i(k_i - 1)$ of N_i. In addition, global clustering coefficient; $\bar{C} = \frac{1}{v}\sum_{i=0}^{v} C_i$ shows clustering coefficient of a network while C_i represents a high clustering coefficient of a node. Thus, high C_i represents a high clustering node. Moreover, the local property (particular node property) of the clustering coefficient; C_i was statistically compared by Wilcoxon rank-sum test using R programming.

Identification of heat shock proteins subnetwork on the conditional transcriptional regulatory networks

Functionally annotated cassava HSPs, which were retrieved from Phytozome (Goodstein et al., 2012), were mapped to each of the reconstructed conditional TRN. Subnetworks of the regulatory system were then generated based on the mapped HSPs and their first-neighbor genes, after which the subnetworks were topologically analyzed.

RESULTS

Conditional TRNs of cassava resistance and susceptibility to CBSV

During transcriptome data analysis, a biological replicate of the susceptible cultivar, 'Albert', showed low reads (approximate 12% of its reads passed the set criteria) at 45 DAG in the control condition and was removed from the analysis. Around 4,500 to 5,000 significant genes were identified in each condition; 4,667 in RC, 4,737 in RT, 4,421 in SC, and 5,018 in ST. Approximately 60-70% of the identified significant genes (2,500 - 3,000 genes) were mapped to the global TRN from PlantRegMap to represent the conditions (Figure 1). In total, the gene-gene interaction of conditional TRNs contained 6,000 to 9,500 interactions — 3,174 genes and 9,535 edges in TRN-RC; 3,030 genes and 8,063 edges in TRN-RT; 2,537 genes and 5,994 edges in TRN-SC; and 3,091 genes and 7,941 edges in TRN-ST. Although the total number of significant genes identified for each condition was comparable, the actual number that was mapped to the global TRN varied. For example, of the total 4,667 significant genes identified for the RC condition, 4,279 genes were mapped to the global TRN, of which only 3,174 of them that were connected were retained in TRN-RC.

When comparing the number of edges per node for each condition, we found that the ration was higher in TRN-RC than in TRN-SC (3.0 > 2.4), and it was also slightly higher in TRN-RT than in TRN-ST (2.7 > 2.6). The number of edges per node is indicative of the complexity of the regulation and is probably associated with resistance, as the resistant cultivar had a higher edge to node ratio in control and treatment conditions.

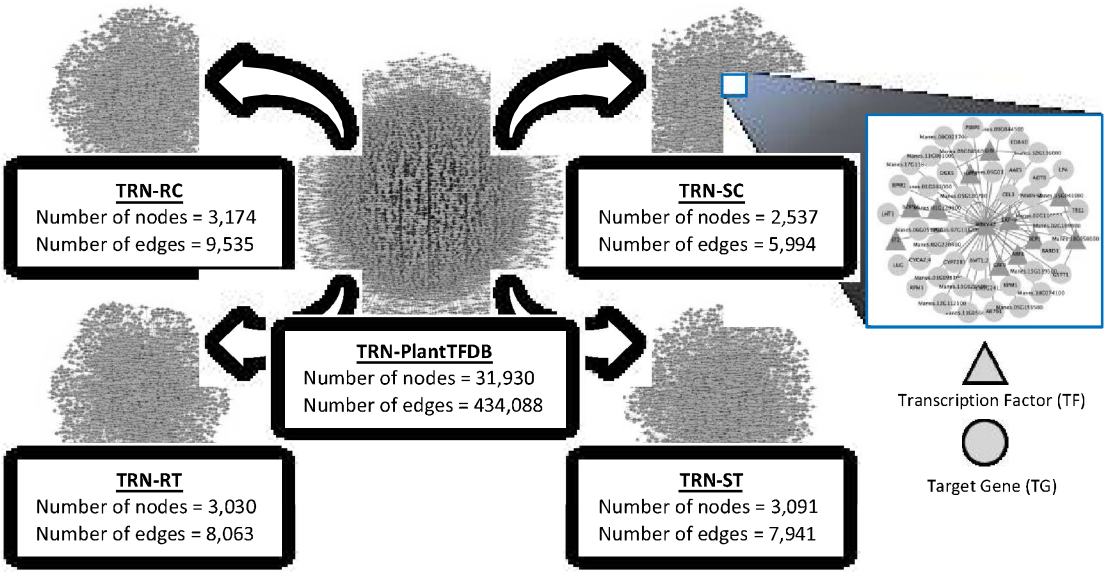

Figure 1. Conditional transcriptional regulatory networks (TRNs) reconstructed by integration of gene expression data to global TRN from PlantTFDB database; TRN of CBSV resistance in control (TRN-RC) and treatment (TRN-RT) conditions and TRN of susceptibility in control (TRN-SC) and in treatment (TRN-ST) conditions.

Impact of cluster-based structure on CBSV resistance and robustness of the system

Topological analysis of the regulatory structure of the reconstructed conditional TRNs revealed that their network diameter, global clustering coefficient, and characteristic path length were comparable, while the number of nodes and edges varied (Table 1). The local clustering coefficient of TRN-RC was significantly different from that of TRN-RT and TRN-SC. Treatment-related differences were found only in the resistant trait. Figure 2 shows the distribution of C_i in all four conditional TRNs. We found that the C_i of TRN-RC was significantly higher than that of TRN-SC (p-value = 1.84×10^{-9}) and TRN-RT (p-value = 1.34×10^{-5}), which may be indicative of its robustness, as C_i is related to the cluster-based structure. The clustering coefficient of the resistant trait was higher than that of the susceptible trait in the control condition. This difference between both traits in naïve condition agrees with the theory of biological robustness by Kitano (2004), which explained that the robust property of systems in naïve conditions is to guard against unpredictable perturbations. The observed differences in conditional TRNs complexity, i.e., the distribution of clustering coefficients, could be explained by variations in node and edge numbers, as the latter has implications for TF regulation of TGs and TF-encoding genes.

Table 1. Global network properties of TRNs

	TRN-PlantTFDB	TRN-RC	TRN-RT	TRN-SC	TRN-ST
Dimension of network					
Number of nodes	31,930	3,174	3,030	2,537	3,091
- Number of TF nodes	2,062	289	243	217	256
Number of edges	434,088	9,535	8,063	5,994	7,941
Network diameter	5	6	6	6	6
Global clustering coefficient ($\bar{C}$)	0.33	0.12	0.10	0.12	0.11
Characteristic path length	2.53	3.16	3.34	3.36	3.34

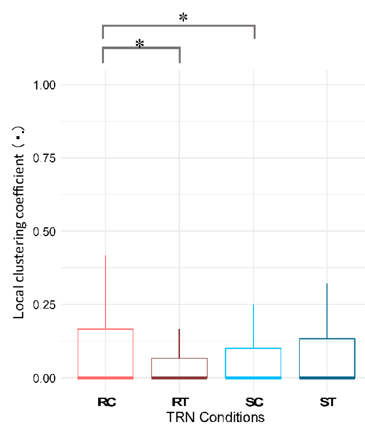

Figure 2. Boxplot of local clustering coefficients (C_i) of TRNs; TRN-RC, TRN-RT, TRN-SC, and TRN-ST

The proposed models of the transcription regulation structure of resistant and susceptible traits were compared with the reconstructed HSP-based subnetworks of the conditional TRNs (figures 3A and 3B). Our proposed models showed a complex regulation of TGs and TF-encoding genes by many TFs.

HSP subnetwork Complexity: a case study on the regulation of heat shock proteins by transcription factors.

Heat shock proteins have been associated with CBSV resistance in cassava, and the genes have been found to be differentially expressed in a resistant cultivar at the early stage of the infection (Amuge et al., 2017). In addition, HSPs have been linked to Potyviral replication in plant cells (Aranda et al., 2002). The proposed models of transcription regulation were reconstructed using HSPs and HSP-associated chaperons as TGs, and their first neighbors as TFs, after which the $\bar{C}$ of each subnetwork was calculated. Results showed that the $\bar{C}$ of the resistant trait was higher than that of the susceptible trait in control and treatment conditions (figure 3).

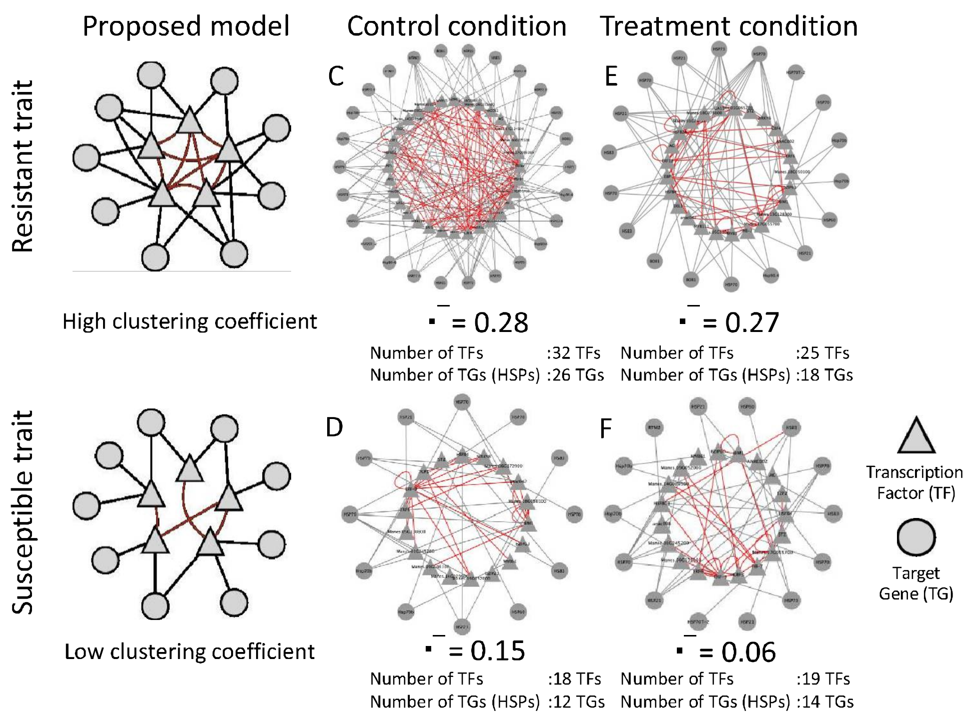

Figure 3. Proposed models of transcriptional regulation underlying CBSV resistance (3A; high clustering coefficient) and susceptibility (3B; low clustering coefficient) in cassava. Triangular nodes represent transcription factors (TFs), and circular nodes represent target genes (TGs). Black edges represent transcriptional regulation between TF and TG and red edges represent transcriptional regulation between TFs and TF-encoding genes. $\bar{C}$ represents global clustering coefficient of each HSP subnetwork.

DISCUSSION

Robustness is a property of organismal systems, in a naïve condition, that enables them to maintain their function under perturbation (Carlson and Doyle, 2002; Kitano, 2004). This property was found in the transcriptional regulatory network structure of the CBSV-resistant cultivar. The cluster-based topology enhances the connectivity of hub and peripheral genes, such that when some of them succumb to perturbation, the functional integrity of the system is preserved. This structure and other related structure such as alternative regulation, redundancy regulation, and the module structure have also been reported in organisms such as *E. coli, S. cerevisiae,* and Arabidopsis (Kurata et al., 2006; Levy and Siegal, 2008; Wu et al., 2009; Benítez and Alvarez-Buylla, 2010). However, some topological structures relating to the cooperation of biological components and their properties, e.g., the feed forward loop—which can reduce noise signals and maintain systems' function in a high noise environment, and the positive feedback loop—which can memorize and react to a few perturbation signals (Freeman, 2000), were not included in this study. The work focused only on the statistical topology, and other approaches such as decomposition, *in silico* simulation, and theoretical modeling can be used to deepen our understanding of transcriptional regulation and gene cooperation.

CONCLUSION

The cluster-based structure is an important structure to improve the robustness of organismal systems. The CBSV resistant cassava, 'Namikonga', was found to contain this structure, as indicated by the high degree of clustering coefficient. The susceptible cultivar, 'Albert', showed a low degree of clustering coefficient. Moreover, the cluster-based structure between both traits was found different in the naïve condition. This study deepens our knowledge of the transcriptional regulation mechanisms underlying CBSV resistance in cassava and the importance of gene cooperation. It also forms a basis for further research towards cassava improvement.

ACKNOWLEDGEMENTS

The authors would like to thank Systems Biology and Bioinformatics Laboratory and Pilot Plant Development and Training Institute (PDTI), King Mongkut's University of Technology Thonburi (KMUTT) for the high-performance computing systems. We are thankful for the partial postgraduate scholarship for TJ from the School of Bioresources and Technology, KMUTT. We would like to thank Mr. Oscar Obidiegwu for language proofreading.

REFERENCES

Amuge, T., Berger, D.K., Katari, M.S., Myburg, A.A., Goldman, S.L. and Ferguson, M.E. 2017. A time series transcriptome analysis of cassava (Manihot esculenta Crantz) varieties challenged with Ugandan cassava brown streak virus. Scientific Reports. 7(1): 1–21.

Anders, S., Pyl, P.T. and Huber, W. 2015. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 31(2): 166–169.

Andrews., S. 2010. FastQC: a quality control tool for high throughput sequence data.

Aranda, M.A., Escaler, M., Wang, D. and Maule, A.J. 2002. Induction of HSP70 and polyubiquitin expression associated with plant virus replication. Proceedings of the National Academy of Sciences. 93(26): 15289–15293.

Assenov, Y., Ramírez, F., Schelhorn, S.E.S.E., Lengauer, T. and Albrecht, M. 2008. Computing topological parameters of biological networks. Bioinformatics. 24(2): 282–284.

Barabási, A. and Oltvai, Z.N. 2004. Network biology: understanding the cell's functional organization. Nature Review Genetics. 5(February):

Barábsi, A.-L. and Albert, R. 1999. Emergence of scaling in random networks. Science. 286(October): 509–513.

Benítez, M. and Alvarez-Buylla, E.R. 2010. Dynamic-module redundancy confers robustness to the gene regulatory network involved in hair patterning of Arabidopsis epidermis. BioSystems. 102(1): 11–15.

Bolger, A.M., Lohse, M. and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30(15): 2114–2120.

Carlson, J.M. and Doyle, J. 2002. Complexity and robustness. Proceedings

of the National Academy of Sciences. 99(Supplement 1): 2538–2545.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 29(1): 15–21.
Freeman, M. 2000. Feedback control of intercellular signalling in development. Nature. 408(6810): 313–319.
Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D.S. 2012. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Research. 40(D1): 1178–1186.
Graaf, A. Van Der and Franke, L. 2018. Gene co-expression analysis for functional classification and gene – disease predictions. 19(September 2016): 575–592.
Hillocks, R.J. and Maruthi, M.N. 2015. Post-harvest impact of cassava brown streak disease in four countries in eastern Africa. Food Chain. 5(1–2): 116–122.
Jeong, H., Mason, S.P., Barabási, A.L. and Oltvai, Z.N. 2001. Lethality and centrality in protein networks. Nature. 411(6833): 41–42.
Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J. and Gao, G. 2017. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research. 45(D1): D1040–D1045.
Kawuki, R.S., Kaweesi, T., Esuma, W., Pariyo, A., Kayondo, I.S., Ozimati, A., Kyaligonza, V., Abaca, A., Orone, J., Tumuhimbise, R., Nuwamanya, E., Abidrabo, P., Amuge, T., Ogwok, E., Okao, G., Wagaba, H., Adiga, G., Alicai, T., Omongo, C., Bua, A., Ferguson, M., Kanju, E. and Baguma, Y. 2016. Eleven years of breeding efforts to combat cassava brown streak disease. Breeding Science. 66(4): 560–571.
Kitano, H. 2004. Biological robustness. Nature Reviews Genetics. 5(11): 826–837.
Kurata, H., El-Samad, H., Iwasaki, R., Ohtake, H., Doyle, J.C., Grigorova, I., Gross, C.A. and Khammash, M. 2006. Module-based analysis of robustness tradeoffs in the heat shock response system. PLoS Computational Biology. 2(7): 0663–0675.
Levy, S.F. and Siegal, M.L. 2008. Network hubs buffer environmental variation in Saccharomyces cerevisiae. PLoS Biology. 6(11): 2588–2604.
Love, M.I., Huber, W. and Anders, S. 2014. Moderated estimation of fold

change and dispersion for RNA-seq data with DESeq2. Genome Biology. 15(12): 1–21.

Mangana, S. 2003. Cassava Brown Streak Virus Disease Research in Northern Mozambique. In: Cassava Brown Streak Virus Disease : Past , Present and Future. 14–17.

MtundaI, K., MuhannaI, M., RayaP, M. and KanjuQ, E. 2003. Current Status of Cassava Brown Streak Virus Disease in Tanzania. In: J.P. Legg and R.J. Hillocks (Editors) Cassava Brown Streak Virus Disease: Past, Present and Future. Natural Resources International Limited, 7–11.

Saithong, T., Saerue, S., Kalapanulak, S., Sojikul, P., Narangajavana, J. and Bhumiratana, S. 2015. Gene co-expression analysis inferring the crosstalk of ethylene and gibberellin in modulating the transcriptional acclimation of cassava root growth in different seasons. PLoS ONE. 10(9): 1–22.

Shannon, P., Markiel, A., Owen Ozier, 2, Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research.(13): 2498–2504.

Tomlinson, K.R., Bailey, A.M., Alicai, T., Seal, S. and Foster, G.D. 2018. Cassava brown streak disease: historical timeline, current knowledge and future prospects. Molecular Plant Pathology. 19(5): 1282–1294.

Watts, D.J. and Strogatz, S.H. 1998. Collectivedynamics of 'small-world' networks. Nature. 393(6684): 440–442.

Wu, J., Tan, S.Y., Liu, Z., Tan, Y.J. and Lu, X. 2017. Enhancing structural robustness of scale-free networks by information disturbance. Scientific Reports. 7(1): 1–13.

Wu, Y., Zhang, X., Yu, J. and Ouyang, Q. 2009. Identification of a Topological Characteristic Responsible for the Biological Robustness of Regulatory Networks. 5(7).