Shittu, U. 2022
J. Basic Appl. Res biomed 9(1): 1-4

10.51152/jbarbiomed.v9i1.221

JBAAR
Journal of Basic and Applied Research in Biomedicine
ISSN 2413-7014

Original Article

Microarray Gene Expression Data Generation and Pre-Processing of *Moringa Oleifera* Leaves for the Improvement of Medicinal Use

Umar Shittu

Department of Biology, Isa Kaita College of Education Dutsin-ma Katsina State, Nigeria

*Corresponding Author: shittuumarkk@gmail.com

Received: 5-3-2023
Revised: 11-4-2023
Published: 19-4-2023

Keywords:
Microarray, Gene expression data, Moringa oleifera leaves, Bioconductor R package

Abstract: *Moringa oleifera* is a plant species belonging to the family name called Moringaceae widely cultivated for human use. This study aimed to generate microarray gene expression data from the leaves of the Moringa oleifera plant and explore the usage of some tools available in the Bioconductor R package for the quality control. Six (6) young *Moringa Oleifera* leaves (YMOL) samples and six (6) old Moringa oleifera leaves (OMOL) samples were collected from the plant and processed for microarray data generation. Microarray gene expression raw data from the leaves of the Moringa oleifera plants were generated, each in a CEL file format and the usage of some tools available in R programming Bioconductor open source and development software project were explored for the quality control of the data. Affycoretools were installed in the R environment for pre-processing of microarray raw data. AffyQCReport tools were used to generate a comprehensive quality control (QC) report for the microarray unnormalized raw data in PDF format. It is recommended that Gene chip robust multiarray analysis (GCRMA) method can be used for visual inspection, background correction, normalization and summarization of this microarray raw data. The normalized microarray raw data can be used through the genetic engineering to improve the Moringa oleifera plant medicinal values in order to solve some medical problems especially with patients suffering from diabetes and hypertension and also can be of enormous importance in the fields of pharmacy and medicine at large.

Cite this article as: Shittu, U. (2022) Microarray Gene Expression Data Generation and Pre-Processing of Moringa Oleifera Leaves for the Improvement of Medicinal Use. Journal of Basic and Applied Research in Biomedicine, 9(1): 1-4 10.51152/jbarbiomed.v9i1.221

INTRODUCTION

Moringa oleifera is a plant species belonging to the genus *Moringa* with a family name called Moringaceae widely cultivated for its young seed pods and leaves used as vegetables, for traditional herbal medicine, and is also used for water purification (Kasolo *et al.*, 2010). The common names of the plant include *Moringa*, drumstick tree, and horseradish tree benzoyl tree is a fast-growing drought-resistant tree and is particularly suitable for dry regions, as it can be grown using rainwater without expensive irrigation techniques (Leone *et al.*, 2015). It is grown mainly in semiarid, tropical, and subtropical regions and tolerates a wide range of soil conditions, but prefers a neutral to slightly acidic (pH 6.3 to 7.0), well-drained sandy or loamy soil (Ted *et al.*, 2011). Microarray technology is one of the technologies that contribute to the development of studies in the field of Molecular Biology. Microarray is a new method for the detection of thousands of DNA or RNA at one time. This technology deals with the collections of DNA or RNA samples from microarray experiments and uses these samples to measure the expression levels of large numbers of genes or the whole genome of an organism or identify differential expressed genes or determine common expression patterns among the genes at a time. Microarray experiment involves the growth of an organism, collection of tissues, extraction of RNA or DNA, hybridization and scanning process to generate microarray data (Miller *et al.*, 2009).

Current scientific data on the corrective potential of *M. oleifera* leaves in chronic dyslipidemia and hyperglycemia as symptoms of diabetes and cardiovascular disease (CVD) risk. The studies reported in experimental animals and humans, although limited in number and variable in design, seem concordant in their support for this potential (Majambu, 2012). *Moringa oleifera* plant has many different applications in the medical field; a tree possesses a wide range of medicinal and therapeutic properties such as anti-inflammatory, anti-microbial, anti-fibrotic, anti-hyperglycaemic, anti-tumor, anti-cancer, and anti-oxidant properties (Ibrahim *et al.*, 2015). The aqueous extract of *M. oleifera* leaves have been demonstrated to exhibit a protective effect on ulcerated gastric tissue induced by aspirin, cerebral nodular lesion and cold stress in rats, wound healing property in rats, significant hypoglycemic and antidiabetic potential, antifertility activity, and the regulatory control on thyroid hormone status in adult Swiss rats (Saurabh *et al.*, 2016). *Moringa oleifera* is an important medicinal plant in the ayurvedic system (Ahmad *et al.*, 2014). It can be proved by several scientific studies also. It has several medicinal properties like hepatoprotective, anti-microbial, anti-cancer, antidiabetic, anti-inflammatory, anti-pyretic, analgesic activity, hypocholesterolemic effect, cardio protective property, anti-asthmatic, water purification, and antioxidant properties (Sanjive, 2017). However, despite the wide range of medicinal and therapeutic properties of *Moringa oleifera* for curing different infections discovered by previous researchers, still, the major problem now associated with especially patients having diabetes and hypertension is that they cannot rely on using *Moringa oleifera* plant only to maintain their glucose and blood pressure level within the control range, unless when they add other synthetic drugs. While the frequent use of some synthetic drugs daily can lead to complications in many organs within the human system, especially with those patients having the problem from an early age. On the other hand, DNA materials can be used by genetic engineering to improve the plant species that have medicinal values in order to solve some medical problems (Tharachand *et al.*, 2012). The overall aim of this research is to generate microarray gene expression data from the leaves of the *Moringa oleifera* plant and explore the usage of some tools available in the Bioconductor R package for the quality control of the data and also for the improvement of the medicinal value of the plant at large.

MATERIALS AND METHODS

Materials used for this study were Seeds of *Moringa Oleifera* plant, Planting land area space (at least four square

meters), Watering container, Leaves collection kit, Laptop computer system (core i5), Modem and internet data. Six (6) young *Moringa Oleifera* leaves (YMOL) samples and six (6) old *Moringa Oleifera* leaves (OMOL) samples were collected for microarray data generation. The leaves samples were dried under shadow and also processed in to powdered form. DNA extraction from the samples involves cell lysis by physical or chemical tissue extract and washing with detergents for separation of DNA by specific adsorption to membrane. The media was then treated with proteases for protein denaturation and removal followed by washing for removal of other cellular contaminants and also treated with RNase for RNA removal which result in generating Purified DNA. After extraction of purified DNA from the samples, one sample is labeled with a dye that fluoresces green and the other sample is labeled with a dye that fluoresces red. The Agilent RNA Spike-In Kit was developed to provide positive controls for monitoring the microarray workflow from sample amplification and labeling to microarray processing. DNA to DNA hybrids was formed by binding of single-stranded DNA molecule with a complementary strand. Scanning was made for image processing and then microarray raw data of *Moringa Oleifera* leaves were generated from the samples, each in a CEL file format (Mantione *et al.,* 2014). The microarray data analysis was performed with R programming Bioconductor open source and development software project.

RESULTS AND DISCUSSION

Two different array of *Moringa Oleifera* leaves were used for this study such as old *Moringa Oleifera* leaves (OMOL) and young *Moringa Oleifera* leaves (YMOL), each array with six (6) samples as shown in table 1. These samples were in the CEL file format required in the Bioconductor R package for the analysis of genome or microarray data.

Table 1: Young and old *Moringa Oleifera* leave samples

Array Index	Array Name
1	OMOL1.CEL
2	OMOL2.CEL
3	OMOL3.CEL
4	OMOL4.CEL
5	OMOL5.CEL
6	OMOL6.CEL
7	YMOL1.CEL
8	YMOL2.CEL
9	YMOL3.CEL
10	YMOL4.CEL
11	YMOL5.CEL
12	YMOL6.CEL

Images of some randomly selected arrays of microarray *Moringa Oleifera* leave samples raw data were visualized using affycoretools for the purpose of visual inspection. The essence of image visualization is to find out if there are technical problems eventually occurring only in one region of the array or not (Jennifer *et al.,* 2013). From these arrays visualized above in figure 1, it indicates that there were little technical problem associated with array because it shows noise by having small white dots in the array when compare with the array of equal colour intensity in all region of the array. In this case there is need to normalize this data.

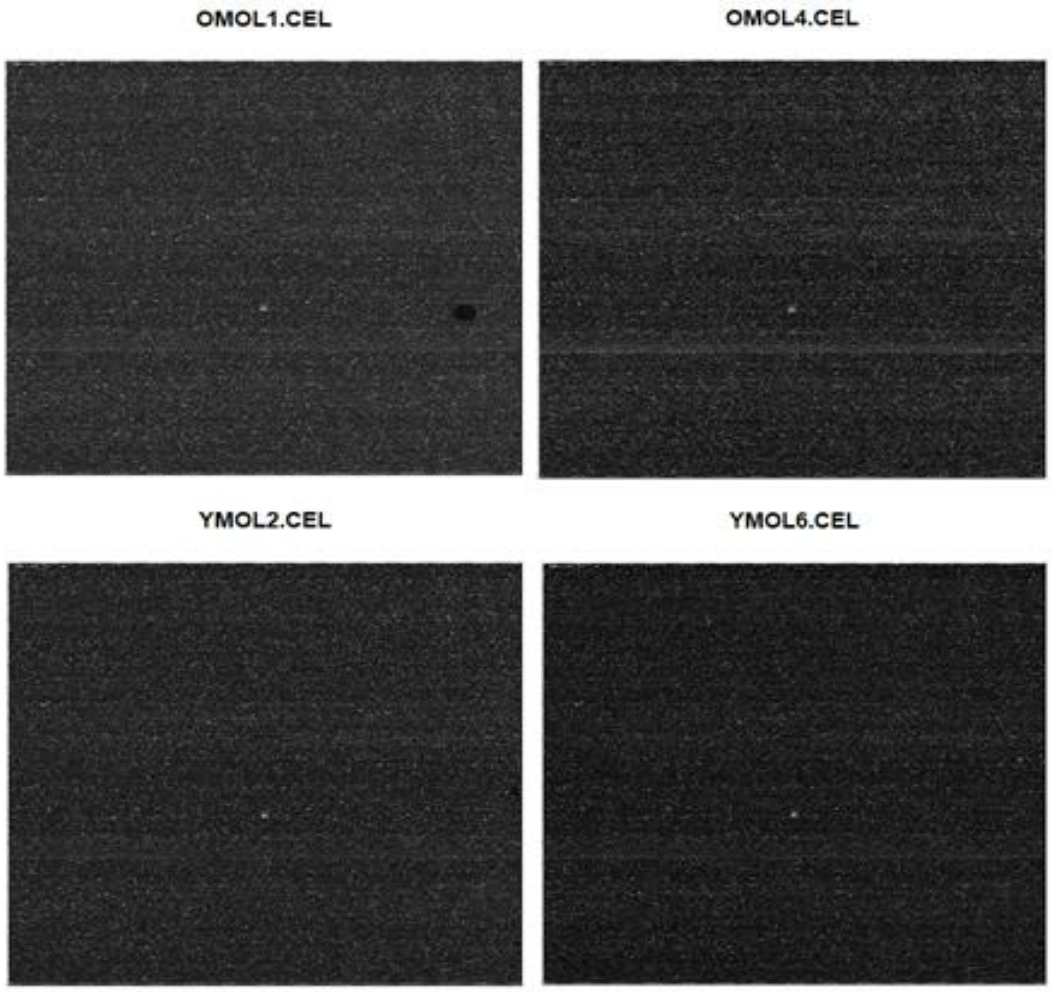

Figure 1: Microarray image visualization of *Moringa Oleifera* leave samples

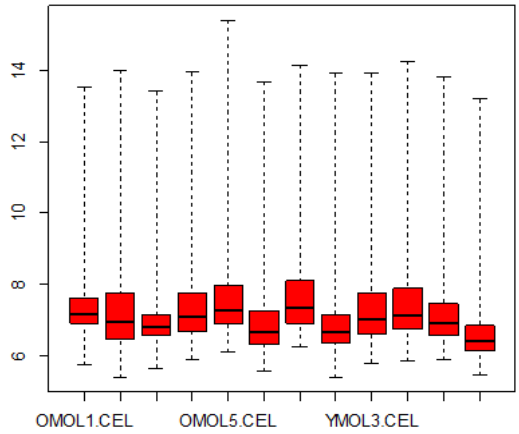

Figure 2: Box plot of *Moringa Oleifera* leaves Microarray raw data of two different arrays.

The distribution of the intensities in each array of all the *Moringa Oleifera* leaves microarray raw data samples in Figure 2 indicates the need of appropriate data normalization, but no evidence of an outlier, because no-one among the arrays from the data set shows the uniqueness about the distribution of intensities. An outlier of microarray data usually occurred as a result of experimental errors or variability.

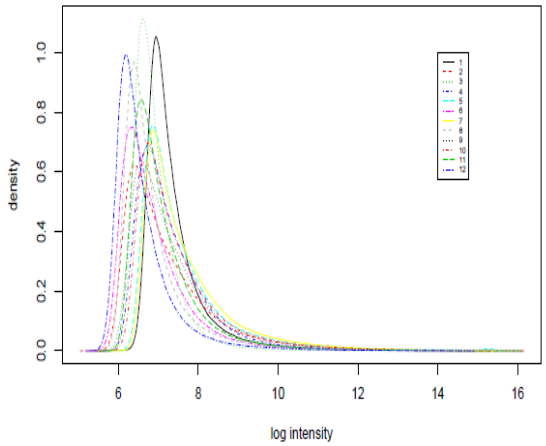

Figure 3: Histogram plot of *Moringa Oleifera* leaves Microarray raw data of two different arrays.

A histogram plot of array log intensity against density levels in Figure 3 provides information with a graphical representation of the density and log intensity distribution of microarray *Moringa Oleifera* leaves raw data in order to observe also whether there is need for data normalization or not. The histogram plot of these arrays indicates overlapping between the samples, old *Moringa oleifera* leaves (OMOL3) sample three (3) overlapped with having the highest density value and old *Moringa oleifera* leaves (OMOL2) sample two (2) also shown the highest log intensity value, because of these problems of overlapping observed associated with the raw data, it clearly indicates that, there is a need for data normalization before applying any advance genome data analysis for better and reliable results.

Simpleaffy plot is a general quality control statistic which provides visual representation of the microarray raw data by using beta-actin and GAPDH. Beta- actin (β-Actin) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were considered as constitutive housekeeping genes for RT-PCR and used to normalize changes in specific gene expression, because they are usually expressed in most of the cell types and are relatively long genes (Thomas *et al.*, 2010). QC plot in Figure 4 shows the 3′ to 5′ ratios for gene hybridization with specific control to the array type, the percentages of present gene calls and background for β-Actin and GAPDH. Plotted triangles represent Actin values while plotted circles represent GAPDH values (Jennifer *et al.*, 2013). It visualized the first array from the bottom of the plot to the last array at the top which corresponds to the order of the samples with ratios and percentages, the dotted horizontal lines separate this plot into rows one for each array. The dotted vertical lines provide a scale from -3 to 3. A line to the left corresponds to a down-scaling, to the right, to an up-scaling. According to Affymetrix GAPDH and β-Actin values that are considered a potential outlier when the ratio > 1.25 and are coloured red, The blue stripe in the image represents the range where scale factors are within 3-fold of the mean for all chips, If any scale factors fall outside this 3-fold region, they are all coloured red and unacceptable but if fall within the scale, are coloured blue and consider acceptable (Jarno, 2008).

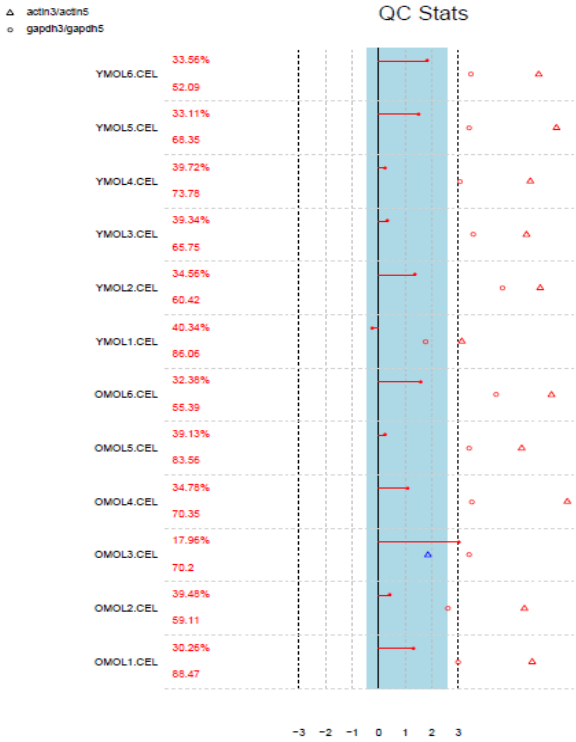

Figure 4: QC plot of *Moringa Oleifera* leaves Microarray raw data of two different arrays.

In Figure 4 only old *Moringa Oleifera* leaves sample three (3) (OMOL3) beta-actin value fall within the scale and considered acceptable, but with unacceptable GAPDH value with all others remaining samples fall outside the scale to the right side known as up-scaling is abnormal and considered as potential variability. This plot also indicates the need of an appropriate normalization method, because only one among samples was found with acceptable beta-actin 3′ to 5′ ratio value, but all other samples with unacceptable values and also all the samples with unacceptable GAPDH values.

RNA degradation analysis was performed on this microarray raw data to assess the quality of RNA and gives a good indication of the quality of the sample that has been hybridized to the array. RNA degradation analysis is usually start from the 5′ end to 3′ end of the molecule, a strong degradation would result with the values for the probes closer to the 5’ end and when the degradation is progressing to 3′ end the probe set should elevated (Jarno, 2008). In figure 5, it indicates a good quality of RNA in all the twelve (12) samples of *Moringa Oleifera* leaves microarray raw data of two different arrays. Because as the RNA degradation was progressing from 5′end to 3′ end, there was an increase in probe set number and mean intensity.

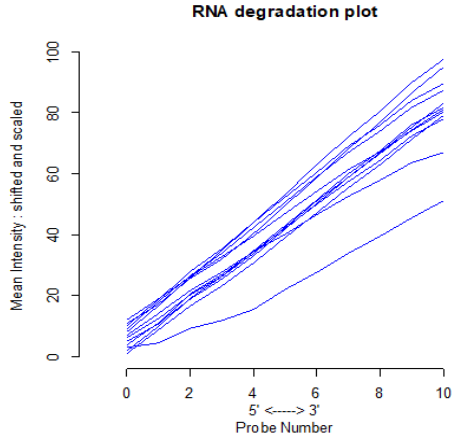

Figure 5: RNA degradation plot of *Moringa Oleifera* leaves Microarray raw data of two different arrays.

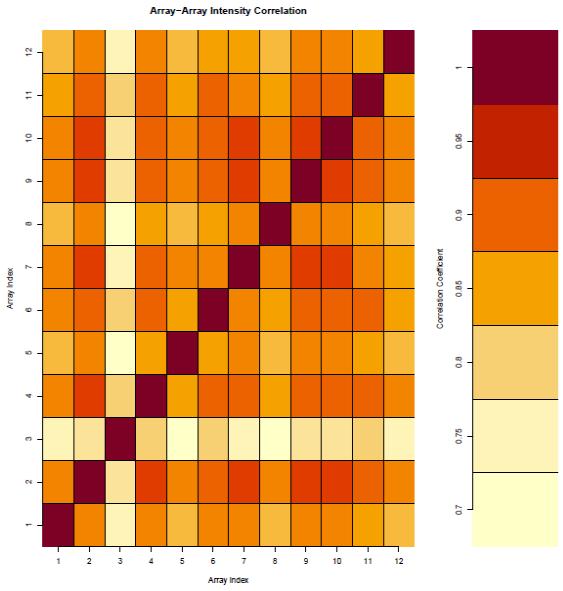

Figure 6: A heat map of the array to array Spearman rank correlation coefficients of *Moringa Oleifera* leaves Microarray raw data of two different arrays.

In figure 6 correlation coefficient analysis was performed to detect outlier arrays, error in hybridizations, and then to get valuable information about phenotypic characteristics of the microarray raw data. This heat map of an array to array correlation coefficients indicates a good quality of this data since the smallest correlation coefficient was 0.7, but only old *Moringa Oleifera* leaves sample three (3) (OMOL3) array has quality problems with very high signals, high variability, and strong background, also appeared different in this plot, correlating poorly with other arrays. It was recorded that, pairs of arrays had a stronger correlation within the tissues than the correlation between the tissues. Samples from similar tissues or treatments tend to have a higher correlation coefficient. The better way to identify which array among many arrays associated with the problem is to correlate all the arrays with each other (Wilson *et al.*, 2004).

CONCLUSION

Microarray gene expression data from the leaves of the *Moringa oleifera* plants were generated and the usage of some tools available in R programming Bioconductor open source and development software project were explore for the quality control of the data. Affycoretools were installed in the R environment for pre-processing of microarray raw data. AffyQCReport tools were used to generate a comprehensive quality control (QC) report for the microarray unnormalized raw data in PDF format. It is recommended that Gene chip robust multiarray analysis (GCRMA) method can be used for visual inspection, background correction, normalization and summarization of *Moringa oleifera* leaves microarray raw data. GCRMA method provides highly precise desirable estimates of gene expression and if the affy library has been loaded into memory, GCRMA is ignoring the mismatch intensities and taking into account the probe sequence information and allows an efficient filtering of irrelevant probe sets. This method can also normalize the microarray raw data to improve the quality of the data for advanced analysis and to store the data as an expression set. The normalized *Moringa oleifera* leaves microarray raw data can be used through the genetic engineering to improve the *Moringa oleifera* plant medicinal values in order to solve some medical problems especially with patients suffering from diabetes and hypertension and also can be of enormous importance in the fields of pharmacy and medicine at large.

REFERENCES

Ahmad, F. A. R., Muhammad, D. I., Saie, B. K., (2014). *Health benefits of Moringa oleifera.* Asian Pac J Cancer Prev, 15 (20), 8571-8576.

Ibrahim Majali, Osama Y. Althunibat, HaithamN. Qaralleh (2015). *Antimicrobial and Immunomodulatory activities of Moringa peregrine-MINIREVIEW.* Journal of Basic and Applied Research 1 (2): 55-61www.jbaar.com

Jarno Tuimala (2008). DNA microarray data analysis using Bioconductor. CSC, the Finnish IT centre for Science, CSC – IT Center for Science Ltd. 2008 ISBN 978-9525520-34-7http://www.csc.fi/oppaat/R

Jennifer, V. P. and Alex, G. L. (2013). Microarray Technique, Analysis, and Applications in Dermatology. Journal of Investigative Dermatology (2013) 133, e7. doi:10.1038/jid.2013.64 Jones and Bartlett (2010). Series in biomedical informatics, R for Medicine and Biology, 2010 ISBN 978-0-7637-5808-0.

Kasolo, J. N., Bimenya G.S., Ojok, L. (2010). Phytochemicals and uses of Moringa oleifera leaves in Ugandan rural communities. J Med Plants Res, 4, 753-7.

Leone, A., Spada, A., Battezzati, A., Schiraldi, A., Aristil, J., Bertoli, S., (2015). *"Cultivation, Genetic, Ethnopharmacology, Phytochemistry and Pharmacology of Moringa oleifera Leaves"*: An Overview". Int J Mol Sci. **16** (6): 12791–835. doi:10.3390/ijms160612791.

Majambu Mbikay (2012). *Therapeufic potential of M. oleifera leaves in chronic hyperglycemia and dyslipidemia: a review.* published 01 March,2012. doi: 10.3389/fphar.2012.00024.

Mantione K. J., R. M. Kream, H. Kuzelova, R. Ptacek, J. Raboch, *(*2014) Comparing bioinformatic gene expression profiling methods: Microarray and RNA-Seq. *Medical Science Monitor Basic Research* 20: 138–142. https://doi.org/10.12659/MSMBR.892101.

Miller M. B., Tang Y. W. (2009). **"***Basic concepts of microarrays and potential Applications in clinical microbiology".*Clinical Microbiology Review, v.22 (4); PMC2772365.

Sanjive Kumar (2017). *Medicinal Importance Of Moringa Oleifera :* Drumstick Plant. Indian J.Sci.Res. Department of Chemistry A P G College, Allahabad.U.P.,India. *16 (1):* 129-132, 2017. ISSN: 2250-0138(Online). ISSN: 0976-2876 (Print)

Saurabh Dixit, Astha Tripathi, Astha Tripathi (2016). *Medicinal Properties Of Moringa Oleifera:* A Review. E-ISSN 2348-6457. Volume-3, Issue-2 April- 2016 www.ijesrr.org Email-editor@ijesrr.org

Ted Radovich, C.R. Elevitch, E. D. (2011). *"Farm and Forestry Production and Marketing Profile for Moringa* (revised February 2011)" (PDF). In: Specialty Crops for Pacific Island Agroforestry.

Tharachand C, Immanuel Selvaraj C and Mythili M. N. (2012). *Molecular markers in characterization of medicinal plants: An overview.* Division of Plant Biotechnology, School of Biosciences and Technology, VIT University, Vellore 632014, India. Research in Plant Biology, 2(2): 01-12, 2012.

Thomas Girke, UC Riverside (2010). R & Bioconductor Manual. Institute for Integrative genome Biology.

Wilson CL, Miller CJ (2005). "Simpleaffy: a Bioconductor package for Affymetrix Quality Control and data analysis". Bioinformatics 21(18): 3683 -5 doi:10.1093/bioinformatics/bti605. PMID 16076888.