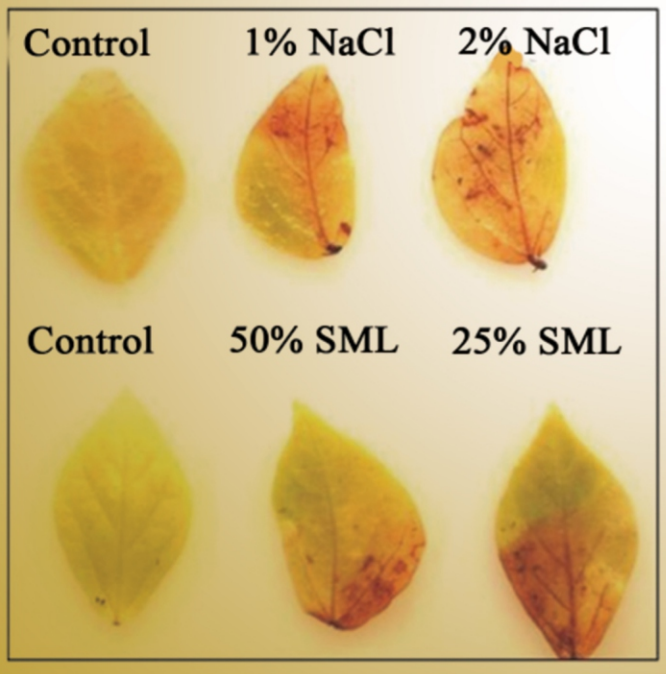
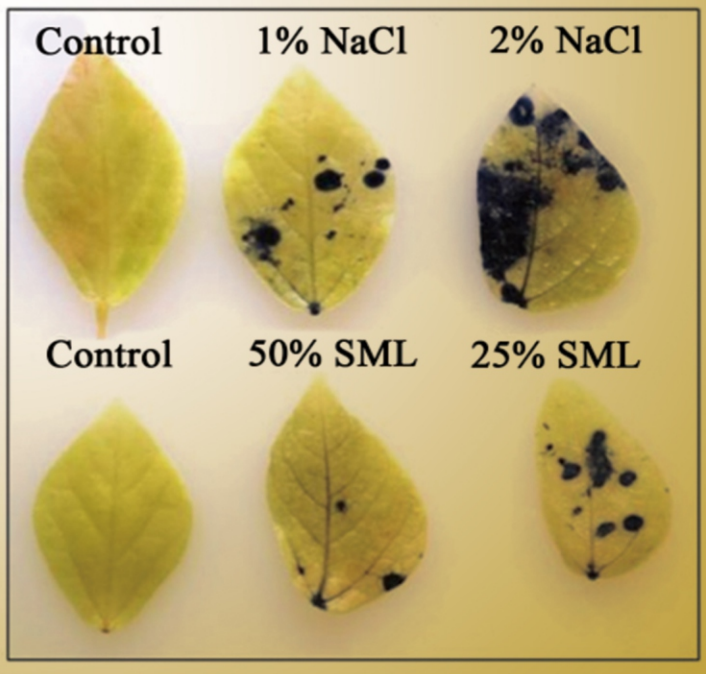
ISSN: 2161-6620 Volume 6, Number 4, October 2016

Scientific Research Publishing

American Journal of Molecular Biology





ISSN: 2161-6620

www.scirp.org/journal/ajmb

Journal Editorial Board

ISSN Print: 2161-6620 ISSN Online: 2161-6663
http://www.scirp.org/journal/ajmb

Editor-in-Chief

Editorial Board

Scientific Research Publishing

American Journal of Molecular Biology, 2016, 6, 125-169
http://www.scirp.org/journal/ajmb
ISSN Online: 2161-6663
ISSN Print: 2161-6620

Table of Contents

Volume 6 Number 4 October 2016

Expression Analysis of WRKY Transcription Factor Genes in Response to Abiotic Stresses in Horsegram (*Macrotyloma uniflorum* (Lam.) Verdc.)

K. Kiranmai, L. R. Gunupuru, A. Nareshkumar, V. A. Reddy, U. Lokesh, M. Pandurangaiah, B. Venkatesh, T. V. Kirankumar, C. Sudhakar.....125

The Potential of DNA Barcode-Based Delineation Using Seven Putative Candidate Loci of the Plastid Region in Inferring Molecular Diversity of Cowpea at Sub-Species Level

P. Okoth, J. Muoma, M. Emmanuel, W. Clabe, D. O. Omayio, P. O. Angienda.....138

Genotyping of *E. coli* Isolated from Urinary Tract Infection Patients Containing B-Lactamase Resistance Gene CTX-M Group 1 in Sanandaj Medical Health Centers

A. Omati, K. Davari, B. Shokrolahi.....159

The figure on the front cover is from the article published in American Journal of Molecular Biology, 2016, Vol. 6, No. 4, pp. 125-137 by Kurnool Kiranmai, *et al.*

American Journal of Molecular Biology (AJMB)
Journal Information

SUBSCRIPTIONS

The *American Journal of Molecular Biology* (Online at Scientific Research Publishing, www.SciRP.org) is published quarterly by Scientific Research Publishing, Inc., USA.

Subscription rates:
Print: $79 per issue.
To subscribe, please contact Journals Subscriptions Department, E-mail: sub@scirp.org

SERVICES

Advertisements
Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)
Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.
E-mail: sub@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:
E-mail: ajmb@scirp.org

Scientific Research Publishing

American Journal of Molecular Biology, 2016, 6, 125-137
http://www.scirp.org/journal/ajmb
ISSN Online: 2161-6663
ISSN Print: 2161-6620

Expression Analysis of WRKY Transcription Factor Genes in Response to Abiotic Stresses in Horsegram (*Macrotyloma uniflorum* (Lam.) Verdc.)

Kurnool Kiranmai[1], Lokanadha Rao Gunupuru[2], Ambekar Nareshkumar[3], Vennapusa Amaranatha Reddy[1], Uppala Lokesh[1], Merum Pandurangaiah[4], Boya Venkatesh[1], Tanguturi Venkata Kirankumar[1], Chinta Sudhakar[1*]

[1]Department of Botany, Sri Krishnadevaraya University, Anantapur, India
[2]Molecular Plant-Microbe Interactions Laboratory, University College Dublin, Dublin, Ireland
[3]Department of Crop Physiology, GKVK, Bangalore, India
[4]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China
Email: *chintasudhakar@yahoo.com

How to cite this paper: Kiranmai, K., Gunupuru, L.R., Nareshkumar, A., Reddy, V.A., Lokesh, U., Pandurangaiah, M., Venkatesh, B., Kirankumar, T.V. and Sudhakar, C. (2016) Expression Analysis of WRKY Transcription Factor Genes in Response to Abiotic Stresses in Horsegram (*Macrotyloma uniflorum* (Lam.) Verdc.). *American Journal of Molecular Biology*, **6**, 125-137.
http://dx.doi.org/10.4236/ajmb.2016.64013

Received: July 20, 2016
Accepted: October 10, 2016
Published: October 13, 2016

Abstract

Drought and salt stress are two major environmental constraints that limit the productivity of agriculture crops worldwide. WRKY transcription factors are the plant-specific transcription factors that regulate several developmental events and stress responses in plants. The WRKY domain is defined by a 60-amino acid conserved sequence named WRKYGQK at N-terminal and a Zinc Finger-like motif at the C-terminal. WRKY genes are known to respond several stresses which may act as negative or positive regulators. The function of most of the WRKY transcription factors from non-model plants remains poorly understood. This investigation shows the expression levels of eight WRKY transcription factor genes from horsegram plant under drought and salt stress conditions. The increase in mRNA transcript levels of WRKY transcription factor genes was found to be high in drought stressed plants compared to salt-stressed plants. The levels of MDA which indicates the lipid peroxidation were less in drought stress. More ROS is produced in salt stress conditions compared to drought. The results show that the expression of WRKY transcription factors in drought stress conditions is reducing the adverse effect of stress on plants. These results also suggest that, during abiotic stress conditions such as drought and salt stress, WRKY transcription factors are regulated at the transcription level.

Keywords

Drought, Salt, Lipid Peroxidation, Plant Abiotic Stress, qRT-PCR,

Reactive Oxygen Species, WRKY

1. Introduction

Plants in their natural environmental conditions are always subjected to various biotic and abiotic stresses. Abiotic stresses are the primary reason behind the decline of crop quality and yield worldwide [1]. Drought and salt stress are two major abiotic stresses that limit crop productivity. Excessive amounts of salts in the agriculture soil cause both osmotic and ionic stresses [2]. The rapid spread of drought and soil salinity in many regions worldwide may cause serious salt contamination in more than 50% of all arable lands by the year 2050 [3]. Transcription factors (TFs) play a vital role in stress response by regulating their expression in target gene through interaction with specific *cis*-acting elements present in the promoter of the gene [4] [5]. WRKY TFs are the largest superfamily of TFs specific to plants. They are classified into three groups based on the number of WRKY domains and nature of their zinc-finger motif. Group I contains two WRKY conserved domains and a classical Zinc finger motif. Group II contains single WRKY domain and a classical zinc finger motif. Group II is divided into five or more subgroups based on short conserved structural WRKYGQK motif. Group III protein of WRKY superfamily contains a single WRKY domain and a modified zinc finger motif C_2-CH rather than classical C_2-H_2. Group II WRKY TFs containing WRKYGQK heptapeptide with Zinc finger CX4-5CX22-23HHX1H is the largest group in most of the plants [6] [7].

A wide range of WRKY TFs has been identified from different plants 72 WRKYs from *Arabidopsis* and 96 from rice [7] and 52 WRKYs in papaya [8] have been identified. Most of the WRKYs are found to play a role in biotic stress response, some in abiotic stress and few WRKYs are known for their role in both biotic and abiotic stress response. During normal growth conditions, WRKY TFs regulate plant processes like senescence, trichome development [9] biosynthesis of secondary metabolites [6] and dormancy [10]. Many research groups have investigated the role of WRKY TFs in biotic and abiotic stress in various plants. Overexpression of WRKY33 a WRKY TF family protein, providing resistance to necrotrophic fungi species such as *Botrytis cinerea* and *Alterneria brassicicola* has been reported in *Arabidopsis* [11]. WRKY TFs like *HvWRKY*33 involving in cold and drought tolerance in barley [12], oxidative stress response WRKY TF like *FcWRKY*40 from *Fortunella crassifolia* [13], and *ZmWRKY*33 conferring tolerance to salt stress were reported in *Arabidopsis* [14]. 46 different WRKY TFs were identified from *Brassica napus* expressing during infection with fungal pathogens *Sclerotinia sclerotiorum* and *Alternaria brassicae*, and ABA, JA, SA and ET treatments [7]. However, the function of each WRKY in various stresses is poorly understood. It is important to get further insight and investigate the response of WRKY TFs under abiotic stress conditions.

Membrane damage is taken as a parameter to determine the lipid destruction of the membrane under various stress conditions. Lipids peroxidation is considered as the

most desecrate process known to occur in every living organism. During lipid peroxidation, small hydrocarbon fragments namely Malondialdehyde (MDA) is formed from polyunsaturated precursors [15]. Lipid peroxidation takes place when Reactive oxygen species (ROS) levels are reached above-threshold, directly affecting normal cellular functioning and also aggravating the oxidative stress through the production of lipid-derived radicals [16]. Level of ROS is controlled by plant cells through detoxification of excess ROS during stress. To help in the detoxification of excess ROS, plants have evolved strong antioxidant defence mechanism [17].

Horsegram belongs to the family Fabaceae and is comparatively a hardy drought tolerant crop which grows well under adverse conditions. The crop accounts for approximately 5% - 10% of the India's pulses, with annual production of about 0.65 million tonnes. It is considerably used as dry fodder, cattle feed, cover crop and for water conservation in the semi-arid region. Horsegram is widely grown in the semi-arid regions of India and generally considered as protein rich poor man's crop. It grows well under dry conditions with the need of low rainfall and marginal soil fertility. Depending on its well adaptive growth conditions and life cycle, there is a high possibility that horsegram plant contains a variable number of genes that can be used to provide stress tolerance. Due to the higher similarity in genome among pulse crops, comprehension of the underlying genetic mechanism of various stress related genes from this plant will be of significant advantage to transform these genes to other legume crops.

The present investigation had made an attempt to identify, the abiotic stress responsive role of horsegram WRKY TFs by qRT-PCR expression analysis under drought and salinity stresses. These experimental results, making feasible to use horsegram WRKY TF genes to develop abiotic stress tolerant crop plants in near future.

2. Materials and Methods

2.1. Plant Material and Stress Treatments

Horsegram seeds were sown in pots containing mixed soil (soil: vermiculite in 3:1 ratio) and allowed to grow in a greenhouse under natural photoperiod and temperature conditions for 19 days. To impose salt stress 19 days, old plants were irrigated with 1% and 2 % NaCl solution. The electrical conductivity of the soil saturation extract was 4.0 $mS{\cdot}cm^{-2}$ and care was taken to avoid drainage of solution during the treatment by giving the water slightly less than field capacity; stress was maintained for four days. The electrical conductivity was monitored and adjusted using conductivity bridge (Elico, Hyderabad, India) on alternate days. For drought stress, the plants were withheld from irrigation until to reach 50% and 25% soil moisture level (SML) and maintained at the same level for three days by measuring SML regularly by following standardised gravimetric method. The samples from stressed and non-stressed plants were collected and flash frozen in liquid nitrogen. Each experiment was repeated at least three times.

2.2. Determination of Growth under Stress

After the stress treatments, plants were carefully uprooted from the soil, to avoid any

breakage of lateral roots and washed thoroughly under running tap water. Plant growth was determined by measuring the root length, shoot length, and plant fresh weight.

2.3. Histochemical Localization of ROS

ROS are produced as a result of stress which leads to oxidative damage. The levels of ROS species like $O_2^{\bullet -}$ and H_2O_2, produced during the stress treatments were detected using Diaminobenzidine (DAB) and Nitoblurtetrazolium chloride (NBT) [18]. For detection of $O_2^{\bullet -}$ the leaf material was immersed in a solution of 0.1% NBT in 50 mM phosphate buffer (pH 7.8) at room temperature. The immersed leaves were illuminated with two 20-W fluorescent tubes for 2 - 4 h until the dark blue spots appear. For localization of H_2O_2, excised leaves were immersed in a 0.1% DAB solution in 50 mM phosphate buffer (pH 3.8) and incubated at room temperature for 8 h until brown spots appear. The leaves were then bleached with 50% ethanol to visualize the spots.

2.4. Determination of MDA Content

The levels of lipid peroxidation were measured [19] by estimating the MDA content, which is a secondary product of polyunsaturated fatty acids and 2-thiobarbituric acid reactive substances. The MDA content was estimated using 2-thibarbutiric acid at 400, 532 and 600 nm [19]. Five hundred milligrams of leaf tissue was used for the extraction of MDA. One millilitre of the supernatant was mixed with 1 ml of 20% TCA containing 0.65% thiobarbituric acid in a clean glass tube, in another tube 1 ml of supernatant was treated with 1 ml of 20% TCA alone and mixed thoroughly. The mixtures were heated at 95˚C for 30 min and absorbance was recorded at 532 nm (MDA), 440 nm (carbohydrates) and 600 nm (phenylpropanoid pigments). The MDA content was determined by its molar extinction coefficient (155 $mM^{-1} \cdot cm^{-1}$) using the formula.

$$\left[\left(\text{Abs532}+\text{TBA}\right)-\left(\text{Abs600}+\text{TBA}\right)-\left(\text{Abs532}-\text{TBA}\right)-\left(\text{Abs600}-\text{TBA}\right)\right]=\text{A}$$

$$\left[\left(\text{Abs440}+\text{TBA}-\text{Abs600}+\text{TBA}\right)0.0571\right]=\text{B}$$

$$\left[\text{MDA equivalents}\left(\mu\text{mol}\cdot\text{ml}^{-1}\right)=\left(\text{A}-\text{B}/157000\right)103\right]=\text{C}$$

$$\text{MDA equivalents}\left(\mu\text{mol}\cdot\text{g}^{-1}\right)=\left(\text{C}\times 15\times 1/0.5\right).$$

2.5. Isolation of RNA and cDNA Synthesis

Total RNA was isolated from the frozen leaf material using TRIzol reagent. RNA concentration was accurately quantified by the UV-spectrophotometric measurement and 3 µg of total RNA was separated on 0.8% agarose gel containing formaldehyde to check RNA integrity. DNA contamination was removed with TURBO DNA-free kit according to the instructions in the manual. Two hundred nanograms of RNA was used to construct cDNA by Thermo Scientific RevertAid Reverse Transcriptase enzyme and oligodT primers at 42˚C. Obtained cDNA was subsequently used as a template for qRT-PCR analysis [20].

2.6. Quantitative Real-Time PCR Analysis

Eight WRKY TF genes were selected to study the expression levels under drought and salt stress conditions by the qRT-PCR analysis. For qRT-PCR assay gene, WRKY specific expression primers were designed using the Primer Express 5.0 software at annealing temperature of 56˚C - 60˚C. Primer sequences used for the amplification were listed in Table 1. qRT-PCR was performed using Applied biosystem power SYBR-green mix. The appearance of PCR products was monitored by detecting the increase in fluorescence caused by binding of SYBR green dye. The length of amplification fragment was about 150 bp. Each reaction was repeated at least three times. The expression analysis was conducted in an Applied Biosystem StepOne Realtime PCR-System thermal cycling block. Conditions maintained are, 1min at 95˚C; 1 min at 57˚C and 1 min at 72˚C for 40 cycles, followed by melting curves analysis at 95˚C for 1 min. Expression of *MuActin* gene was used as internal control for normalization. Relative expression of mRNA was calculated using $2^{-\Delta\Delta ct}$ formula [21].

2.7. Data Analysis

All physiological and biochemical data were analysed using SPSS version 16.0. Data presented were calculated using One-way ANOVA and statistical analysis was carried out by using Post hoc multiple comparisons Duncan's test at a significance level of $P < 0.05$.

Table 1. Primers sequences used for the amplification of gene in qRT-PCR analysis.

Primers		Sequence
WRKY 3	F	5'ATGGGCAGAAACATGTTAAA3'
	R	5'TGAGTAACGGCGATTTGGTT3'
WRKY 12	F	5'CCCGAGGAGTTATTACAGATGCA33'
	R	5'CCGATAGTCGTCCCACTCTCTT3'
WRKY 40	F	5'GTCGGAGAGTAGCTCAACGG3'
	R	5'CCATCTTTCACAACGAGGGT3'
WRKY 75	F	5'-GAGGGATATGATAATGGGTCG3'
	R	5'CGACCACTTCTTGGTCCAC3'
WRKY 53	F	5'CCAGAATCTCCGGCATCG3'
	R	5'CTCCTTGGGAATTTGGCGCC3'
WRKY57	F	5'GCAGATGCACGGTGAAGAAG3'
	R	5'GGGAATCCAATGGTTTGATGGC3'
WRKY 65	F	5'GCCTAAACCTAAACCCGAGC3'
	R	5'-TCTCCCATCGGAAAGAACAC-3'
WRKY 33	F	5'AAGGAGAGGATGGTTACAATTGGA3'
	R	5'CGAGGATTCTCACTTCCTTTCAC3'
Actin	F	5'TCCATAATGAAGTGTGATGT3'
	R	5'GGACCTGACTCGTCATACTC3'

3. Results

3.1. Effect of Stress on Plant Growth

Effect of stress is usually perceived as a decrease in photosynthesis and growth of the plant. Growth parameters were calculated for uprooted plants. In general, the overall growth of plant was decreased under both stress conditions. However, the inhibitory effect of salt stress was more on the growth of horsegram when compared to drought. Further, more reduction in the growth of the plant was observed with increased levels of stress. The decrease in root growth was up to 44% in drought and 60% in salt stress. Similarly, the shoot length was decreased up to 22% in drought and 48% under salt stress conditions (**Figure 1**). There was 37% reduction in the plant total biomass in drought stress conditions and up to 78% under salt stress conditions (**Figure 2**). The depletion of chlorophyll and leaf drying was also observed at 2% NaCl stress which leads to plant dying. Dying of leaves and eventually entire plant due to high salt concentration in plants was also reported [22].

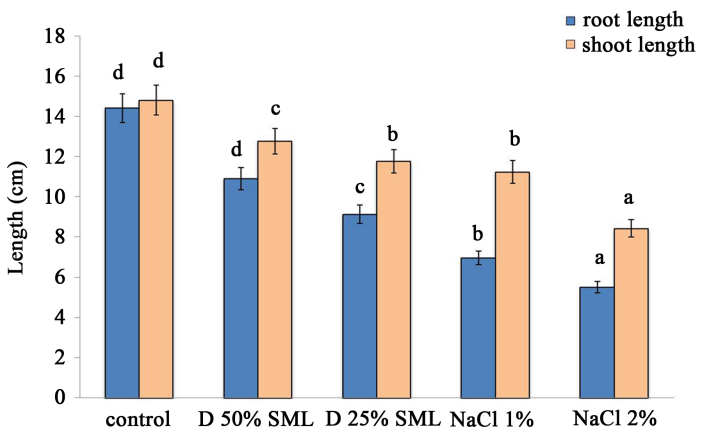


Figure 1. Varying root length and shoot length of horsegram plant under control, drought and salt stress conditions (±SD). And letters shown above the bars are significantly different at $P < 0.05$ (DMR).

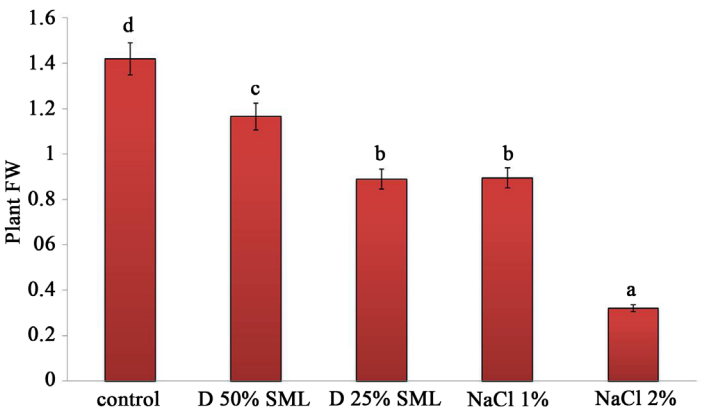


Figure 2. Fresh biomass of horsegram plants under control, drought, and salt stress conditions (±SD), and letters shown above the bars are significantly different at $P < 0.05$ (DMR).

3.2. Effect of Drought and Salt Stress on MDA Levels

MDA, the product of lipid peroxidation is often used as an indicator to check the extent of oxidative stress in plant [23]. The formation of ROS increases the levels of cellular lipid peroxidation, such increase in the production of ROS and peroxidation levels differ from species and severity of stress [24]. Total MDA content was estimated from the leaf tissue of control and stressed plants. The results suggest that the MDA content was significantly increased under both the stresses and an increase in the MDA content was observed proportional to the intensity of stress applied. However, the level of MDA is more in salt stress than in drought stress conditions (**Figure 3**). An increase in the MDA concentration under abiotic stress conditions like Pb stress was previously reported in groundnut plants by Nareshkumar *et al.* [25].

3.3. Histochemical Detection of O_2^- and H_2O_2 in Leaves

In situ histochemical staining assay was employed to detect the accumulation of O_2^- and H_2O_2, the two important ROS molecules produced during stress. Under control conditions there was no accumulation of free radicals, O_2^- and H_2O_2 was observed. However, remarkable differences were observed in the accumulation of ROS under stress conditions. Blue spots are the indicator of O_2^- and brown spots indicate the H_2O_2 accumulation. The results showed that the accumulation of ROS was more under salt stress conditions compared to drought (**Figure 4**).

3.4. Expression Analysis of WRKY TFs under Drought and Salt Stresses

qRT-PCR analysis was conducted to check the expression levels of different WRKY TF genes under drought and salt stress conditions. The results showed the different levels of regulation in the WRKY genes under different abiotic stresses conditions. In drought stress conditions, the highest level of expression was observed in *WRKY*3 and *WRKY* 75 was found to be negatively regulated. There was an increase in the transcription levels of different genes from 1.2 fold in *WRKY*33 to 12 folds in *WRKY*3 under drought

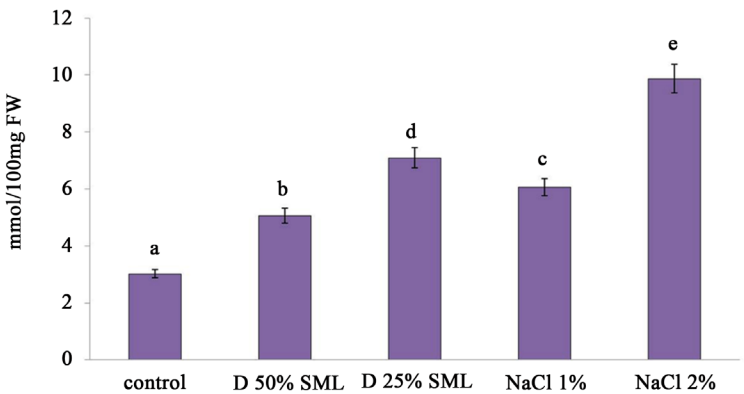
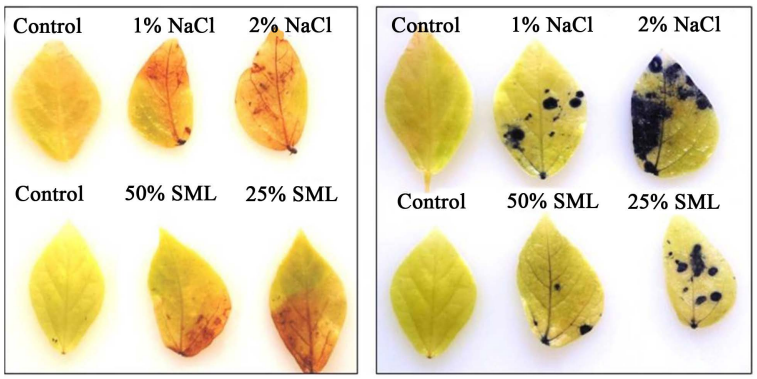


Figure 3. MDA content in leaf samples of horsegram under control, drought and salt stress conditions. The data represented is the mean of three different experiments ±SD, and the letters shown above the bars are significantly different at $P < 0.05$ (DMR).



(a) (b)

Figure 4. *In situ* histochemical detection of O_2^- and H_2O_2 under control, drought and salt stress conditions. (a) Brown spots indicate the H_2O_2 accumulation which is stained with DAB. (b) Blue colour spots indicate O_2^- accumulation stained with NBT. The intensity of colour spots is seen under salt stress conditions. Control plants did not show any colour due to the absence of ROS.

stress conditions at 25% SML. In salt stress, at 2% NaCl, the expression levels of genes were increased up to 6.1 folds in *WRKY*33. However, the negative regulation was also recorded in *WRKY*40 and *WRKY*53 genes whose transcription levels were downregulated by 0.63 fold in *WRKY*53 and 3.8 folds in *WRKY*40 genes (**Figure 5**).

4. Discussion

Abiotic stresses are the major limiting factors that reduce the productivity of crop worldwide. In the present investigation, we are delineated representing the role of WRKY TF gene from horsegram under drought and salt stress conditions. WRKY TFs are the largest superfamily TF that is specific to plants. WRKY recognize and bind to specific TTGAC (C/T) W-box elements found in promoters of a large number of plant defence related genes [6] [27] [28]. The first WRKY cDNAs were cloned from sweet potato, wild oat, parsley and *Arabidopsis*, based on the ability to bind specifically to the DNA sequence motif TTGACCT, which is known as the W box [29] [30] [31]. Current information available is suggesting that a large number of WRKY TFs plays a pivotal role in pathogen-induced defence mechanism [32]-[34]. However, there are fewer reports on the involvement of WRKY TFs in drought and salt stress response. In earlier studies expression of six WRKY TFs under drought stress have been reported [8] in *Carica papaya*. *AtWRKY*57 conferring drought tolerance by elevated ABA levels was reported [35]. A group II WRKY TF named *WRKY*58 was characterized from maize enhancing the tolerance towards drought and salt stress [36]. There are reports that few WRKY genes can be induced by both biotic and abiotic stress conditions. Zheng *et al*. [11] have reported that *AtWRKY*33 is a multifunctional gene that involves both in biotic and abiotic stresses.

In this report, we studied eight WRKY TF genes from horsegram, under drought and salt stress. WRKY3 gene has shown the highest expression level with 12 folds increase in drought stress among eight genes selected. Strong *VpWRKY*3 transcription levels are

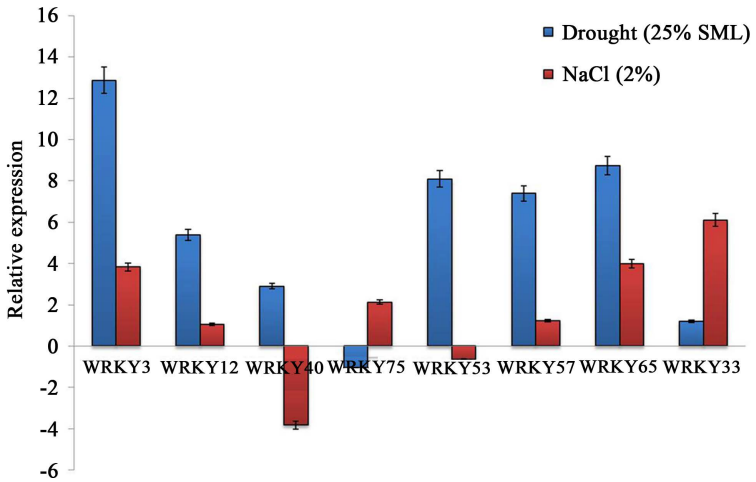


Figure 5. qRT-PCR analysis of WRKY TFs in horsegram plants after imposing drought and salt stress. The relative expression of WRKY TF was more under drought stress conditions.

detected under abiotic stress conditions like high and low temperature, NaCl and drought in transgenic tobacco lines [26]. So we speculate that *WRKY*3 genes might be involved in stress response and tolerance under drought. All eight WRKY TFs studied are up-regulated in drought stress except the *WRKY*75 whose mRNA transcript levels are found to be down-regulated by one fold. In salt stress conditions the expression of *WRKY*33 is more with 6.1 folds increase in the mRNA transcript levels and two genes WRKY 40 and *WRKY* 53 were found to be down-regulated. In earlier reports, the overexpression of *DgWRKY*1 from chrysanthemum has shown better tolerance to salt stress in transgenic tobacco plants [37].

The stress leads to trigger some of the key enzymes of antioxidant defence system. To resist oxidative damage in plants the antioxidant enzymes and certain metabolites; play a vital role leading to adaptation and the ultimate survival under stress [38]. In the present study, we speculate that the expression of WRKY TFs in turn regulating the expression of other stress related antioxidative genes under drought stress conditions. The expression of antioxidative enzymes enhances the scavenging activity in plants and reduces the ROS produced under stress. The low levels of MDA in drought stress conditions also represent the less lipid peroxidation and membrane damage. The overexpression of *TaWRKY*10 has been reported earlier to enhance the expression level of stress related antioxidant enzyme like SOD, CAT, and GPX in transgenic tobacco plant. The expression of *FcWRKY*40 altering the expression levels of two ROS scavenging genes like SOD and POD were reported previously [13]. Our results suggest that the up-regulation of WRKY TF genes contributes to plant defence under drought and salt stress conditions. But the genes we studied are upregulated well in drought stress conditions than in salt stress conditions.

5. Conclusion

Based on the result obtained from this study we speculate that horsegram WRKY TF

genes are promising to stress responsible gene for modulation and enhancement of abiotic stress tolerance in sensitive crops. In this study, an attempt was made to identify the abiotic stress responsive WRKY TF genes from horsegram plant under various abiotic stress conditions. Furthermore, studies will be carried out to develop abiotic stress tolerant transgenic crops by genetic manipulation of WRKY TF genes and assess the function of these genes in abiotic stress tolerance.

Acknowledgements

Financial support from Council of Scientific and Industrial research (CSIR: 38(1305)/11/EMR-11), Department of Science and Technology (DST: SR/SO/PS/0001/2011) and University Grant Commission, New Delhi are greatly acknowledged.

References

[1] Wang, Y. and Frei, M. (2011) Stressed Food—The Impact of Abiotic Environmental Stresses on Crop Quality. *Agriculture, Ecosystems & Environment*, **141**, 271-286. http://dx.doi.org/10.1016/j.agee.2011.03.017

[2] Khelil, A., Menu, T. and Ricard, B. (2007) Adaptive Response to Salt Involving Carbohydrate Metabolism in Leaves of a Salt-Sensitive Tomato Cultivar. *Plant Physiology and Biochemistry*, **45**, 551-559. http://dx.doi.org/10.1016/j.plaphy.2007.05.003

[3] Wang, W., Vinocur, B. and Altman, A. (2003) Plant Responses to Drought, Salinity and Extreme Temperatures: Towards Genetic Engineering for Stress Tolerance. *Planta*, **218**, 1-14. http://dx.doi.org/10.1007/s00425-003-1105-5

[4] Tran, L.S.P., Nakashima, K., Sakuma, Y., Simpson, S.D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2004) Isolation and Functional Analysis of *Arabidopsis* Stress-Inducible NAC Transcription Factors That Bind to a Drought Responsive Cis-Element in the Early Responsive to Dehydration Stress 1 Promoter. *Plant Cell*, **16**, 2481-2498. http://dx.doi.org/10.1105/tpc.104.022699

[5] Shinozaki, K., Yamaguchi-Shinozaki, K. and Seki, M. (2003) Regulatory Network of Gene Expression in the Drought and Cold Stress Responses. *Current Opinion in Plant Biology*, **6**, 410-417. http://dx.doi.org/10.1016/S1369-5266(03)00092-X

[6] Eulgem, T., Rushton, P.J., Robatzek, S. and Somssich, I.E. (2000) The WRKY Superfamily of Plant Transcription Factors. *Trends in Plant Science*, **5**, 199-206. http://dx.doi.org/10.1016/S1360-1385(00)01600-9

[7] Yang, B., Jiang, Y.Q., Rahman, M.H., Deyholos, M.K. and Kav, N.N.V. (2009) Identification and Expression Analysis of WRKY Transcription Factor Genes in Canola (*Brassica napus* L.) in Response to Fungal Pathogens and Hormone Treatments. *BMC Plant Biology*, **9**, 68. http://dx.doi.org/10.1186/1471-2229-9-68

[8] Pan, L.J. and Jiang, L. (2014) Identification and Expression of the *WRKY* Transcription Factors of *Carica papaya* in Response to Abiotic and Biotic Stresses. *Molecular Biology Reports*, **41**, 1215-1225. http://dx.doi.org/10.1007/s11033-013-2966-8

[9] Johnson, C.S., Kolevski, B. and Smyth, D.R. (2002) TRANSPARENT TESTA GLABRA2, a Trichome and Seed Coat Development Gene of *Arabidopsis*, Encodes a WRKY Transcription Factor. *The Plant Cell*, **14**, 1359-1375. http://dx.doi.org/10.1105/tpc.001404

[10] Pnueli, L., Hallak-Herr, E., Rozenberg, M., Cohen, M., Goloubinoff, P., Kaplan, A. and Mittler, R. (2002) Molecular and Biochemical Mechanisms Associated with Dormancy and

Drought Tolerance in the Desert Legume *Retama raetam*. *Plant Journal*, **31**, 319-330. http://dx.doi.org/10.1046/j.1365-313X.2002.01364.x

[11] Zheng, Z., Qamar, S.A., Chen, Z. and Mengiste, T. (2006) Arabidopsis WRKY33 Transcription Factor Is Required for Resistance to Necrotrophic Fungal Pathogens. *Plant Journal*, **48**, 592-605. http://dx.doi.org/10.1111/j.1365-313X.2006.02901.x

[12] Mare, C., Mazzucotelli, E., Crosatti, C., Francia, E., Stanca, A.M. and Cattivelli, L. (2004) Hv-WRKY38: A New Transcription Factor Involved in Cold- and Drought-Response in Barley. *Plant Molecular Biology*, **55**, 399-416. http://dx.doi.org/10.1007/s11103-004-0906-7

[13] Gong, X., Hu, J. and Liu, J. (2014) Cloning and Characterization of FcWRKY40, A WRKY Transcription Factor from *Fortunella crassifolia* Linked to Oxidative Stress Tolerance. *Plant Cell Tissue and Organ Culture*, **119**, 197-210. http://dx.doi.org/10.1007/s11240-014-0526-0

[14] Li, H., Gao, Y., Xu, H., Dai, H., Deng, D. and Chen, J. (2013) ZmWRKY33, a WRKY Maize Transcription Factor Conferring Enhanced Salt Stress Tolerances in Arabidopsis. *Plant Growth Regulation*, **70**, 207-216. http://dx.doi.org/10.1007/s10725-013-9792-9

[15] Garg, N. and Manchanda, G. (2009) ROS Generation in Plants: Boon or Bane? *Plant Biosystems*, **143**, 8-96. http://dx.doi.org/10.1080/11263500802633626

[16] Montillet, J.L., Chamnongpol, S., Rusterucci, C., Dat, J., van de Cotte, B., Agnel, J.P., Battesti, C., Inze, D., Van Breusegem, F. and Triantaphylides, C. (2005) Fatty Acid Hydroperoxides and H_2O_2 in the Execution of Hypersensitive Cell Death in Tobacco Leaves. *Plant Physiology*, **138**, 1516-1526. http://dx.doi.org/10.1104/pp.105.059907

[17] Yannarelli, G.G., Noriega, G., Batlle, A. and Tomaro, M. (2006) Heme Oxygenase Up-Regulation in Ultraviolet-B Irradiated Soybean Plants Involves Reactive Oxygen Species. *Planta*, **224**, 1154-1162. http://dx.doi.org/10.1007/s00425-006-0297-x

[18] Romero-Puertas, M.C., Rodriguez-Serrano, M., Corpas, F.J., Gomez, M., Del Rio, L.A. and Sandalio, L.M. (2004) Chadmium-Induced Subcellular Accumulation of $O_2^{\cdot-}$ and H_2O_2 in Pea Leaves. *Plant Cell & Environment*, **27**, 1122-1134. http://dx.doi.org/10.1111/j.1365-3040.2004.01217.x

[19] Hodges, D.M., Delong, J.M., Forney, C.F. and Prange, R.K. (1999) Improving the Thiobarbitric Acid Reactive Substance Assay for Estimating Lipid Peroxidation in Plant Tissues Containing Anthocyanin and Other Interfering Compounds. *Planta*, **207**, 604-611. http://dx.doi.org/10.1007/s004250050524

[20] Hu, W., Yuan, Q., Wang, Y., Cai, R., Deng, X., Wang, J., Zhou, S., Chen, M., Chen, L., Huang, C., Ma, Z., Yang, G. and He, G. (2012) Overexpression of a Wheat Aquaporin Gene, *TaAQP*8, Enhances Salt Stress Tolerance in Transgenic Tobacco. *Plant and Cell Physiology*, **53**, 2127-2141. http://dx.doi.org/10.1093/pcp/pcs154

[21] Livak, K.J. and Schmittgen, T.D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods*, **25**, 402-408. http://dx.doi.org/10.1006/meth.2001.1262

[22] Munns, R. (1993) physiological Process Limiting Plant Growth Saline Soils Some Dogmas and Hypotheses. *Plant Cell & Environment*, **16**, 15-24. http://dx.doi.org/10.1111/j.1365-3040.1993.tb00840.x

[23] Chen, J., Shiyab, S., Han, F.X., Monts, D.L., Waggoner, A.W. and Su, Z.Y. (2009) Bioaccumulation and Physiological Effects of Mercury in *Pteris vittata* and *Nephrolepis exaltata*. *Ecotoxicology*, **18**, 110-121. http://dx.doi.org/10.1007/s10646-008-0264-3

[24] Wang, X. and Han, J. (2009) Changes of Proline Content, Activity, and Active Isoforms of

Antioxidative Enzymes in Two Alfalfa Cultivars under Salt Stress. *Agriculture Sciences in China*, **8**, 431-440. http://dx.doi.org/10.1016/S1671-2927(08)60229-1

[25] Nareshkumar, A., Nagamallaiah, G.V., Pandurangaiah, M., Kiranmai, K., Amarana-thareddy, V., Lokesh, U., Venkatesh, B. and Sudhakar, C. (2015) Pb-Stress Induced Oxidative Stress Caused Alterations in Antioxidant Efficacy in Two Groundnut (*Arachis hypogaea* L.) Cultivars. *Agricultural Sciences*, **6**, 1283-1297. http://dx.doi.org/10.4236/as.2015.610123

[26] Zhu, Z., Shi, J., Cao, J., He, M. and Wang, Y. (2012) *VpWRKY*3, a Biotic and Abiotic Stress-Related Transcription Factor from Chinese Wild *Vitis pseudoreticulata*. *Plant Cell Reports*, **11**, 2109-2120. http://dx.doi.org/10.1007/s00299-012-1321-1

[27] Jing, S.J., Zhou, X., Song, Y. and Yu, D.Q. (2009) Heterogonous Expression of *OsWRKY*23 Gene Enhances Pathogen Defence and Dark Induced Leaf Senescence in Arabidopsis. *Plant Growth Regulation*, **58**, 181-190. http://dx.doi.org/10.1007/s10725-009-9366-z

[28] Yang, P., Chen, C., Wang, Z., Fan, B. and Chen, Z. (1999) A Pathogen- and Salicylic Acid-Induced WRKY DNA-Binding Activity Recognizes the Elicitor Response Element of the Tobacco Class I Chitinase Gene Promoter. *Plant Journal*, **18**, 141-149. http://dx.doi.org/10.1046/j.1365-313X.1999.00437.x

[29] Ishiguro, S. and Nakamura, K. (1994) Characterization of a cDNA Encoding a Novel DNA-Binding Protein, SPF1, That Recognizes SP8 Sequences in the 5' Upstream Regions of Genes Coding for Sporamin and β-Amylase from Sweet Potato. *Molecular and General Genetics*, **244**, 563-571. http://dx.doi.org/10.1007/BF00282746

[30] Rushton, P.J., Torres, J.T., Parniske, M., Wernert, P., Hahlbrock, K. and Somssich, I.E. (1996) Interaction of Elicitor-Induced DNA-Binding Proteins with Elicitor Response Elements in the Promoters of Parsley PR1 Genes. *EMBO Journal*, **15**, 5690-5700.

[31] De Pater, S., Greco, V., Pham, K., Memelink, J. and Kijne, J. (1996) Characterization of a Zinc-Dependent Transcriptional Activator from Arabidopsis. *Nucleic Acids Research*, **24**, 4624-4631. http://dx.doi.org/10.1093/nar/24.23.4624

[32] Oh, S.K., Baek, K.H., Park, J.M., Yi, S.Y., Yu, S.H., Kamoun, S. and Choi, D. (2008) Capsicum Annuum WRKY protein *CaWRKY*1 Is a Negative Regulator of Pathogen Defence. *New Phytologist*, **177**, 977-989. http://dx.doi.org/10.1111/j.1469-8137.2007.02310.x

[33] Cai, M., Qiu, D., Yuan, T., Ding, X., Li, H., Duan, L., Xu, C., Li, X. and Wang, S. (2008) Identification of Novel Pathogen-Responsive Cis-Elements and Their Binding Proteins in the Promoter of *OsWRKY*13, a Gene Regulating Rice Disease Resistance. *Plant Cell and Environment*, **31**, 86-96.

[34] Yu, F.F., Huaxia, Y.F., Lu, W.J., Wu, C. and Guo, X.Q. (2012) *GhWRKY*15, a Member of the WRKY Transcription Factor Family Identified from Cotton (*Gossypium hirsutum* L.), Is Involved in Disease Resistance and Plant Development. *BMC Plant Biology*, **12**, 144. http://dx.doi.org/10.1186/1471-2229-12-144

[35] Jiang, Y.J., Liang, G. and Yu, D.Q. (2012) Activated Expression of WRKY57 Confers Drought Tolerance in Arabidopsis. *Molecular Plant*, **5**, 1375-1388. http://dx.doi.org/10.1093/mp/sss080

[36] Cai, R., Zhao, Y., Wang, Y., Lin, Y., Peng, X., Li, Q., Chang, Y., Jiang, H., Xiang, Y. and Cheng, B. (2014) Overexpression of a Maize *WRKY*58 Gene Enhances Drought and Salt Tolerance in Transgenic Rice. *Plant Cell Tissue and Organ Culture*, **119**, 565-577. http://dx.doi.org/10.1007/s11240-014-0556-7

[37] Liu, Q., Xu, K., Pan, Y., Jiang, B., Liu, G., Jia, Y. and Zhang, H. (2014) Functional Analysis of a Novel Chrysanthemum WRKY Transcription Factor Gene Involved in Salt Tolerance.

Plant Molecular Biology Reports, **32**, 282-289.
http://dx.doi.org/10.1007/s11105-013-0639-3

[38] Zhang, F.Q., Wang, Y.S., Lou, Z.P. and Dong, J.D. (2007) Effect of Heavy Metal Stress on Antioxidative Enzymes and Lipid Peroxidation in Leaves and Roots of Two Mangrove Plant Seedlings (*Kandelia candel* and *Bruguiera gymnorrhiza*). *Chemosphere*, **67**, 44-50.
http://dx.doi.org/10.1016/j.chemosphere.2006.10.007

Scientific Research Publishing

American Journal of Molecular Biology, 2016, 6, 138-158
http://www.scirp.org/journal/ajmb
ISSN Online: 2161-6663
ISSN Print: 2161-6620

The Potential of DNA Barcode-Based Delineation Using Seven Putative Candidate Loci of the Plastid Region in Inferring Molecular Diversity of Cowpea at Sub-Species Level

Patrick Okoth[1], John Muoma[2], Mulaya Emmanuel[3], Wekesa Clabe[4], Dennis O. Omayio[2], Paul O. Angienda[1]

[1]Department of Zoology, School of Physical and Biological Sciences, Maseno University, Maseno, Kenya
[2]Department of Biological Sciences, Masinde Muliro University of Science and Technology, Kakamega, Kenya
[3]Centre for Global Health Research, Kenya Medical Research Institute (KEMRI), Kisumu, Kenya
[4]Department of Biochemistry and Biotechnology, Kenyatta University, Nairobi, Kenya
Email: okothpatrick@mmust.ac.ke

How to cite this paper: Okoth, P., Muoma, J., Emmanuel, M., Clabe, W., Omayio, D.O. and Angienda, P.O. (2016) The Potential of DNA Barcode-Based Delineation Using Seven Putative Candidate Loci of the Plastid Region in Inferring Molecular Diversity of Cowpea at Sub-Species Level. *American Journal of Molecular Biology*, **6**, 138-158.
http://dx.doi.org/10.4236/ajmb.2016.64014

Received: September 12, 2016
Accepted: October 10, 2016
Published: October 13, 2016

Abstract

The novelty and suitability of the mitochondrial gene *CO*1 in DNA barcoding as a reliable identification tool in animal species are undisputed. This is attributed to its standardized sequencing segment of the mitochondrial cytochrome c oxidase-1 gene (*CO*1) which has the necessary universality and variability making it a generally acceptable barcode region. CO1 is a haploid single locus that is uniparentally-inherited. Protein-coding regions are present in high-copy numbers making it an ideal barcode. The mitochondrial oxidase subunit I (*COI*) gene is a robust barcode with a suitable threshold for delineating animals and is not subject to drastic length variation, frequent mononucleotide repeats or microinversions. However, a low nucleotide substitution rate of plant mitochondrial genome [mtDNA] precludes the use of *CO*1 as a universal plant DNA barcode and makes the search for alternative barcode regions necessary. Currently, there exists no universal barcode for plants. The plastid region reveals leading candidate loci as appropriate DNA barcodes yet to be explored in biodiversity studies in Kenya. Four of these plastid regions are portions of coding genes (*matK*, *rbcL*, *rpoB*, and *rpoC*1), and three noncoding spacers (*atpF-atpH*, *trnH-psbA*, and *psbK-psbI*) which emerge as ideal candidate DNA loci. While different research groups propose various combinations of these loci, there exists no consensus; the lack thereof impedes progress in getting a suitable universal DNA barcode. Little research has attempted to investigate and document the applicability and extend of effectiveness of different DNA regions as barcodes to delineate cowpea at subspecies level. In this study we sought to test feasibility of the seven putative

candidate DNA loci singly and in combination in order to establish a suitable single and multi-locus barcode regions that can have universal application in delineating diverse phylogeographic groups of closely related Kenyan cowpea variants. In this study, our focus was based on genetic parameters including analyses of intra- and infra-specific genetic divergence based on intra- and infra-specific K2P distances; calculation of Wilcoxon signed rank tests of intra-specific divergence among loci and coalescence analyses to delineate independent genetic clusters. Knowledge of DNA candidate loci that are informative will reveal the suitability of DNA barcoding as a tool in biodiversity studies. Results of this study indicate that: *matK*, *trnH-psbA*, *psbK-psbL*, and *rbcL* are good barcodes for delineating intra and infraspecific distances at single loci level. However, among the combinations, *matK + trnH-psbA*, *rpoB + atpF-atpH + matK* are the best barcodes in delineating cowpea subvariants. *rbcL* gene can be a suitable barcode marker at single locus level, but overall, multi locus approach appears more informative than single locus approach. The present study hopes to immensely contribute to the scanty body of knowledge on the novelty of DNA barcoding in cataloguing closely related cowpea variants at molecular level and hopes to open up future research on genomics and the possibility of use of conserved regions within DNA in inferring phylogenetic relationships among Kenyan cowpea variants.

Keywords

DNA Barcoding, Plastid Region, DNA Sequencing, Intergenic Spacers, cp DNA, Molecular Phylogenetics, Intraspecific, Infraspecific

1. Background

DNA-barcoding is a technique used for the taxonomic characterization and phylogenetic analysis of organisms and entails the use of defined regions within the DNA genetic material of an organism, which though exposed to evolution mechanism, is conserved between and within the species. This region serves as a tool to uniquely identify two individuals with unique ancestral-lineage. DNA barcoding is a sequence-based identification system that may be constructed of one or several loci taken together as a complementary unit in delineating relationships and inferring patterns of change among related organisms. It employs short highly variable regions of the genome to delineate organism that are closely related. In animals including human, cytochrome oxidase I (coI-gene) has vastly been utilized as a unique barcoding region for phylogenetic analysis. The mitochondrial coxidase subunit I (COI) gene has demonstrated significant reliability and recoverability notwithstanding its limited application in the plant kingdom [1]. The Consortium for the Barcode of Life (CBOL) plant-working group proposes seven Barcode regions for use as barcoding markers [2]. The feasibility of DNA barcoding and the use of plastid regions in biodiversity studies can be an important tool of utility at molecular level. Plastid DNA candidate loci are universally present and conserved in the plant target lineages and can provide a rapid and reproducible

molecular identification away from the Linnaean system of nomenclature. The informativeness of each barcode region was therefore explored singly and in combination with a view to assessing the feasibility of these candidate loci in infra-specific and intra-specific discrimination of phylogeographic groups among Kenyan cowpea variants. This was informed by the fact that chloroplast genes exist in a single copy and are conserved among plant eukaryotic genomes. These regions undergo limited mutation overtime and are considered novel in revealing evolutionary divergences and ancestral lineages within species. A single and multilocus approach was explored using seven chloroplast genic candidate DNA regions (*rbcL* gene, *atpF-atpH* spacer, *matK* gene, *rpoB* gene, *rpoC*1 gene, trnH-*psbA* spacer and psbK-*psbI* spacer.

2. Single-Locus DNA Barcode Typing

2.1. *MatK*

matK is considered an important tool in plant evolutionary studies and systematics. matK gene loci is the only putative group II intron maturase encoded in the chloroplast genome of plants and is the only plastid gene containing this putative maturase domain in higher plants [3]. *matK* is Maturase-Kinase gene, a plastid gene responsible for the chloroplast post-transcriptional processing. It has an unusual evolutionary tempo, with relatively high substitution rates at both nucleotide and amino acid levels according to [4]. The strong phylogenetic signal from *matK* gene renders it invaluable gene loci in plant systematics and evolutionary studies at various evolutionary depths. This gene is proposed as the only chloroplast-encoded group II intron maturase, and is suggested to play a role in chloroplast post-transcriptional processing. The ability of Maturase kinase sequence to solely work as a barcoding candidate was assayed together with the six other candidates in the present study to delineate distance characterization. matK has in the recent past emerged as an invaluable locus in plant biodiversity and systematics based on its highly informative ability to decode phylogenetic distinctiveness unlike other candidate loci [5]. This genelocus has 1500 base pair nested in the group II intron of the 50 and 30 exons of trnK in the large single copy region of the chloroplast genome of green plants. matK gene sequence is one of the seven putative gene loci widely utilized in the DNA barcoding of land plants [6]. Phylogenetic analysis based on matK, against other candidate genes has demonstrated excellent parsimony informative characters with significantly more phylogenetic structure pereach parsimony-informative site contrary to the highly conserved chloroplast/plastid region. matK sequence information has been reported to generate robust phylogenies [7] and is considered to have reliable evolutionary rate, suitable length and good inter specific divergence as well as a low transversion rate [8]. *matK* is however difficult to amplify universally demonstrating that *matK* barcode albeit informative, may be inadequate and inconclusive when used in isolation as a universal barcode. This study therefore considered matK alongside other six barcode.

2.2. *rbcL*

A region of the chloroplast gene *rbcL*—RuBisCo large subunit has been considered an

ideal candidate barcode region in plants and is famed as the most abundant protein on earth. The region, RuBisCo (Ribulose-1, 5-bisphosphate carboxylase oxygenase) is used in the catalysis of the first step of carbon fixation and is a target region in phylogenetic investigations due to its easy amplification, sequencing and alignment. Many taxonomists consider *rbcL* gene as ideal DNA barcoding region for plants at both family and generic level. However, *rbcL* sequences have the limitation of slow evolution. The rbcL locus has the lowest divergence of plastid genes in flowering plants according to [1]. Studies by (CBOL Plant Working Group [9]-[12] report modest discriminatory power of this locus. Other studies however indicate that rbcL remains one of the best candidate barcodes based on the straightforward recovery of the gene sequence, easy accessibility and discriminatory power [13].

2.3. *trnH-psbA*

The *trnH-psbA* region is a straightforward region easily amplifiable across land plants, and is one of the most variable intergenicspacers [14]. It has been used successfully in a range of barcoding studies. [15] report that the *trnH-psbA*, non-coding intergenic region exhibits significant sequence divergence with notable insertion/deletion rates. Studies by [16] indicate that this plastid region has highly conserved coding sequences that makes it an attractive marker. These attributes make *trnH-psbA* an important plant barcode for species discrimination [15]. However, the complex molecular evolution and considerable length variation of *trnH-psbA* limits it as a barcode singly [17]. However, trnH-*psbA* is reported to suffer high rates of insertion or deletion in larger families of angiosperms. The trnH-*psbA* putative gene loci albeit a standard barcode region in most plants has been reported to suffer frequent inversions in some lineages of plants and singly as a barcode marker, may result to overestimation of genetic divergence and consequently inaccurate assignment of phylogenetic position [18].

2.4. *atpF-atpH*

The second International Barcode of Life Conference proposes that at pF-atpH intergenic spacer is a potential plant barcode region [9]. The fact that atpF-atpH marker has not been widely used in studies of plant systematic and phylogeographics has led to paucity of data on its performance as barcodes. However, the CBOL Plant Working Group indicate that atpF-atpH has relatively modest discriminatory power, intermediate sequence quality and universality and could be used as a plant DNA barcode. Recent studies document positive reports on the performance of atpF-atpH as a plant barcode region [19]. Ki-Joong Kim, pers. Comm reports usefulness of atpF-atpH on the Korean flora biodiversity studies. Studies on duckweeds [20] also demonstrated that *atpF-atpH*, a noncoding spacer could serve as a universal DNA barcoding marker for species-level identification. In their study, the utility of this non cording region in identification of new species by reason of its ease of amplification, straightforward sequence alignment and rates of DNA variation was reported [20]. In the same study, it's documented that DNA barcoding made significant contribution to the taxonomical structure in duckweeds as opposed to the less informative morphological classification and

therefore recommends *atpF-atpH* as an important barcode region in biodiversity studies. The current study therefore seeks to among others test the in formativeness of this barcode region in delineating cowpea diversity at sub-species level.

2.5. Multi-Locus Candidate DNA Barcode Typing [MLA]

Lack of adequate variation within single loci makes it difficult to get a universal plant barcode comparable to CO1 in animal species [11]; [12]; [15]; [21]-[24]. Amulti-locus approach has been suggested as ideal in delineating species [9]; [12]; [15]; [25]-[28]. Various combinations of plastid loci have been proposed by many studies because combined barcodes exhibit satisfactory discrimination as opposed to single-locus approaches; *rbcL* + *trnH-psbA* [15], or *rpoC*1 + *matK* + *trnH-psbA* or *rpoC*1 + *rpoB* + *matK* [21] and *matK* + *atpF-atpH* + *psbK-psbL* or *matK* + *atpFatpH* + *trnH-psbA* [29]. Previous studies by [15] support the earlier observation that trnH-psbA coupled with rbcL can correctly delineate and discriminate among related species. In the same study, a combination of the non-coding trnH-psbA spacer region and a portion of the coding rbcL locus are considered ideal two-locus global land plant barcode that provides the necessary universality and species discrimination that meets good threshold of CBoL. The Consortium for the Barcode of Life-Plant Working Group (CBOL) recommends a two-locus combination of *matK* and *rbcL* as suitable plant barcode with a discriminatory efficiency of 72% [9]. A multi-locus approach has been suggested in typing plant species by taxonomists [12]; [21]; [30]; [31]. However, CBOL demonstrated that multiple loci approach did not clearly improve the species-level discrimination. Accordingly whole-plastid genome sequence has been suggested by other studies as appropriate in plant identification [32]-[34]. Overall, a MLA exhibit higher discriminatory ability as opposed to single-locus barcodes in most studies. Accordingly, several combinations of plastid loci have been suggested among them *rbcL* + *trnH-psbA* [15], rpoC1 + *matK* + *trnH-psbA* or rpoC1 + *rpoB* + *matK* [18] and *matK* + *atpF-atpH* + *psbK-psbL* or *matK* + *atpFatpH* + *trnH-psbA* [29]. The current study seeks to test this at sub species level. In this study, we sought to test these combinations in delineating cowpea accessions at varietal level. An investigation of the relevance of DNA barcoding to correctly delineate and discriminate between closely related cowpea variants is therefore presented for biodiversity analysis and to evaluate the overall utility of chloroplast DNA barcode candidates in reconstructing their ancestry.

3. Statistical Analysis

BIOEDIT Software was used to analyze raw sequence data chromatograms by trimming and assembling. MEGA 6.06 software was then used for multiple sequence alignment. The UPGMA was used for clustering and generating an accurate topology with reference to the molecular clock based on the greatest similarity amongst pairs [35]. Distance estimation model used the Kimura 2 Parameter (K2P) model. This was because K2P distances are best when distance is low especially in highly similar sequences [36]. This was further informed by the fact that our sequences were from closely related

cowpea sub variants. The K2P distances generated were exported into an Excel format which was uploaded into GraphPad software and STATA for further statistical analysis. GraphPad Prism v. 7 was then used to plot the histograms on Intra-specific and Infra-specific distances after importing the excel file containing the K2P distances. Wilcoxon Signed Ranks test was performed to separately compare the intra-specific and infra-specific distances of between markers at single and multi-locus level. The choice of Wilcoxon test was informed by the fact that data was non-parametric and did not assume a normal Gaussian distribution [37]. STATA 13 was then used to perform the Moods median Test and the Wilcoxon Two Sample Test for **Table 7**. Moods median test was used because the data was not normally distributed and yet the mean would not be a good representative of "location" [37]; hence the need to employ a statistical tool which utilized the median as a better estimator of "location". The Moods Median test was then preferred to the Sign test. The Wilcoxon two sample tests acted as a quality control for the median test by comparing the *P*-values of both for significance.

4. Materials and Methods

4.1. DNA Extraction, PCR Amplification, and Sequencing

57 cowpea accessions from different phylogeographic locations of Kenya deposited in the National Gene Bank of Kenya were used for this study. The accessions were collections from different agro-ecological zones of Kenya. Seeds were planted in the green house at Masinde Muliro University of Science and Technology under strict conditions. Young leaves were sampled eight days after planting in 1.5 mL Eppendorf tubes. They were immediately frozen in liquid nitrogen and stored at −80˚C. Leave samples were then manually ground using a micropestle. DNA quality and quantity control was done using Nano-drop spectrophotometer. DNA was normalized by adjusting its concentration to 25 ngμL in an optical 96-well Reaction plates using sterile de-ionized water. Total DNA was extracted using Qiagen plant DNA extraction protocol as per the manufacturer's guidelines with slight modifications. Total DNA was extracted from each sample and quantified using the DNA Nano Drop ND-1000 by Thermo Fisher at the Kenya Medical Research Institute [KEMRI-Kisumu Kenya]. This was followed by amplification of three intergenic spacers' *atpF-atpH*, *psbK-psbI* and *trnH-psbA* and four genes: *matK*, *rbcL*, *rpoB*, *rpoC*1using the primers as shown in **Table 1**. The amplicons were resolved on 3% agarose gel at 80 V for 48 minutes. The gels were observed for bands using a UV trans-illuminator (FotoDyne model 3 3500 Foto-Prep). Photographs of the bands were taken using the software "Strata-gene Eagle View" that was integrated with the digital camera on the UV trans-illuminator. The bands containing DNA of interest which is 400 bp long was excised and the DNA purified using the DNA purification kit from Qiagen. Sequenced products were then analyzed using an automatic sequencer, ABI3730XL (Applied Biosystems). The sequence chromatograms were analyzed and the terminals were trimmed using the BioEdit software (Thomas Hall & Abbott). Similarity searches were conducted based on Basic Local Alignment Search Tool (BLAST) located at www.ncbi.nih.gov/blast; with the parameters set as follows: database-non redundant;

Table 1. List of cpDNA genes/intergenic spacers amplified in the present study including primers and approximate amplicon lengths.

Genes/intergenic pacers	Primer pair (5′-3′)	Amplicon length	T_a (˚C)	Source
atpF-atpH	ACTCGCACACACTCCCTTTCC GCTTTTATGGAAGCTTTAACAAT	621 bp	48˚C	Ki-Joong Kim; kimkj@KOREA.AC.KR
rpoc1	GGCAAAGAGGGAAGATTTCG CCATAAGCATATCTTGAGTTGG	490 bp	53˚C	http://www.kew.org/barcoding/protocols.html
rpoB	ATGCAACGTCAAGCAGTTCC CCGTATGTGAAAAGAAGTATA	490 bp	51˚C	http://www.kew.org/barcoding/protocols.html
matK	CGTACAGTACTTTTGTGTTTACGAG ACCCAGTCCATCTGGAAATCTTGGTTC	892 bp	49.5˚C	Ki-Joong Kim; kimkj@KOREA.AC.KR
psbK-psbI	TTAGCCTTTGTTTGGCAAG AGAGTTTGAGAGTAAGCAT	576 bp	60˚C	Ki-Joong Kim; kimkj@KOREA.AC.KR
rbcL	GTAAAATCAAGTCCACCRCG ATGTCACCACAAACAGAGACTAAAGC	596 bp	50˚C	David Erickson; ERICKSOND@si.edu
trnH-psbA	GTTATGCATGAACGTAATGCTC CGCGCATGGTGGATTCACAATCC	812 bp	50˚C	David Erickson; ERICKSOND@si.edu

search-megablast and the expectant value set at 10^{-9}. The sequence that had the lowest expectant value (E-value) and the highest identity score was considered to be a similar sequence. The NCBI taxonomy tool (https://www.ncbi.nih.gov/taxonomy) was then used to determine the complete classification of the sequence which was confirmed in the International Plant Names Index website (https://www.ipni.org/ipni/plantnamessearchpage.do). The sequences were then named as per the molecular characterization. Three subspecies were identified namely: *Vigna unguiculata* (L.) *Walp*. Subsp. *Cylindrica*; *Vigna unguiculata var. serotina Bertoni*; *Vigna unguiculata subvar. Deflexa Bertoni*. Also included in our analysis were two other species namely *Rhynchosia minima* and *Vigna-luteola* to act as out groups for the phylogenetic analysis. The sequences were then assembled into a single Fasta file format in BioEdit software version 7 (Thomas Hall & Abbott). This was followed by performing a local alignment using muscle in Mega 6 software with UPGMA as the clustering method. The intra-specific and infra-specific distances were determined using Kimura-2-parameters in mega 6. Since all the sequences were from the genus *Vigna unguiculata* but different sub-variants; therefore in Mega 6; the intraspecific distances were determined as "between the various group mean distances" and the infraspecific sequences inferred based on "within group mean distance".

4.2. Sequence Alignment, Analysis and Amplification Efficiency

Consensus sequences were generated and sequences of the candidate DNA barcodes aligned using muscle. Genetic distance matrices were calculated on the basis of Kimura 2-Parameter (K2P) substitution model for the seven chloroplast candidate DNA loci and the average values between subpopulations inferred. Combined DNA barcode sequences showed significant intra specific but low infra-specific variation rates (**Table 3** and **Table 4**). Average infra and intra-specific distance, mean theta and coalescent depth were calculated to determine infra and intra-specific variation [**Table 3** and **Table 4**]. Wilcoxon signed-rank tests were performed. The distribution of intra-specific versus infra specific variability was evaluated by assessment of the DNA barcoding

gaps. The average intra-specific distance, mean theta, and coalescent depth were calculated to evaluate the intra-specific variation [Table 3 and Table 4]. Wilcoxon signed-rank tests were used [Table 5 and Table 6]. Infra- and intra-specific genetic divergences were calculated based on each putative candidate loci. To characterize intra-specific divergence it was necessary to invoke three different metrics. Genetic distances between cowpea variants were used to characterize intra-specific divergence. For each barcode, pairwise distances were calculated with the simplest K2P model followed by Wilcoxon Signed Rank Tests to compare infra- and intra-specific variability for every barcodes following Kress and Erickson [Table 5 and Table 6. DNA barcoding gaps were evaluated by comparing the distribution of infra- versus intra-specific divergences. Median and Wilcoxon Two-Sample Tests were used to evaluate any overlaps in the distributions with a view to establishing a suitable single and multilocus barcode for cowpea at sub-species level [Table 6].

5. Results and Discussion

5.1. DNA Barcoding Success and Levels of Variability

Overall, PCR amplifications were largely successful and any low quality sequences and dubious amplicons were excluded from the analyses [Table 2]. Similar observations were made by [11]; [15]. However, all the primer pairs designed for each DNA region proved highly successful [Table 1].

5.2. Evaluating the Feasibility of Using DNA Barcodes in Delineating Subspecies of Cowpea

Infra- and Intra-Specific Diversities

Performances of each of the seven candidate DNA barcode loci was assessed by means of intra- and infra-specific diversity calculated from K2P (Kimura's two parameters) pairwise distance matrices [12]. The highest intraspecific diversity was reached by rbcL [Table 3] followed by matK and psbL-psbK respectively. However, the lowest intraspecific distance was reported by rpoB [Table 3 and Table 4]. The mean coalescent depth was slightly superior to the average of overall intraspecific distances because it takes into consideration only the highest distance. Results showed the highest mean of intraspecific differences was recorded by rbcL (Table 4).

Three different metrics were used to characterize intra and infra-specific divergence between average of all the pairwise distances between all individuals sampled within the samples and mean theta with theta being the average pairwise distances calculated for each sample and the coalescent depth [Table 3 and Table 4]. To test the applicability of

Table 2. Sequence analysis and PCR amplification performance efficiency of seven candidate plastid DNA regions (atpF-atpH spacer, matK gene, rbcL gene, rpoB gene, rpoC1 gene, psbK-psbI spacer, and trnH-psbA spacer).

	matK	psbK-psbL	trnH-psbA	atpF-atpH	rpoB	rpoC1	rbcL
Success of Sequencing (%)	100	100	100	100	100	100	100
Amplification Success (%)	96%	86%	96%	94%	92%	92%	93%

Table 3. Measurements of intra- and infra-specific K2P distances matrices for four potential barcode regions, three intergenic spacers and multi locus combinations

	matK	trnH-psbA	atpF-atpH	psbK-psbL	4 loci	matK + trnH-psbA	matK + atpF-atpH + trnH-psbA	matK + psbK-psbL + trnH-psbA	matK + atpF-atpH	matK + psbK-psbL	matK + psbK-psbL + atpF-atpH
Mean of all intra-specific distances	0.2776	0.1148	0.1619	0.2377	0.8042	0.8214	0.6915	0.9144	0.3681	0.6992	1.0186
St. deviation±	0.2095	0.089	0.0378	0.2195	0.209	0.076	0.2530	0.2340	0.305	0.423	0.346
Mean of all infra-specific distances	0.0773	0.0116	0.1615	0.1103	0.7878	1.2656	0.9112	0.8594	0.1271	0.7476	0.9266
St. deviation±	0.1155	0.0108	0.1156	0.2129	0.7146	0.4456	0.8282	0.6822	0.1538	0.6130	0.8124
Mean Theta	0.1319	0.1003	0.1902	0.2869	0.7287	0.9241	0.3318	0.8678	0.1689	0.7671	0.8987
St. deviation±	0.0682	0.109	0.0182	0.2435	0.141	0.001	0.068	0.154	0.077	0.164	0.154
Mean coalescent depth	0.0296	0.0153	0.0471	0.036	0.0316	0.0227	0.0123	0.0112	0.1171	0.0134	0.0138
St. deviation±	0.0713	0.0289	0.1157	0.098	0.0621	0.0975	0.0413	0.0399	0.9045	0.0438	0.0461
Number of measurements for all infraspecific distances	344	460	222	196	5121	1853	3323	3300	1325	1216	2543
Number of measurements for all intraspecific distances	1107	980	339	270	9745	3910	6705	6304	2501	2340	4484

Legend: **Table 3** above and **Table 4** below: The intra-specific and infra-specific genetic divergences were calculated for each DNA barcode. Measures used include: the average pair wise distances between all the sampled within subspecies having a minimum of two representatives; the "mean-theta" where theta is the average pair wise distance computed for each species having more than one representative thus eliminating partiality due to uneven sampling among taxa; and the average coalescent depth which is the depth of a node linking all samples. The aforementioned parameters were deemed important in characterizing infra-specific divergence.

Table 4. Measurements of intra- and infra-specific K2P distances matrices for three potential barcode regions, three intergenic spacers and multi locus combinations

	rpoB	rpoC1	rbcL	psbK-psbL + atpF-atpH + matK	trnH-psbA + atpF-atpH	rpoB + atpF-atpH + matK	rpoC1 + atpF-atpH + matK
Mean of all intra-specific distances	0.0969	0.1561	0.5015	1.0186	0.6084	0.6824	0.7514
St. deviation±	0.0591	0.038	0.2494	0.346	0.211	0.508	0.42
Mean of all infra-specific distances	0.0892	0.1505	0.1784	0.92662	0.9138	0.3448	1.2786
St. deviation±	0.1075	0.1077	0.2080	0.81239	0.8469	0.1518	0.6381
Mean Theta	0.1229	0.1589	0.4556	0.8987	0.5179	0.2799	1.0389
St. deviation±	0.0502	0.048	0.0733	0.154	0.116	0.063	0.184
Mean coalescent depth	0.0168	0.0144	0.0147	0.0132	0.0114	0.0131	0.0137
St. deviation±	0.0724	0.0610	0.0705	0.0449	0.0427	0.0450	0.0452
Number of measurements for all infraspecific distances	229	217	419	2543	1304	2636	2462
Number of measurements for all intraspecific distances	366	344	907	4478	2538	4745	4739

the seven loci at single and multi-loci level for sub-species identification, BLAST1 searches and the nearest genetic distance were used (Table 3 and Table 4). The goal was to identify the most informative single locus candidate DNA barcode gene markers that show the best discriminatory power at the varietal level in common cowpea accessions. Our results revealed that at intraspecific level, rbcL [50.15%] possessed the highest identification efficiency among the seven loci followed by matK [27.76%] and *psbK-psbL* [23.77%] respectively at single locus level [Table 2 and Table 3]. The lowest intra-specific variation was reported by rpoB [9.69%]. On the other hand, the highest infraspecific variation was reported by rbcL [17.84%] followed by *atpF-atpH* [16.5%], rpoC1 [15.05%] and subsequently psbK-psbL [11.03%] respectively. The lowest infra-specific variation was reported by trnH-psbA [1.16%]. Overall, the means of infra-specific variation were significantly lower than those of intraspecific variation as expected. To circumvent the challenges associated with single locus approach, this study undertook a multilocus analysis (MLA) as a useful option in delineating closely related cowpea variants based on the following multigenic combination *matK + atpF-atpH*; *matK + trnH-psbA*; *matK + atpF-atpH + psbKpsbI*; *matK + psbK-psbI*; *matK + trnH-psbA + atpFatpH* and *matK + trnH-psbA + psbKpsbI.* It is worth noting that the success rates of combined barcodes were higher than those of the single locus for intraspecific variation as well as infraspecific variation as expected with matK + psbK-psbL + atpF-atpH and psbK-psbL + atpF-atpH + matK giving the highest identification at 108.6% [Table 3 and Table 4]. Overall, however, the means of all infraspecific distances were significantly lower than those of intraspecific distances as expected [Table 3 & Table 4]. Overall, the findings of this study reveal the novelty of chloroplast genes in delineating the diversity of cowpea at sub-species level. The superiority of *rbcL* as a suitable single loci candidate in delineatingcowpea variants is demonstrated and continues to reveal multigenic combinations as being equally informative.

5.3. DNA Barcode Gap Analysis

5.3.1. Single Locus Approach [SLA]

The infra and intraspecific genetic divergence was inferred based on the seven candidate DNA barcode loci at a scale of 0.001 distance units [Figure 1 for single loci and Figure 2 for multiple loci combinations]. The goal was to identify the most informative single locus candidate DNA barcode gene markers that show the best discriminatory power at varietietal level in common cowpea. DNA barcoding gaps were evaluated by comparing the distribution of intra-versus infra-specific divergences [38]. Median and Wilcoxon Two-Sample Tests were used to evaluate whether these distributions overlapped. The assessment of the informativeness of each candidate loci was therefore done by analyses of intra and infraspecific K2P distance matrices. The purpose was to delineate the barcoding gap. It is noteworthy that the distance distribution for each single loci gene displayed the characteristic peaks [Figure 1 & Figure 2]. In this study however, no distinct barcoding gaps typical of CO1 may have been reported but it lends credence to a clearly defined range where the infraspecific variation is significantly

lower than the intraspecific divergence as expected. Out of the seven candidate loci, rbcL loci reveal a relatively well separated distribution followed by matK at single locus level [**Figure 1** & **Figure 2**]. Furthermore, it was confirmed that the intraspecific divergences of all the seven loci was significantly higher than that of the corresponding infraspecific variations by Wilcoxon two-sample tests [**Table 5** and **Table 6**].

5.3.2. Multi Locus Approach [MLA]

Previous studies have raised concerns about SLA in discriminating between closely related organisms [9]; [12]; [15]; [25]; [26] and [28]. To circumvent this, a multilocus approach (MLA) was employed in delineating infra and intra-specific genetic divergencebetween the samples based on the following multigenic combination *matK + atpF-atpH*; *matK + trnH-psbA*; *matK + atpF-atpH + psbKpsbI*; *matK + psbK-psbI*; *matK + trnH-psbA + atpFatpH* and *matK + trnH-psbA*. Overall, the findings of this study reveal the informativeness of chloroplast genes in delineating the diversity of cowpea at sub-species level and continue to reveal further that MLA is more informative [**Tables 3-5**]. The superiority of *rbcL*as a suitable single loci candidate in delineating cowpea variants is demonstrated, and continues to reveal multigenic combinations as being more informative. Two clear peaks appear distinguishable albeit some overlap of intra- and infra-specific distances [**Figure 2**]. These observations are confirmed by median and Wilcoxon two samples statistical tests differentiating the medians for the former and the medians plus the distributions between the infra- and intra-specific distances for the latter **Table 5** and **Table 6**]. For each distribution, Median and Wilcoxon two sample tests were largely significant which agrees with studies by [12].

The histograms above give a visual impression of the bar-coding gap for each potential marker. A good marker for DNA bar-coding should have "good" gap and no overlaps between the peaks [12]; [28], [38]. However, a marker with overlaps between the extreme peaks is not necessarily a bad candidate for bar-coding. Therefore to test this, a statistical tool was employed to evaluate each of the markers and the different marker combinations as a potential barcode candidate despite the overlaps and issues with the bar-coding gap. The use of the Wilcoxon signed rank tests was specifically due to the fact that our data on the Kimura Two Parameters Pairwise Distances (K2P) do not assume a normal Gaussian distribution as well as sample means did not assume a normal distribution. So instead of testing for the difference between the means we tested for the difference in distribution. Consider the case of matK vs *atpF-atpH* below; W+ = 25350, W− = 31600, N = 1446, $P = 0.0808$, matK = atpF-atpH. Looking at the histograms, the marker *atpF-atpH* seems to be more superior to *matK* because *matK* has a lot of overlaps. However upon conducting the Wilcoxon Signed Sum Rank test comparing the data on the two markers, we find that there is no significant difference in the distribution of the datasets within the two markers since $P = 0.0808$ which is more than the *P value* threshold at 95% confidence ($P = 0.05$).

In **Table 7** above, a comparison of intra and infraspecific distance medians was explored in order to determine the best barcode marker. In this table, the *P value* for the median test is the probability that the difference between intra and infraspecific

Table 5. Wilcoxon signed-ranks test for intra-specific pair-distances.

matK vs trnH-psbA	W+ = 276,200, W− = 203,600, N = 2087, $P < 0.0001$	matk >> trnH-psbA
matK vs atpF-atpH	W+ = 25,350, W− = 31,600, N = 1446, $P = 0.0808$	matk = atpF-atpH
matK vs psbK-psbL	W+ = 14,830, W− = 21,480, N = 1377, $P = 0.0092$	matK < psbK-psbL
trnH-psbA vs atpF-atpH	W+ = 28,180, W− = 29,110, N = 1319, $P = 0.7963$	trnH-psbA = atpF-atpH
trnH-psbA vs psbK-psbL	W+ = 17,120, W− = 19,200, N = 1250, $P = 0.4146$	trnH-psbA = psbK-psbL
atpF-atpH vs psbK-psbL	W+ = 18,780, W− = 13,100, N = 609, $P = 0.0142$	atpF-atpH > psbK-psbL
4 loci vs matK + trnH-psbA	W+ = 878,500, W− = 5,154,000, N = 13,655, $P < 0.0001$	4 loci <<< matK_trnH-psbA
4 loci vs matK + trnH-psbA + atpF-atpH	W+ = 4,479,000, W− = 2,667,000, N = 10,725, $P < 0.0001$	4 loci >>> matk_trnH-psbA_atpF-atpH
4 loci vs matK + trnH-psbA + psbK-psbL	W+ = 2,338,000, W− = 6,205,000, N = 16,049, $P < 0.0001$	4 loci <<< matK_trnH-psbA_psbK-psbL
4 loci vs matK + atpF-atpH	W+ = 3,256,000, W− = 307,800, N = 12,246, $P = 0.4089$	4 loci = matK_atpF-atpH
4 loci vs matK + pbsK-pbsl	W+ = 7404, W− = 24,220, N = 12,085, $P < 0.0001$	4 loci <<< matK_psbK-psbL
4 loci vs matK + psbK-psbL + atpF-atpH	W+ = 912,700, W− = 2,068,000, N = 14,229, $P < 0.0001$	4 loci <<< matK_psbK-psbL_atpF-atpH
matK + trnH-psbA vs matK + trnH-psbA + atpF-atpH	W+ = 5466000, W− = 480300, N = 6411, $P < 0.0001$	matK_trnH-psbA >>> matK_trnH-psbA_atpF-atpH
matK + trnH-psbA vs matK + trnH-psbA + psbK-psbL	W+ = 4,442,000, W− = 1,748,000, N = 10,214, $P < 0.0001$	matK_trnH-psbA >>> matK_trnH-psbA_psbK-psbL
matK + trnH-psbA vs matK + atpF-atpH	W+ = 2,030,000, W− = 169,900, N = 6411, $P < 0.0001$	matK_trnH-psbA >>> matK_atpF-atpH
matK + trnH-psbA vs matK + psbK-psbL	W+ = 1,234,000, W− = 746,900, N = 6250, $P < 0.0001$	matK_trnH-psbA >> matK_psbK-psbL
Matk + trnH-psbA vs matK + atpF-atpH + psbk-psbL	W+ = 4,805,000, W− = 1,242,000, N = 8394, $P < 0.001$	matK_trnH-psbA >> matK_atpF-atpH_psbK-psbL
rpoB vs rbcL	W+ = 6460, W− = 57,440, N = 1273, $P < 0.0001$	rpoB << rbcL
rpoB vs rpoC1	W+ = 18,320, W− = 27,740, N = 710, $P = 0.002$	rpoB < rpoC1
rbcL vs rpoC1	W+ = 48,890, W− = 10,110, N = 1251, $P < 0.0001$	rbcL >>> rpoC1
rbcL_atpF-atpH_matK vs rpoB_atpF-atpH_matK	W+ = 18,800,000, W− = 2,636,000, N = 11,217, $P < 0.0001$	rbcL_atpF-atpH_matK >>> rpoB_atpF-atpH_
rbcL_atpF-atpH_matK vs rpoC1_atpF-atpH_matK	W+ = 1,747,000, W− = 3292,000, N = 11,211, $P < 0.0001$	rbcL_atpF-atpH_matK <<< rpoC1_atpF-atpH_matK
matK vs rpoB	W+ = 42,170, W− = 23,900, N = 1473, $P < 0.0001$	matK >>> rpoB

Continued

matK vs rbcL	W+ = 76,380, W− = 332,700, N = 2014, $P < 0.0001$	matK <<< rbcL
matK vs rpoC1	W+ = 29,400, W− = 29,600, N = 1451, $P = 0.9573$	matK = rpoC1
rpoC1_atpF-atpH_matK vs rpoB_atpF-atpH_matK	W+ = 2,413,000, W− = 1,843,000, N = 9484, $P < 0.0001$	rpoC1_atpF-atpH_matK >>> rpoB_atpF-atpH_matK

Legend: From the table above, N refers to the total number of pairwise comparisons while W+ is the sum of positive runs for the first column while W− is the sum of negative runs for the second column. The efficiency of one marker, or a combination of markers was evaluated in determining intraspecific distances. Results here indicate that matK >> trnH-psbA [$P < 0.0001$] while psbK-psbL seems to be a relatively better barcode than matK, $P = 0.0092$. rbCL is a better barcode than rpoC1 where $P < 0.0001$. In the multi locus approach, the two combinations matK + psbK-psbL and the three loci combination matK + _psbK-psbL + _atpF-atpH were superior to the 4loci [matK + trnH-psbA + atpF-atpH + psbK-psbL] approach [$P < 0.0001$] in delineating intraspecific distances. However, the 4loci performed better than the combination matK + trnH-psbA + atpF-atpH, $P < 0.0001$.

Table 6. Wilcoxon signed rank tests for infra-specific differences among loci.

Wilcoxon signed-ranks test—infra-specific pair-distances		
matK vs trnH-psbA	W+ = 32,340, W− = 27,000, N = 804, $P = 0.1487$	matK = trnH-psbA
matK vs atpF-atpH	W+ = 8512, W− = 13,220, N = 566, $P = 0.0067$	matK < atpF-atpH
matK vs psbK-psbL	W+ = 13,150, W− = 5375, N = 540, $P < 0.0001$	matK >>> psbK-psbL
trnH-psbA vs atpF-atpH	W+ = 10340, W− = 13750, N = 682, $P = 0.0700$	trnH-psbA = atpF-atpH
trnH-psbA vs psbK-psbL	W+ = 15,500, W− = 3615, N = 656, $P < 0.0001$	trnH-psbA >>> psbK-psbL
atpF-atpH vs psbK-psbL	W+ = 11,400, W− = 4886, N = 418, $P < 0.01$	atpF-atpH >> psbK-psbL
4 loci vs matK + trnH-psbA	W+ = 1,064,000, W− = 1,047,000, N = 6974, $P < 0.0001$	4 loci >>> matK_trnH-psbA
4 loci vs matK + trnH-psbA + atpF-atpH	W+ = 1,127,000, W− = 605,200, N = 5581, $P < 0.0001$	4loci >>> matK_trnH-psbA_psbK-psbL
4 loci vs matK + trnH-psbA + psbK-psbL	W+ = 967,800, W− = 2,305,000, N = 8421, $P < 0.0001$	4 loci >>> matK_trnH-psbA_psbK-psbL
4 loci vs matK + atpF-atpH	W+ = 183,300, W− = 215,900, N = 6446, $P = 0.0281$	4 loci < matK_atpF-atpH
4 loci vs matK + pbsK-pbsl	W+ = 7798, W− = 7778, N = 6337, $P = 0.9987$	4 loci = matK_psbK-psbL
4 loci vs matK + psbK-psbL + atpF-atpH	W+ = 689,300, W− = 716,100, N = 7664, $P = 0.4903$	4 loci = matK_psbK-psbL_atpF-atpH
matK + trnH-psbA vs matK + trnH-psbA + atpF-atpH	W+ = 1,446,000, W− = 42,530, N = 3178, $P < 0.0001$	matK_trnH-psbA >>> matK_trnH-psbA_atpF-atpH
matK + trnH-psbA vs matK + trnH-psbA + psbK-psbL	W+ = 1,085,000, W− = 434,400, N = 5153, $P < 0.0001$	matK_trnHpsbA >> matK_trnH-psbA_psbK-psbL
matK + trnH-psbA vs matK + atpF-atpH	W+ = 734000, W− = 41,630, N = 3178, $P < 0.0001$	Matk_trnH-psbA >>> matK_atpF-atPH
matK + trnH-psbA vs matK + psbK-psbL	W+ = 400,900, W− = 223,500, N = 3069, $P < 0.0001$	MatK_trnH-psbA >> matK_psbK-psbL
matk + trnH-psbA vs matK + atpF-atpH + psbk-psbL	W+ = 1,357,000, W− = 155,500, N = 4398, $P < 0.0001$	matK_trnH-psbA >>> matK_atpF-atpH_psbK-psbL
rpoB vs rbcL	W+ = 6142, W− = 18,830, N = 648, $P < 0.0001$	rpoB <<< rbCL

Continued

rpoB vs rpoC1	W+ = 9007, W− = 7829, N = 446, $P = 0.4116$	rpoB = rpoC1
rbCL vs rpoC1	W+ = 15,840, W− = 7601, N = 636, $P < 0.0001$	rbCL >>> rpoC1
rbcL_atpF-atpH_matK vs rpoB_atpF-atpH_matK	W+ = 464,600, W− = 824,200, N = 5894, $P < 0.0001$	rbCL_atpF-atpH_matK <<< rpoB_atpF-atpH_matK
rbcL_atpF-atpH_matK vs rpoC1_atpF-atpH_matK	W+ = 553,200, W− = 985,900, N = 5720, $P < 0.0001$	rbcL_atpF-atpH_matK <<< rpoC1_atpF-atpH_matK
matK vs rpoB	W+ = 10,340, W− = 11,810, N = 573, $P = 0.4056$	matK = rpoB
matK vs rbcL	W+ = 20,870, W− = 36,760, N = 763, $P < 0.0001$	matK <<< rbcL
matK vs rpoC1	W+ = 8329, W− = 13,830, N = 688, $P = 0.0018$	matK << rpoC1
rpoC1_atpF-atpH_matK vs rpoB_atpF-atpH_matK	W+ = 753,200, W− = 657,200, N = 5098, $P = 0.0123$	rpoC1_atpF-atpH_matK > rpoB_atpF-atpH_matK

Legend: From the table above, N refers to the total number of pairwise comparisons while W+ is the sum of positive runs for the first column while W− is the sum of negative runs for the second column. In this table, a comparison of one markervs another marker and or combinations in order to determine which marker is superior to the other. Results here indicate that among single loci matK and trnH-psbA are superior to psbK-psbL [$P < 0.0001$]. rbcL is superior to rpoB and rpoC1 [$P < 0.0001$]. Amongst the combinations, the 4 loci [matK + trnH-psbA + atpF-atpH + psbL-psbK] approach was superior to all the other two loci combinations $P < 0.0001$ except for matK + atpF-atpH [$P = 0.0281$]. For two loci combinations, matK = trnH-psbA was greater than the combinations matK + atpF-atpH and matK + psbK-psbL. Overall matK, trnH-psbA and rpoB would seem the best marker for determining infraspecific distances.

Table 7. Median and Wilcoxon two sample statistical tests applied to the distributions of intra- and infra-specific K2P distances for each potential DNA barcode. In this case #A refers to the number of pairwise comparisons for intraspecific while #B is the number of pairwise comparisons for infraspecific distances Barcoding gaps were assessed with Median and Wilcoxon Two sample statistical Tests. Wilcoxon test served as a QC for median test. These observations were confirmed by two sample statistical tests differentiating the means for the former and the means plus distributions between the intra and infraspecific distance for the latter.

K2P distributions	Median test	Wilcoxon Two Sample Test
matK	#A = 1107 #B = 344, Median = 0.00733, $P = 0.0387$	#A = 1107 #B = 344, W = 27,800, $P = 1.66e{-}05$
trnH-psbA	#A = 980 #B = 460, Median = 0.01723, $P = 0.0002$	#A = 980 #B = 460, W = 56,980, $P = 0.0003$
atpF-atpH	#A = 339 #B = 222, Median = 0.0000, $P = 0.8473$	#A = 339 #B = 222, W = 7369, $P = 0.1523$
psbK-psbL	#A = 269 #B = 196, Median = 0.00494, $P < 0.0001$	#A = 269 #B = 196, W = 10,740, $P < 0.0001$
4 loci	#A = 9745 #B = 5122, Median = −214748364, $P = 0.922$	#A = 9745 #B = 5122, W = 745,900, $P < 0.0001$
matK + trnH-psbA	#A = 3909 #B = 1853, Median = 0.61702, $P = 0.0035$	#A = 3909 #B = 1853, W = 600,500, $P < 0.0001$
matK + trnH-pbsA + atpF-atpH	#A = 6704 #B = 3323, Median = −214748364, $P = 0.1013$	#A = 6704 #B = 3323, W = 942,500, $P < 0.0001$
matK + trnH-pbsA + psbK-psbL	#A = 6305 #B = 3300, Median = 0.02004, $P = 0.0551$	#A = 6305 #B = 3300, W = 2,145,000, $P < 0.0001$
matK + atpF-atpH	#A = 2501 #B = 1325, Median = 0.0000, $P = 0.6871$	#A = 2501 #B = 1325, W = 398,500, $P < 0.0001$
matK + psbK-psbL	#A = 2340 #B = 1216, Median = 0.00683, $P = 0.2175$	#A = 2340 #B = 1216, W = 204,700, $P < 0.0001$
matK + atpF-atpH + psbK-psbL	#A = 4484 #B = 2543, Median = 0.0000, $P = 0.0488$	#A = 4484 #B = 2543, W = 658,800, $P < 0.0001$
rpoB	#A = 366 #B = 229, Median = 0.0000, $P = 0.4224$	#A = 366 #B = 229, W = 3350, $P = 0.1446$
rpoC1	#A = 344 #B = 217, Median = 0.0089, $P = 0.5144$	#A = 344 #B = 217, W = 7167, $P = 0.2458$
rbcL	#A = 907 #B = 419, Median = 0.0709, $P < 0.0001$	#A = 907 #B = 419, W = 57,370, $P < 0.0001$
rbcL_atpF-atpH_matK	#A = 6472 #B = 3258, Median = 0.0000, $P = 0.04032$	#A = 6472 #B = 3258, W = 1,229,000, $P < 0.0001$
rpoB_atpF-atpH_matK	#A = 4745 #B = 2636, Median = 0.55038, $P = 0.0011$	#A = 4745 #B = 2636, W = 647,200, $P < 0.0002$
rpoC1_atpF-atpH_matK	#A = 4739 #B = 2642, Median = −107,374,182, $P = 0.6840$	#A = 4739 #B = 2642, W = 728,600, $P < 0.0001$

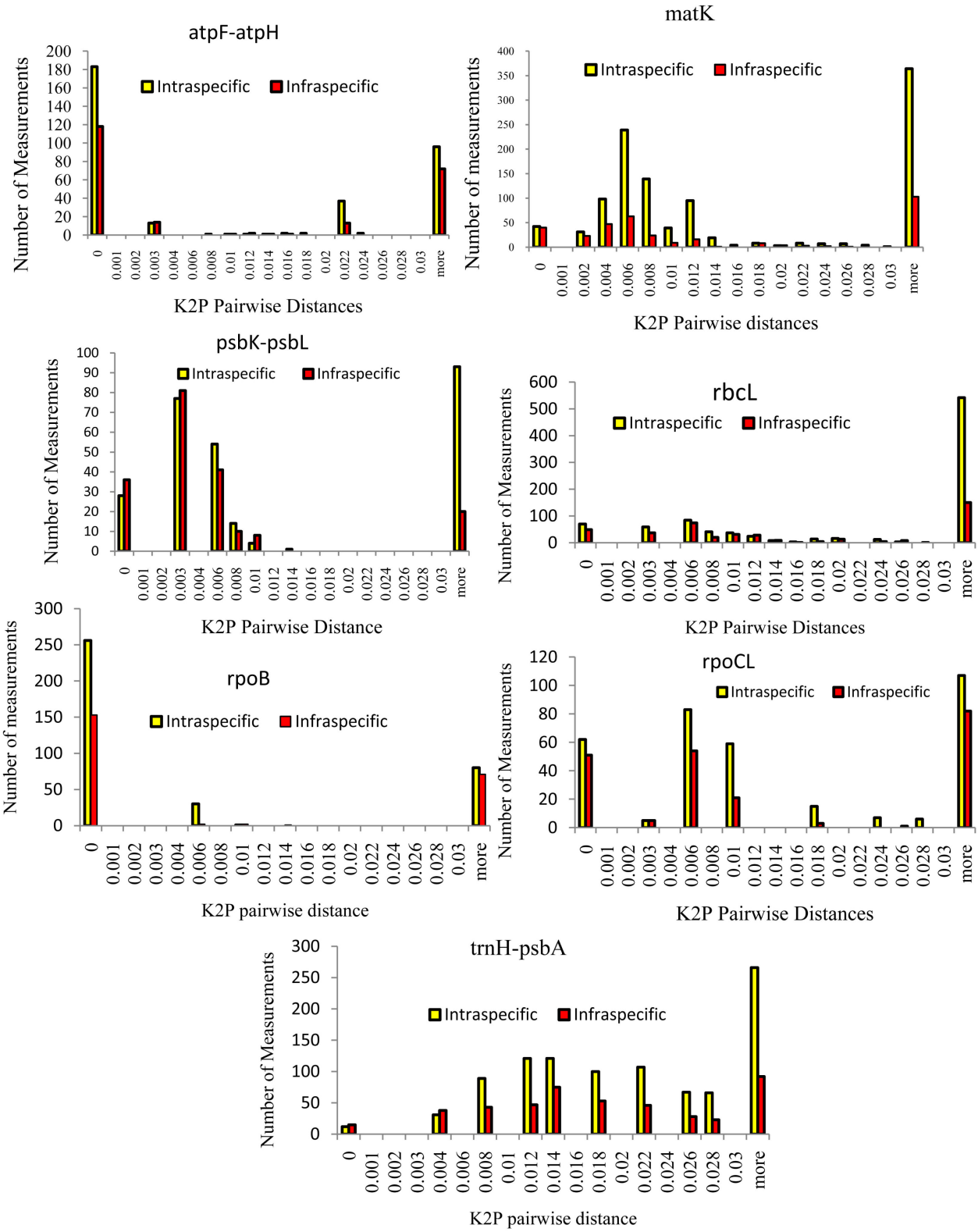
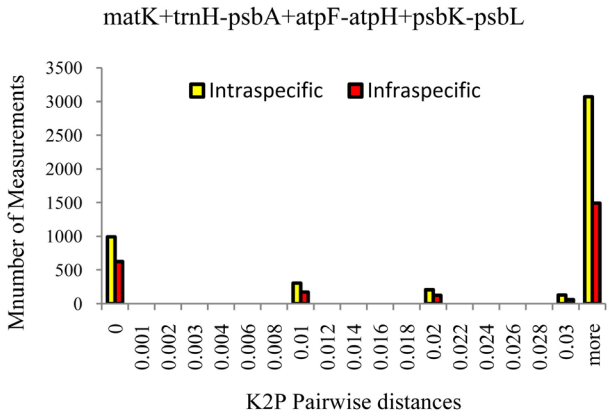
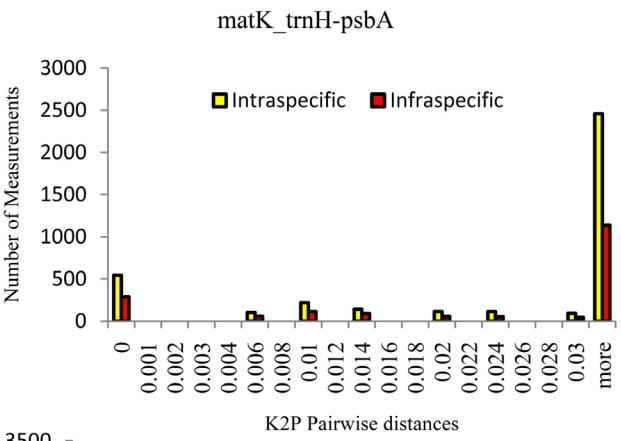
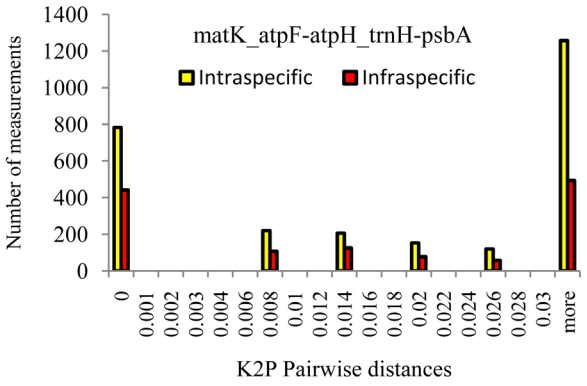
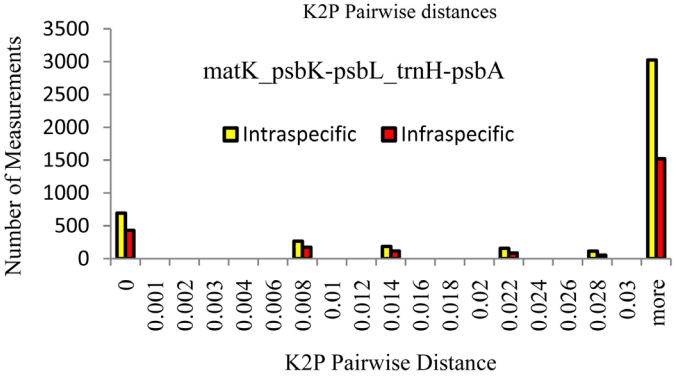


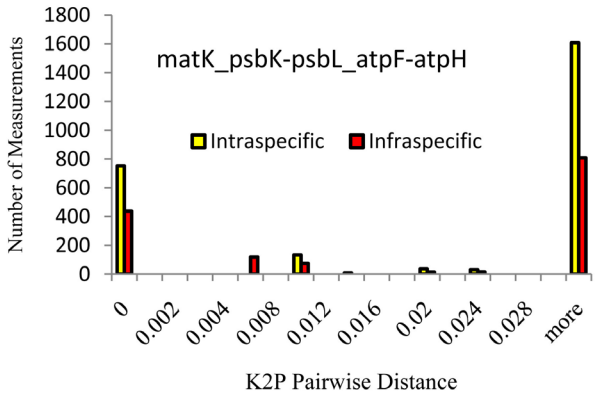
Figure 1. Relative distributions of intraspecific divergence between *Vigna unquiculata* variants (yellow) and infraspecific distances (red) for seven candidate single loci genes *matK*, *rbcL*, *rpoB*, and *rpoC*1, and three noncoding spacers (*atpF-atpH*, *trnH-psbA*, and *psbK-psbL*).

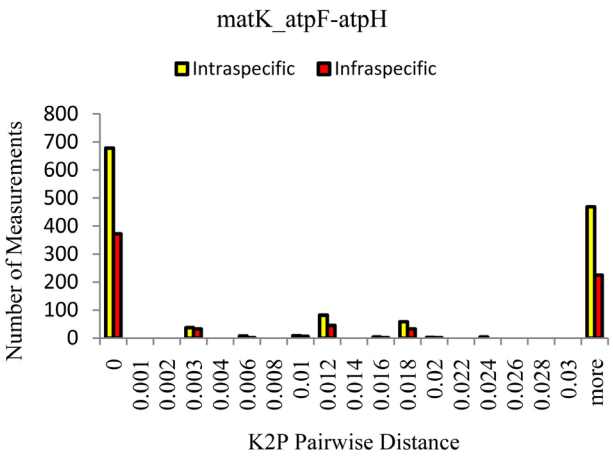

matK+trnH-psbA+atpF-atpH+psbK-psbL
Intraspecific
Infraspecific
Mnumber of Measurements
K2P Pairwise distances

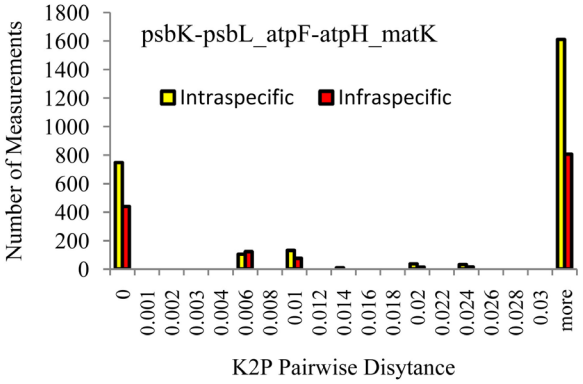

matK_trnH-psbA
Intraspecific
Infraspecific
Number of Measurements
K2P Pairwise distances

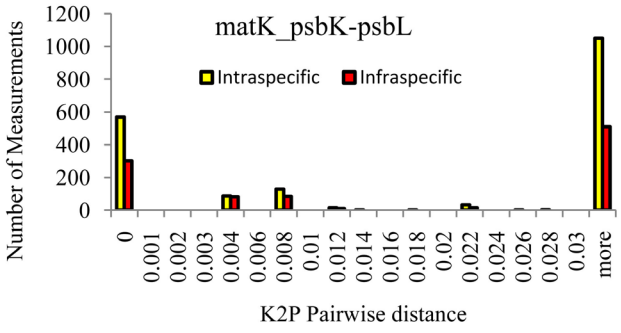

matK_atpF-atpH_trnH-psbA
Intraspecific
Infraspecific
Number of measurements
K2P Pairwise distances


matK_psbK-psbL_trnH-psbA
Intraspecific
Infraspecific
Number of Measurements
K2P Pairwise Distance


matK_psbK-psbL_atpF-atpH
Intraspecific
Infraspecific
Number of Measurements
K2P Pairwise Distance


matK_atpF-atpH
Intraspecific
Infraspecific
Number of Measurements
K2P Pairwise Distance


psbK-psbL_atpF-atpH_matK
Intraspecific
Infraspecific
Number of Measurements
K2P Pairwise Disytance


matK_psbK-psbL
Intraspecific
Infraspecific
Number of Measurements
K2P Pairwise distance

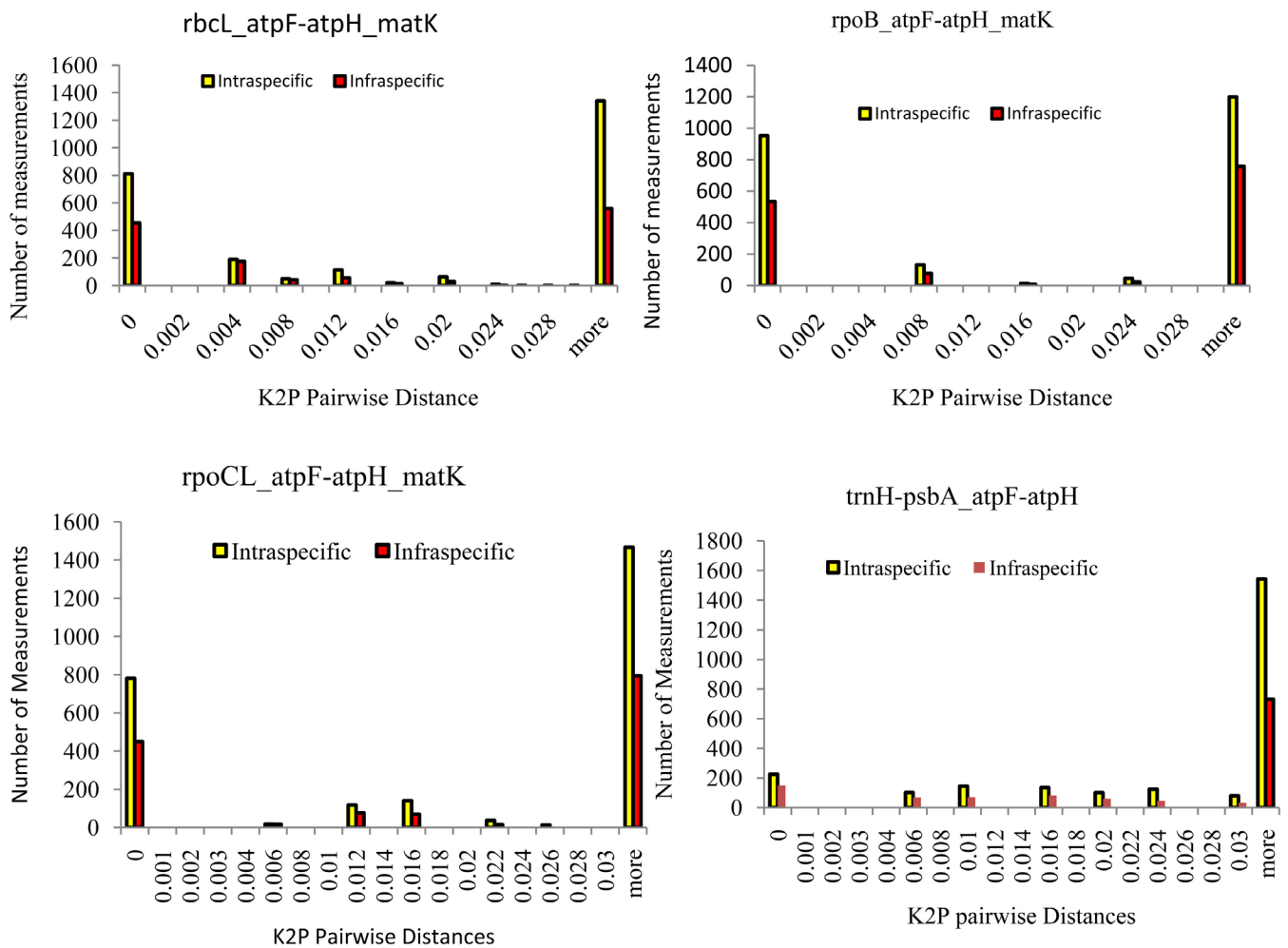


Figure 2. Relative distributions of intraspecific divergence between congeneric (yellow) and infraspecific K2P distances (red) for 12different combinations keeping matK for each.

divergence does not occur for example, the *P value* for *matK is* 0.0387 signifying a 3.87% type two error rate. Therefore, in determining intra and infra specific distances, *matK* has an accuracy of more than 95% where $P < 0.05$. This is confirmed by the Wilcoxon test where $P = 1.66^{-5}$ which is the type one error rate. Compared to the marker *atpF-atpH*, the median test *P value* = 0.8473 signifying an 84.73% error rate with Wilcoxon *P value* = 0.1523 making it an unsuitable marker for delineating intra and infraspecific distances.

6. Conclusion

This study sought to investigate the plastid sequences in Kenyan cowpea variants looking for the loci that could be used for delineating different phylogeographic groups. Based on the results presented here, the study concludes the best locus combinations for DNA-barcoding of Kenyan cowpea. The evidence presented here clearly demonstrates the overall utility of DNA barcoding in delineating molecular diversity of Ke-

nyan cowpea at sub-species level. The results of this study demonstrate the informativeness of plastid region in delineating intra and infra-specific distances at single loci level; *matK*, *trnH-psbA*, *psbK-psbL*, and *rbcL*. rbcL and matK distinguish themselves as ideal barcodes at single loci level. However, among the combinations, *matK + trnH-psbA*, *rpoB + atpF-atpH + matK* appear to be the best barcodes in delineating genetic distances. The current study however demonstrates that using combinations of DNA barcodes [MLA] improves accuracy of delineation. This study therefore recommends a multi locus approach in delineating cowpea at varietal level.

Acknowledgements

The author acknowledges support of the following persons who contributed in many ways to the success of this study; Wenqin Wang, Shang Sheng, ASARECA.

Conflict of Interest Declaration

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. and Janzen, D.H. (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8369-8374. http://dx.doi.org/10.1073/pnas.0503123102

[2] Hollingsworth, P.M., Forrest, L.L., Spouges, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L., Fazekas, A.J., Graham, S.W., James, K.E., Kim, K.-J., Kress, W.J., Schneider, H., van AlphenStahl, J., Barrett, S.C.H., van den Berg, C., Bogarin, D., Burgess, K.S., Cameron, K.M., Carine, M., Chacón, J., Clark, A., Clarkson, J.J., Conrad, F., Devey, D.S., Ford, C.S., Hedderson, T.A.J., Hollingsworth, M.L., *et al.* (2009) A DNA Barcode for Land Plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794-12797. http://dx.doi.org/10.1073/pnas.0905845106

[3] Neuhaus, H. and Link, G. (1987) The Chloroplast tRNALys (UUU) Gene from Mustard (*Sinapsis alba*) Contains a Class II Intron Potentially Coding for a Maturase-Related Polypeptide. *Current Genetics*, **11**, 251-257. http://dx.doi.org/10.1007/BF00355398

[4] Barthet, M.M. and Hilu, K.W. (2007) Expression of matK: Functional and Evolutionary Implications. *American Journal of Botany*, **94**, 1402-1412. http://dx.doi.org/10.3732/ajb.94.8.1402

[5] Mu¨ller, K.F., Borsch, T. and Hilu, K.W. (2006) Phylogenetic Utility of Rapidly Evolving DNA at High Taxonomical Levels: Contrasting matK, trnT-F and rbcL in Basal Angiosperms. *Molecular Phylogenetics and Evolution*, **41**, 99-117. http://dx.doi.org/10.1016/j.ympev.2006.06.017

[6] Babbar, S.B., Raghuvanshi, S., Singh, H.K., Parveen, I. and Malik, S. (2012) An Overview of the DNA Barcoding of Plants. *Phytomorphology*, **62**, 69-99.

[7] Cameron, K.M. (2005) Leave It to the Leaves: A Molecular Phylogenetic Study of Malaxideae (*Orchidaceae*). *American Journal of Botany*, **92**, 1025-1032. http://dx.doi.org/10.3732/ajb.92.6.1025

[8] Min, X.J. and Hickey, D.A. (2007) Assessing the Effect of Varying Sequence Length on

DNA Barcoding of Fungi. *Molecular Ecology Notes*, **7**, 365-373. http://dx.doi.org/10.1111/j.1471-8286.2007.01698.x

[9] CBOL Plant Working Group (2009) A DNA Barcode for Land Plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794-12797. http://dx.doi.org/10.1073/pnas.0905845106

[10] Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y. and Leon, C. (2010) Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLoS ONE*, **5**, e8613. http://dx.doi.org/10.1371/journal.pone.0008613

[11] Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R., Graham, S.W., Newmaster, S.G., Husband, B.C., *et al.* (2008) Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. *PLoS ONE*, **3**, e2802. http://dx.doi.org/10.1371/journal.pone.0002802

[12] Lahaye, R., Van Der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., Maurin, O., Duthoit, S., Barraclough, T.G. and Savolainen, V. (2008) DNA Barcoding the Floras of Biodiversity Hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2923-2928. http://dx.doi.org/10.1073/pnas.0709936105

[13] Blaxter, M.L. (2004) The Promise of DNA Taxonomy. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, **359**, 669-679. http://dx.doi.org/10.1098/rstb.2003.1447

[14] Shaw, J., Lickey, E.B., Schilling, E.E. and Small, R.L. (2007) Comparison of Whole Chloroplast Genome Sequences to Choose Noncoding Regions for Phylogenetic Studies in Angiosperms: The Tortoise and the Hare III. *American Journal of Botany*, **94**, 275-288. http://dx.doi.org/10.3732/ajb.94.3.275

[15] Kress, W.J. and Erickson, D.L. (2007) A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. *PLoS ONE*, **2**, e508. http://dx.doi.org/10.1371/journal.pone.0000508

[16] Shaw, J. and Small, R.L. (2005) Chloroplast DNA Phylogeny and Phylogeography of the North American Plums (*Prunus* Subgenus *Prunus* Section *Prunocerasus*, Rosaceae). *American Journal of Botany*, **92**, 2011-2030. http://dx.doi.org/10.3732/ajb.92.12.2011

[17] Chang (2006) The Chloroplast Genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative Analysis of Evolutionary Rate with That of Grasses and Its Phylogenetic Implications. *Molecular Biology and Evolution*, **23**, 279-291. http://dx.doi.org/10.1093/molbev/msj029

[18] Whitlock, B.A., Hale, A.M. and Groff, P.A. (2010) Intraspecific Inversions Pose a Challenge for the trnH-psbA Plant DNA Barcode. *PLoS ONE*, **5**, e11533. http://dx.doi.org/10.1371/journal.pone.0011533

[19] Nicolalde-Morejón, F., Vergara-Silva, F., González-Astorga, J., Stevenson, D.W., Vovides, A.P., *et al.* (2010) A Character-Based Approach in the Mexican Cycads Supports Diverse Multigene Combinations for DNA Barcoding. *Cladistics*, **26**, 1-15.

[20] Wang, W., Wu, Y., Yan, Y., Ermakova, M., Kerstetter, R. and Messing, J. (2010) DNA Barcoding of the Lemnaceae, a Family of Aquatic Monocots. *BMC Plant Biology*, **10**, 205. http://www.biomedcentral.com/1471-2229/10/205-BMC Plant Biology 10:205 http://dx.doi.org/10.1186/1471-2229-10-205

[21] Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madrinan, S., Petersen, G., Seberg, O., Jorgsensen, T., Cameron, K.M., Carine, M., Pedersen, N., Hedderson, T.A.J., Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L. and Wilkinson, M. (2007) A Proposal for a Standardised Protocol to Barcode All Land

Plants. *Taxon*, **56**, 295-299.

[22] Sass, C., Little, D.P., Stevenson, D.W. and Specht, C.D. (2007) DNA Barcoding in the Cycadales: Testing the Potential of Proposed Barcoding Markers for Species Identification of Cycads. *PLoS ONE*, **2**, e1154. http://dx.doi.org/10.1371/journal.pone.0001154

[23] Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. and Janzen, D.H. (2005) Use of DNA Barcodes to Identify Flowering Plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8369-8374. http://dx.doi.org/10.1073/pnas.0503123102

[24] Newmaster, S.G., Fazekas, A.J. and Ragupathy, S. (2006) DNA Barcoding in Land Plants: Evaluation of rbcL in a Multigene Tiered Approach. *Canadian Journal of Botany*, **84**, 335-341. http://dx.doi.org/10.1139/b06-047

[25] Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madrinan, S., Petersen, G. and Fay, M. (2009) Barcoding of Plants and Fungi. *Science*, **325**, 682-683. http://dx.doi.org/10.1126/science.1176906

[26] Erickson, D.L., Spouge, J., Resch, A., Weigt, L.A. and Kress, J.W. (2008) DNA Barcoding in Land Plants: Developing Standards to Quantify and Maximize Success. *Taxon*, **57**, 1304-1316.

[27] Kane, N.C. and Cronk, Q. (2008) Botany without Borders: Barcoding in Focus. *Molecular Ecology*, **17**, 5175-5176. http://dx.doi.org/10.1111/j.1365-294X.2008.03972.x

[28] Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S. and Francis, C.M. (2004) Identification of Birds through DNA Barcodes. *PLoS Biology*, **2**, e312. http://dx.doi.org/10.1371/journal.pbio.0020312

[29] Pennisi, E. (2007) Taxonomy. Wanted: A Barcode for Plants. *Science*, **318**, 190-191. http://dx.doi.org/10.1126/science.318.5848.190

[30] Kane, N., Sveinsson, S., Dempewolf, H., Yang, J.Y., Zhang, D., Engels, J.M.M. and Cronk, Q. (2012) Ultra-Barcoding in Cacao (*Theobroma* spp., Malvaceae) Using Whole Chloroplast Genomes and Nuclear Ribosomal DNA. *American Journal of Botany*, **99**, 320-329. http://dx.doi.org/10.3732/ajb.1100570

[31] Hebert, P.D.N., Cywinska, A., Ball, S.L. and de Waard, J.R. (2003) Biological Identifications through DNA Barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313-321. http://dx.doi.org/10.1098/rspb.2002.2218

[32] Yang, J.B., Tang, M., Li, H.T., Zhang, Z.R. and Li, D.Z. (2013) Complete Chloroplast Genome of the Genus *Cymbidium*: Lights into the Species Identification Phylogenetic Implications and Population Genetic Analyses. *BMC Evolutionary Biology*, **13**, 84. http://dx.doi.org/10.1186/1471-2148-13-84

[33] Nock, C.J., Waters, D.L., Edwards, M.A., Bowen, S.G., Rice, N., Cordeiro, G.M. and Henry, R.J. (2011) Chloroplast Genome Sequences from Total DNA for Plant Identification. *Plant Biotechnology Journal*, **9**, 328-333. http://dx.doi.org/10.1111/j.1467-7652.2010.00558.x

[34] Sucher, N.J. and Carles, M.C. (2008) Genome-Based Approaches to the Authentication of Medicinal Plants. *Planta Medica*, **74**, 603-623. http://dx.doi.org/10.1055/s-2008-1074517

[35] Brinkman, F.S.L. (2001) Phylogenetic Analysis. In: Baxevanis, A.D. and Francis Ouellette, B.F., Eds., *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, John Wiley & Sons Ltd., Vancouver, Vol. 2, 323-358. http://dx.doi.org/10.1002/0471223921.ch14

[36] Collins, R.A., Boykin, L.M., Cruickshank, R.H. and Armstrong, K.F. (2012) Barcoding's Next Top Model : An Evaluation of Nucleotide Substitution Models for Specimen Identification. *Methods in Ecology and Evolution*, **3**, 457-465. http://dx.doi.org/10.1111/j.2041-210X.2011.00176.x

[37] Whitley, E. and Ball, J. (2002) Statistics Review 6 : Nonparametric Methods. *Critical Care*, **6**, 509-513. http://dx.doi.org/10.1186/cc1820

[38] Meyer, C.P. and Paulay, G. (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. *PLoS Biology*, **3**, 2229-2238. http://dx.doi.org/10.1371/journal.pbio.0030422

American Journal of Molecular Biology, 2016, 6, 159-169
http://www.scirp.org/journal/ajmb
ISSN Online: 2161-6663
ISSN Print: 2161-6620

Genotyping of *E. coli* Isolated from Urinary Tract Infection Patients Containing B-Lactamase Resistance Gene CTX-M Group 1 in Sanandaj Medical Health Centers

Arezoo Omati[1*], Kambiz Davari[2], Borhan Shokrolahi[3]

[1]Department of Biology, Islamic Azad University, Rasht Branch, Rasht, Iran
[2]Department of Biology, Islamic Azad University, Sanandaj Branch, Sanandaj, Iran
[3]Department of Animal Science, Faculty of Agriculture, Islamic Azad University, Sanandaj Branch, Sanandaj, Iran
Email: *arezoo_2013_new@yahoo.com

How to cite this paper: Omati, A., Davari, K. and Shokrolahi, B. (2016) Genotyping of *E. coli* Isolated from Urinary Tract Infection Patients Containing B-Lactamase Resistance Gene CTX-M Group 1 in Sanandaj Medical Health Centers. *American Journal of Molecular Biology*, **6**, 159-169.
http://dx.doi.org/10.4236/ajmb.2016.64015

Received: August 11, 2016
Accepted: October 25, 2016
Published: October 28, 2016

Abstract

CTX-M-producing bacteria are known as a resistant source against *oxyimino-cephalosporin* such as cefotaxime and ceftazidime; although laboratory diagnosis of this gene has not been properly defined. The aims of this study are determining the rates of prevalence of CTX-M and CTX-M group 1 in the *Escherichia coli* (*E. coli*) obtained from urinary tract infections (UTI), and also determining their genetic relationship in the city of Sanandaj. In current study, 180 *E. coli* strains isolated from urinary tract infections were used. Sensitivity to common antibiotics was studied by the disc diffusion method. Phenotypic detection of isolated ESBL-producing starins was done by the combination disc test. CTX-M and CTX-M1 genes were detected using the PCR method and finally, the possible clonal relationship between isolates was determined using the REP-PCR method. 89 samples were ESBL-positive. The PCR assay used for detecting the CTX-M gene, showed that 48 samples out of 180 samples (26.66%) contained that gene; also among these 48 samples, 23 (12.77%) had CTX-M group 1. Based on the REP-PCR assay, 48 genotypes among 48 samples were CTX-M-positive. Results from the REP-PCR assay indicated that the clonal propagation theory of one epidemic strain of *Escherichia coli* is not apply, *i.e.* all CTX-M-producing species are not originated from one single strain and the gene is spread between different isolates. Therefore, hospitals and their employees must be more hygiene and, proper disposal of hospital waste can help to prevent the spread of different resistances.

Keywords

Escherichia coli, Urinary Tract Infection, ESBL, CTX-M, CTX-M Group 1, REP-PCR

1. Introduction

Extended-spectrum beta-lactamases (ESBLs) were reported for the first time in Germany 1983 [1]. The CTX-M family of ESBLs is a serious threat for global health [2] to the extent that in the previous decade, it was described pandemic [3]. CTX-M is the most prevalent ESBL in antrobactericeas that produce nosocomial and society-acquired infections [2] [4]. CTX-M genes are usually found on plasmids and derivative chromosomes of Beta-lactamase genes are from the *Kluyvera* spp. genus which are created by "multiple delivery mechanisms" [2] [5]. Normally, these plasmids are easily spread among microbial populations and they carry the resistant genes against other antibiotics such as aminoglycoside acetyltransferases and dihydropteroate synthases or other beta-lactamases [6]. This increase in resistance, to a large extend, is due to the spread of *E. coli* bacteria and *Klebsiella pneumonia* that carry CTX-M [7]. CTX-M group 1 contains six plasmid-dependent enzymes namely: CTX-M-1, CTX-M-3, CTX-M-10, CTX-M-12, CTX-M-15 and FEC-1 and, unprinted enzymes of CTX-M-22, CTX-M-23 and CTX-M-28 (corresponding gene bank numbers respectively are AY080894, AF488377 and AJ549244) [2]. For epidemiological study and determining the genetic relationships of resistant isolates, a rapid typing method could be a valuable tool. Repetitive Element Palindromic PCR (REP-PCR) is a suitable method for proliferation of repetitive elements of bacterial DNA, with the following characteristics: 1) low costs, 2) high discriminatory power, 3) high speed, and 4) reliable tool for typing and classification of a wide range of Gram-negative and some Gram-positive bacteria [8] [9].

A lot of research has been done on identifying CTX-M in *E. coli*. Woodford *et al*. (2004) conducted a study to identify the CTX-M in *E. coli* isolated from community and hospital in Britain. In this study, 291 CTX-M-producing samples were identified in Britain which 279 sample involved CTX-M 1and 12 samples involved CTX-M-9. The result of dendrogram indicated that 279 CTX-M-producing samples are related with each other's [10].

Leila Nasehi *et al*. (2010) studied the CTX-M, PER, SHV and TEM β-lactamase prevalence in *Lebsiella pneumoniae* isolated from clinical samples in Tehran. The results indicated that the prevalence of blaSHV, blaCTX-M, and blaTEM gens was 7.5%, 16%, 22.5 % and 23%, respectively [11].

According to the above description, the aims of this study are determining the rate of prevalence of CTX-M and CTX-M group 1 genes in the *Escherichia coli* responsible for urinary tract infection and, determining the genetic relationship between isolated starin using the typing method of REP-PCR.

2. Material and Methods

2.1. Sampling

In 2015, 325 urine samples were collected from Sanandaj laboratories. *E. coli* isolated in 180 samples that their presence were confirmed by biochemical test.

2.2. Antibiotic Sensitivity and Phenotypic Identification of ESBLs

The antibiotic sensitivity of the samples was conducted using the disc diffusion method and based on the Clinical and Laboratory Standards Institute (CLSI standards); also, the antibiotics that were used are listed in Table 2. ESBL-producing isolates were detected by the CLSI combination disc test [12]. At first, bacterial suspensions equivalent to Mc-Farland half of resistant isolates were cultured on Molar-Hinton agar media; then, two Ceftazidime and Cefotaxime discs and also, two discs of these materials combined with clavulanic acid were placed 25 mm apart on the media. After incubation, if the difference in diameters of the halos around the combined discs and the halos of the initial discs is ≥5 mm, the isolate is considered as a positive ESBLs phenotype.

2.3. Determining MIC by the E-Test

The E-test was performed with the antibiotics of Ceftazidime and Cefotaxime for all the samples that were detected as ESBLs. In this method, after making a bacterial suspension by the Mc-Farland half method, it was placed on the Molar Hinton agar plate; then, E-test strips, each representing one specific antibiotic, were placed on the Molar Hinton agar and after 24 h of incubation in 37˚C, a triangular growth zones of inhibition was formed. Then by referring to the table provided by the company that had created the E-test strips, the susceptibility of *E. coli* bacteria to the mentioned antibiotics was determined.

2.4. DNA Extraction

DNA of the bacterium was extracted using gram-negative DNA extraction kit (Sina Gene, Iran). The extracted DNA was checked by the agar gel electrophoresis.

2.5. Detection of Resistant Genes

In this study,primers from previous studies were used which their characteristics are listed in Table 1 [13] [14]. The PCR reaction was done with the final volume of 25 µl that contained 12.5 µl of PCR Master Mix (containing DNA polymerase, salts, magnesium, dNTPs and optimized reaction buffer), 1 µl of each primer, 2 µl of the sample's DNA and 8.5 µl distilled water. The conditions of reaction for each primer are shown in Table 1. The product of PCR was analyzed by electrophoresis in agar gel 1.5% and finally it was visible under UV.

2.6. REP-PCR

In order to determine the genetic relationships between ESBL-producing samples, the

Table 1. Primers and the conditions of their reaction.

Product size	PCR condition	Primer (5′→3′)	Target
759 bp	94˚C, 5 min; 35 cycles of 94˚C, 45 s; 58˚C, 45 s; 72˚C, 60 s	Forward ACGCTGTTGTTAGGAAGTG Reverse TTGAGGCTGGGTGAAGT	CTX-M
864 bp	94˚C, 2 min; 30 cycles of 95˚C, 45 s; 58˚C, 30 s; 72˚C, 45 s	Forward GGTTAAAAAATCACTGCGTC Reverse TTGGTGACGATTTTAGCCGC	CTX-M-1 group

typing method of REP-PCR was used. The REP-PCR reaction was done with the final volume of 25 µl containing: 12.5 µl of PCR Master Mix, 1 µl of the sample's DNA, 1 µl of each primer and 9.5 µl distilled water. REP1 (5'-IIIGCGCCGICATCAGGC-3') and REP2 (ACGTCTTATCAGGCCTAC-3') primers were used for amplification of repetitive sequences in the bacterial genome [15]. The reaction took place in XP Thermal Cycler with the following circumstances: Initial denaturation (2 min at 95˚C), then, 35 cycles of denaturation (1 min at 92˚C), annealing (1 min at 40˚C), extension (8 min at 65˚C) and at last, final extension (8 min at 65˚C). The products of REP-PCR were electrophoresed in Agar gel 1.5%. In the end, the bonds became visible by UV ray and then the image was recorded. The dendrogram relating to the analysis of fingerprinting was drawn by the algorithm of the Unweighted Pair-Group Method (UPGMA) using the software NTSYS v2.02e.

3. Results

3.1. Bacterial Isolated

All the samples were separated from different patients that had referred to sanandaj diagnostic laboratories. The rates of prevalence of urinary tract infection based on sex and age are shown in Figure 1 and it is logical that UTIs are more prevalent in women.

3.2. Sample Antibiotic Sensitivity

CLSI standard was used to determine the sensitivity. Samples sensitivity to antibiotics was measured and presented in Table 2 as sensitive, semi sensitive, and resistant.

3.3. Phenotypic Detection of ESBL Producers

89 *E. coli* samples (49.44%) were detected as ESBL producers by the combination disc test (Figure 2). Antibiotic susceptibility of the positive- and negative-ESBL samples are compared in Figure 3.

3.4. Results of E-Test

The results of this E-test for the ESBL-producing samples were in the range of 2 - 4 µg/ml for Cefotaxime and 1 - 16 µg/ml for Ceftazidime (Figure 4).

3.5. Detection of Resistant Genes

Among SBL-producing samples, 48 out of 89 samples contained the CTX-M gene; also 23 out of 48 samples, were detected as CTX-M group 1.

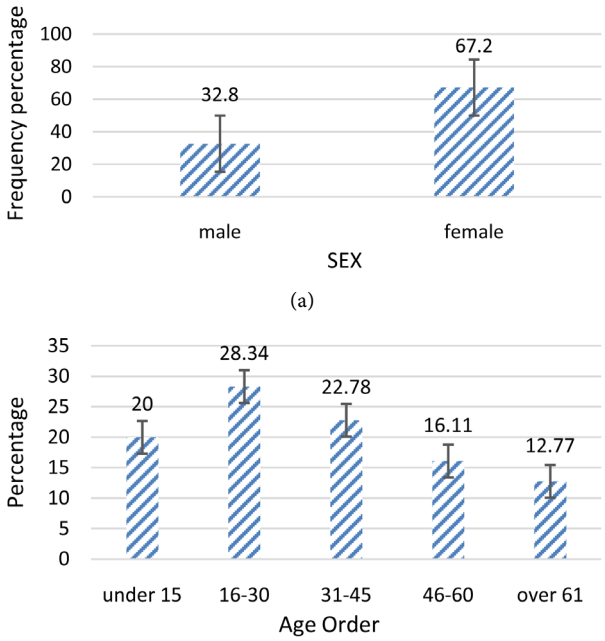


Figure 1. (a) The rate of prevalence of UTI based on sex; (b) The rate of prevalence of UTI based on age.

Figure 2. Phenotypic detection of ESBL producers by the combination disc test.

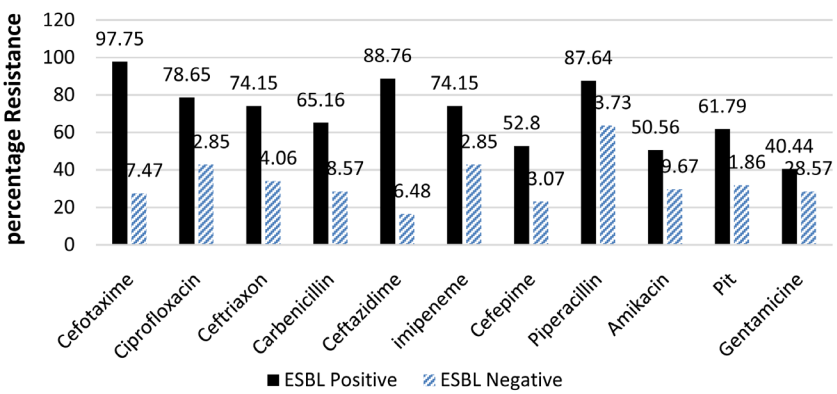


Figure 3. Comparison of the susceptibility profiles of positive and negative ESBL-producing *E. coli.*

Figure 4. Non-growth triangle due to increased antibiotic density around T-Test strip.

Table 2. Results of antibiotic sensitivity determination based on disc diffusion.

	Resistant N. (%)	Intermediate N. (%)	Susceptible N. (%)
Cefotaxime	144 (57.8)	8 (4.4)	68 (37.8)
Ciprofloxacin	87 (48.3)	22 (12.2)	71 (39.4)
Ceftriaxon	72 (40)	25 (13.9)	83 (46.1)
Carbenicillin	67 (37.3)	17 (9.4)	96 (53.3)
Ceftazidime	59 (32.8)	35 (19.4)	86 (41.8)
Imipeneme	81 (45)	24 (13.3)	75 (41.7)
Cefepime	52 (28.9)	16 (8.9)	112 (62.2)
Piperacillin	111 (61.7)	25 (13.9)	44 (24.4)
Amikacin	39 (21.7)	33 (18.3)	108 (60)
Piperacillin-tazobactam	38 (21.1)	46 (25.6)	96 (53.3)
Gentamicine	54 (30)	8 (4.4)	118 (65.6)

3.6. Results of REP-PCR

The next step was determining the genetic relationship between the samples. After drawing the dendrogram for the obtained results from REP-PCR (Figure 5), ESBL-producing samples which were patterned as the samples that had 100% genetic similarity, were considered as one pattern and, other samples each were considered as a separate pattern. Based on this, 89 patterns exist among 89 ESBL-producing samples; so 89 positive-ESBL samples, had 89 different genotypes (Figure 6).

In this dendrogram, 10 clusters which labeled by letters A-J can be observed. Each cluster A-C-D-F-G-H is divided by two sub-clusters.

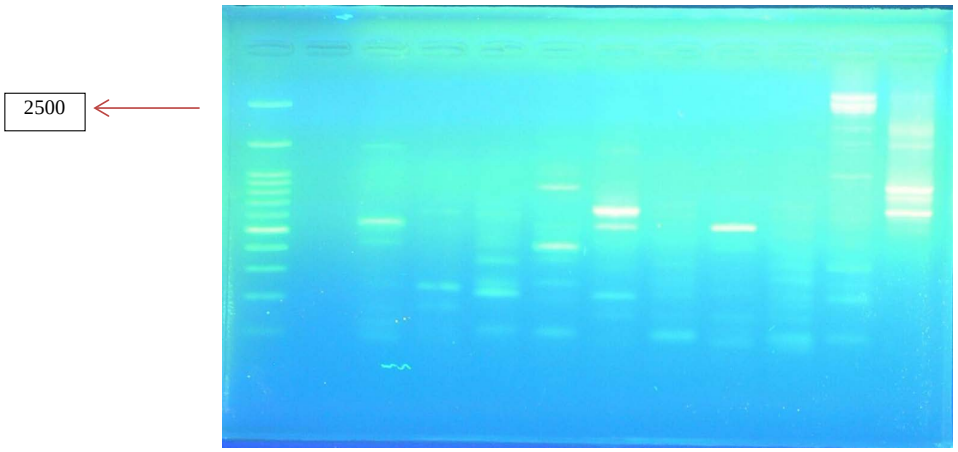


Figure 5. Bands created by rep-PCR for sample type.

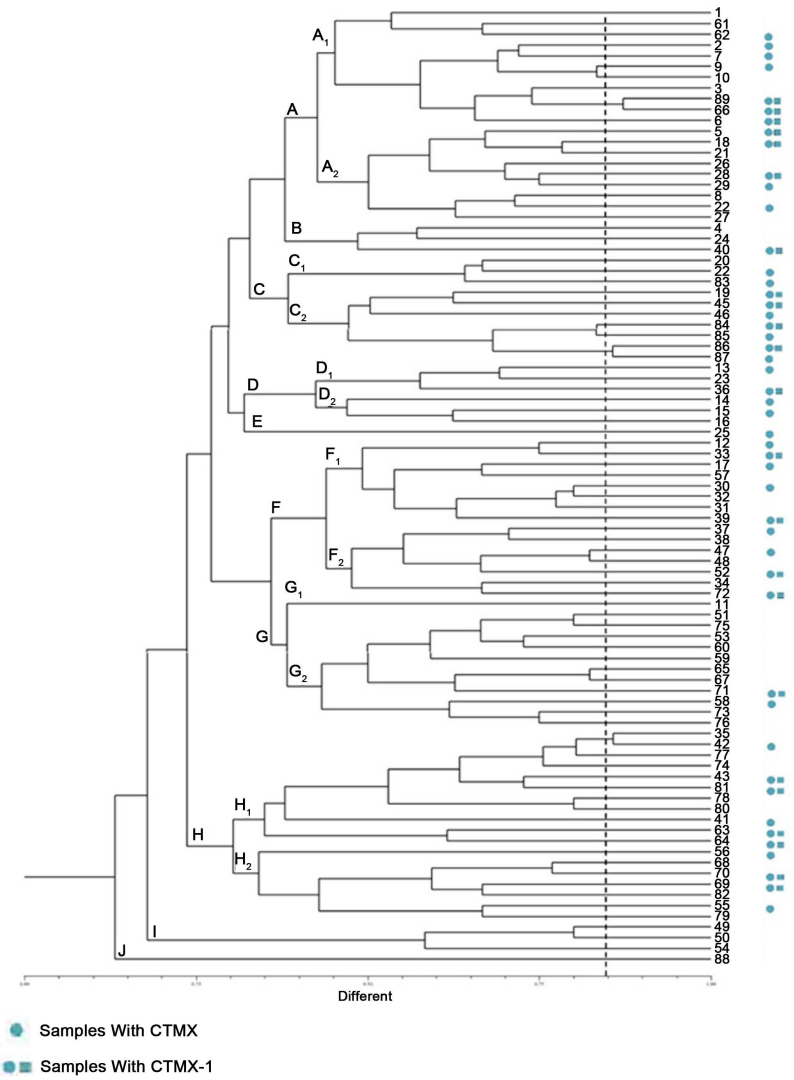


Figure 6. Dendrogram related to rep-PCR analysis shows 89 genetic patterns in 89 ESBL producing samples.

4. Discussion

In 1992, a new type of ESBL that gave high level of resistance against Cefotaxime to bacteria was detected in members of *Antrobactericeas* [16] [17]. This new family of ESBLs are from the class A in the Ambler's classification. As mentioned before, CTX-M is described pandemic. According to several reports, the number of CTX-M B. lactamases is rapidly increasing [18]. In some reports from France [19], Sweden [20], and India [21] as well, the prevalence of this B. In a study performed in Tabriz on 188 *E. coli* separated from urine samples of outpatient and hospitalized patients, it was revealed that 84.1% of the isolates were contained the CTX-M type 1 B. lactamase gene [22]. In order to find a suitable strategy for stopping further spread of this gene, worldwide studies are required. According to **Figure 1**, UTI is most prevalent in ages between 16 - 30 and 31 - 45. This result could be attributed to this fact that most sexual intercourses occur in these periods. Also, UTI was more prevalent in women, which seem logical because of anatomical reasons.

In this study, CTX-M gene was detected in the *E. coli* isolated separated from UTIs in a specified period (2015). 48 out of 89 ESBL-producing samples (53.93%) contained the CTX-M gene. In addition, the results indicated that 23 samples (47.91%) of these 48, contained CTX-M group 1. Our findings showed the high prevalence of CTX-M enzyme in the ESBL-producing *E. coli* in the Sanandaj. Furthermore, it was observed that almost half of these enzymes were from the CTX-M group 1. CTX-M-15, which is in CTX-M group 1, has the most rate of prevalence worldwide [2] [23] [24]. As it can be seen in **Figure 2**, ESBL-producing samples had high resistance against Cefotaxime (97.75%), Ciprofloxacin (78.65%), Ceftriaxone (74.15%), Carbenicillin (65.16%) Ceftazidime (88.76%), Imipenem (74.15%), Piperacillin (87.64%) and Piperacillin-tazobactam (61.79%) and, a moderate resistance against gentamicin (40.44%), Amikacin (50.56%) and Cefepime (52.8%).

In a study in India, resistance of ESBL-producing isolates against non-beta lactam antibiotics were as follows: 93.8% to Ciprofloxacin, 79.1% to Sulfamethoxazole and 14.7% to Amikacin [21]. It is possible that genes that code resistance against these antibiotics are transferred alongside the ESBL genes. In a study conducted in the USA, among 20 isolated bacteria resistant to antibiotics that were separated from patients from hospitals and nursing homes, 17 bacteria contained 54-kb plasmid that endoced the resistance to Ceftazidime by TEM-10.This plasmid was the mediator of resistance against Co-trimoxazole, Gentamicin and Tobramycin [25].

According to reports, CTX-M B. lactamases hydrolyze Cefotaxime more than Ceftazidime [26]. In the present study, 95.83% of the CTX-M B. lactamase-producing isolates that were detected in the study, were resistant to or an intermediary for Cefotaxime, while lack of sensitivity to Ceftazidime was equal to 87.5%.

In order to differentiate between the two following hypotheses, the REP-PCR method was necessary. 1) An epidemic *E. coli* strain had been spread among all the patients, so one ancestral strain is possibly the cause of spread of resistance. 2) CTX-M gene had been spread among different *E. coli* isolated.

This study recognized 48 different genotypes as positive among 48 CTX-M samples; therefore, the results of this experiment indicated that the clonal propagation theory of one epidemic *E. coli* strain is not applicaple. This means that not all the types of CTX-M producers were originated from one single strain and, the gene had been spread among different isolates. Therefore, it can be concluded that one plasmid or mobile genetic element (MGE) containing the CTX-M gene, is responsible for the spread of the gene among different isolated of *E. coli*.

In the drawn dendrogram, it was observed that the samples in Clusters B, E, I and J do not contain the CTX-M-resistant gene and the number of samples in these clusters is very low. For instance, in clusters E and J there are only one sample and in clusters B and I, there are 3 samples and based on that, it can be concluded that the samples without the CTX-M gene, have a lower survival rate and their spread among the patients are lower.

5. Conclusions

In this study genotyping of *E. coli* from urinary tract of infection patients containing B-lactamase resistance gene CTX-M group 1 was assessed. 48 out of 89 ESBL-producing samples (53.93%) contained the CTX-M gene. In addition, the results indicated that 23 out of (47.91%) of these 48 samples, contained CTX-M group 1. The samples without the CTX-M gene, have a lower survival rate and the spread among the patients is lower.

Our findings showed the high prevalence of CTX-M enzyme in the ESBL-producing *E. coli* in the Sanandaj. Also, UTI was more prevalent in women, which seemed logical because of anatomical reasons. According to the obtained results from REP-PCR, it was concluded that the resistant genes were spread among different isolates. So hospitals and their staff must be more hygiene and, proper disposal of hospital waste and using antibiotics only by the doctors' order can help to prevent the spread of resistances.

6. Limitations and Recommendations for Future Research

Limitations of the current study were lack of proper access to some of the reagents and instruments in the tests and due to financial constraints, the study was under-powered, and because of small sample size, it is impossible to generalize the study results, certainly. For future studies, the results of study recommended that in order to generalize the results, study be repeated with big enough population of patients. Also, it is suggested that other ESBL gens prevalence and risk factors related to the spread of ESBL genes can be studied in future.

Acknowledgements

This is a part of Arezoo Omati's M.Sc. thesis. The authors wish to extend their gratitude to Rasht branch of Islamic Azad University for its financial support.

References

[1] Coque, T.M., Oliver, A., Pérez-Díaz, J.C., Baquero, F. and Cantón, R. (2002) Genes Encod-

ing TEM-4, SHV-2, and CTX-M-10 Extended-Spectrum β-Lactamases Are Carried by Multiple *Klebsiella pneumoniae* Clones in a Single Hospital (Madrid, 1989 to 2000). *Antimicrobial Agents and Chemotherapy*, **46**, 500-510. http://dx.doi.org/10.1128/AAC.46.2.500-510.2002

[2] Bonnet, R. (2004) Growing Group of Extended-Spectrum β-Lactamases: The CTX-M Enzymes. *Antimicrobial Agents and Chemotherapy*, **48**, 1-14. http://dx.doi.org/10.1128/AAC.48.1.1-14.2004

[3] Cantón, R. and Coque, T.M. (2006) The CTX-M β-Lactamase Pandemic. *Current Opinion in Microbiology*, **9**, 466-475. http://dx.doi.org/10.1016/j.mib.2006.08.011

[4] Vervoort, J., Baraniak, A., Gazin, M., Sabirova, J., Lammens, C., Kazma, M., *et al*. (2012) Characterization of Two New CTX-M-25-Group Extended-Spectrum β-Lactamase Variants Identified in *Escherichia coli* Isolates from Israel. *PLoS ONE*, **7**, e46329. http://dx.doi.org/10.1371/journal.pone.0046329

[5] Barlow, M., Reik, R.A., Jacobs, S.D., Medina, M., Meyer, M.P., McGowan Jr., J.E., *et al*. (2008) High Rate of Mobilization for blaCTX-Ms. *Emerging Infectious Diseases*, **14**, 423-428. http://dx.doi.org/10.3201/eid1403.070405

[6] Woodford, N., Carattoli, A., Karisik, E., Underwood, A., Ellington, M.J. and Livermore, D.M. (2009) Complete Nucleotide Sequences of Plasmids pEK204, pEK499, and pEK516, Encoding CTX-M Enzymes in Three Major *Escherichia coli* Lineages from the United Kingdom, All Belonging to the International O25:H4-ST131 Clone. *Antimicrobial Agents and Chemotherapy*, **53**, 4472-4482. http://dx.doi.org/10.1128/AAC.00688-09

[7] Nielsen, J.B., Skov, M.N., Jørgensen, R.L., Heltberg, O., Hansen, D.S. and Schønning, K. (2011) Identification of CTX-M15-, SHV-28-Producing *Klebsiella pneumoniae* ST15 as an Epidemic Clone in the Copenhagen Area Using a Semi-Automated Rep-PCR Typing Assay. *European Journal of Clinical Microbiology & Infectious Diseases*, **30**, 773-778. http://dx.doi.org/10.1007/s10096-011-1153-x

[8] Versalovic, J., Schneider, G.M., De Bruijn, F. and Lupski, J.R. (1994) Genomic Fingerprinting of Bacteria Using Repetitive Sequence-Based Polymerase Chain Reaction. *Methods in Molecular and Cellular Biology*, **5**, 25-40.

[9] Olive, D.M. and Bean, P. (1999) Principles and Applications of Methods for DNA-Based Typing of Microbial Organisms. *Journal of Clinical Microbiology*, **37**, 1661-1669.

[10] Woodford, N., Ward, M., Kaufmann, M., Turton, J., Fagan, E., James, D. and Cheasty, T. (2004) Community and Hospital Spread of *Escherichia coli* Producing CTX-M Extended-Spectrum β-Lactamases in the UK. *Journal of Antimicrobial Chemotherapy*, **54**, 735-743. http://dx.doi.org/10.1093/jac/dkh424

[11] Nasehi, L., Shahcheraghi, F., Nikbin, V. and Nematzadeh, S. (2010) PER, CTX-M, TEM and SHV Beta-Lactamases in Clinical Isolates of *Klebsiella pneumoniae* Isolated from Tehran, Iran. *Iranian Journal of Basic Medical Sciences*, **13**, 111-118.

[12] Turnidge, J.D. (2015) Susceptibility Test Methods: General Considerations. Manual of Clinical Microbiology. 11th Edition, American Society of Microbiology.

[13] Barguigua, A., El Otmani, F., Talmi, M., Reguig, A., Jamali, L., Zerouali, K., *et al*. (2013) Prevalence and Genotypic Analysis of Plasmid-Mediated [Beta]-Lactamases among Urinary *Klebsiella pneumoniae* Isolates in Moroccan Community. *Journal of Antibiotics*, **66**, 11-16. http://dx.doi.org/10.1038/ja.2012.91

[14] Ramazanzadeh, R. (2010) Etiologic Agents and Extended-Spectrum Beta-Lactamase Production in Urinary Tract Infections in Sanandaj, Iran. *Eastern Journal of Medicine*, **15**, 57-62.

[15] Bou, G., Cartelle, M., Tomas, M., Canle, D., Molina, F., Moure, R., *et al.* (2002) Identification and Broad Dissemination of the CTX-M-14 Beta-Lactamase in Different *Escherichia coli* Strains in the Northwest Area of Spain. *Journal of Clinical Microbiology*, **40**, 4030-4036. http://dx.doi.org/10.1128/JCM.40.11.4030-4036.2002

[16] Barthélémy, M., Péduzzi, J., Bernard, H., Tancrède, C. and Labia, R. (1992) Close Amino Acid Sequence Relationship between the New Plasmid-Mediated Extended-Spectrum β-Lactamase MEN-1 and Chromosomally Encoded Enzymes of *Klebsiella oxytoca*. *Biochimica et Biophysica Acta*, **1122**, 15-22. http://dx.doi.org/10.1016/0167-4838(92)90121-S

[17] Bauernfeind, A., Holley, M., Jungwirth, R., Mangold, P., Röhnisch, T., Schweighart, S., *et al.* (1992) A New Plasmidic Cefotaximase from Patients Infected with *Salmonella typhimurium*. *Infection*, **20**, 158-163. http://dx.doi.org/10.1007/BF01704610

[18] Al-Jasser, A.M. (2006) Extended-Spectrum Beta-Lactamases (ESBLs): A Global Problem. *Kuwait Medical Journal*, **38**, 171-185.

[19] Lavigne, J.P., Marchandin, H., Delmas, J., Moreau, J., Bouziges, N., Lecaillon, E., *et al.* (2007) CTX-M Beta-Lactamase-Producing *Escherichia coli* in French Hospitals: Prevalence, Molecular Epidemiology, and Risk Factors. *Journal of Clinical Microbiology*, **45**, 620-626. http://dx.doi.org/10.1128/JCM.01917-06

[20] Fang, H., Ataker, F., Hedin, G. and Dornbusch, K. (2008) Molecular Epidemiology of Extended-Spectrum Beta-Lactamases among *Escherichia coli* Isolates Collected in a Swedish Hospital and Its Associated Health Care Facilities from 2001 to 2006. *Journal of Clinical Microbiology*, **46**, 707-712. http://dx.doi.org/10.1128/JCM.01943-07

[21] Goyal, A., Prasad, K.N., Prasad, A., Gupta, S., Ghoshal, U. and Ayyagari, A. (2009) Extended Spectrum Beta-Lactamases in *Escherichia coli* & *Klebsiella pneumoniae* & Associated Risk Factors. *The Indian Journal of Medical Research*, **129**, 695-700.

[22] Soltan Dallal, M.M., Azarsa, M., Shirazi, M.H., Owlia, P., Sabbaghi, A., Shamkani, F., *et al.* (2011) The Prevalence of Extended-Spectrum Beta-Lactamases and CTX-M-1 Producing *Escherichia coli* in Urine Samples Collected at Tabriz City Hospitals. *Tehran University of Medical Sciences*, **69**, 273-278.

[23] Rossolini, G.M., D'Andrea, M.M. and Mugnaioli, C. (2008) The Spread of CTX-M-Type Extended-Spectrum Beta-Lactamases. *Clinical Microbiology and Infection*: *The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, **14**, 33-41. http://dx.doi.org/10.1111/j.1469-0691.2007.01867.x

[24] Rogers, B.A., Sidjabat, H.E. and Paterson, D.L. (2011) *Escherichia coli* O25b-ST131: A Pandemic, Multiresistant, Community-Associated Strain. *The Journal of Antimicrobial Chemotherapy*, **66**, 1-14. http://dx.doi.org/10.1093/jac/dkq415

[25] Wiener, J., Quinn, J.P., Bradford, P.A., Goering, R.V., Nathan, C., Bush, K., *et al.* (1999) Multiple Antibiotic-Resistant Klebsiella and *Escherichia coli* in Nursing Homes. *JAMA*, **281**, 517-523. http://dx.doi.org/10.1001/jama.281.6.517

[26] Bradford, P. (2001) Extended-Spectrum Beta-Lactamases in the 21st Century: Characterization, Epidemiology, and Detection of This Important Resistance Threat. *Clinical Microbiology Reviews*, **14**, 933-951. http://dx.doi.org/10.1128/CMR.14.4.933-951.2001

Call for Papers

American Journal of Molecular Biology (AJMB)

ISSN 2161-6620 (Print) ISSN 2161-6663 (Online)
http://www.scirp.org/journal/ajmb

American Journal of Molecular Biology (AJMB) is an international journal dedicated to the latest advancement of molecular biology. The goal of this journal is to provide a platform for scientists and academicians all over the world to promote, share, and discuss various new issues and developments in different areas of molecular biology. All manuscripts must be prepared in English, and are subject to a rigorous and fair peer-review process. Accepted papers will immediately appear online followed by printed hard copy.

Editorial Board

The journal publishes original papers including but not limited to the following fields:

- Algorithms for Molecular Biology
- Biochemistry and Molecular Biology
- Biophysics and Molecular Biology
- Cellular & Molecular Biology
- Electrophoresis
- Gene Expression
- Genomics
- GH Family
- Insect Biochemistry and Molecular Biology
- Insect Molecular Biology
- Microbiology and Molecular Biology
- Molecular Biology and Evolution
- Molecular Biology of the Cell
- Molecular Biology Techniques
- Molecular Evolution
- Molecular Genetics
- Molecular Neuroscience
- Peptides
- Physiology and Molecular Biology of Plants
- Plant Molecular Biology
- Proteins
- Steroid Biochemistry and Molecular Biology

We are also interested in: 1) Short reports—2 - 5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data; 2) Book reviews—Comments and critiques.

Notes for Intending Authors

Submitted papers should not be previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. For more details, please access the website.

Website and E-Mail

http://www.scirp.org/journal/ajmb E-mail: ajmb@scirp.org

What is SCIRP?

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

What is Open Access?

All original research papers published by SCIRP are made freely and permanently accessible online immediately upon publication. To be able to provide open access journals, SCIRP defrays operation costs from authors and subscription charges only for its printed version. Open access publishing allows an immediate, worldwide, barrier-free, open access to the full text of research papers, which is in the best interests of the scientific community.

- High visibility for maximum global exposure with open access publishing model
- Rigorous peer review of research papers
- Prompt faster publication with less cost
- Guaranteed targeted, multidisciplinary audience

Scientific Research Publishing

Website: http://www.scirp.org
Subscription: sub@scirp.org
Advertisement: service@scirp.org