

DNA Marker Analysis Reveals Genomic Diversity and Putative QTL Associated with Drupe Traits in *Phyllanthus emblica*

S.M.U.P. Mawalagedera¹, P. Janaththani¹, S.W.M.B. Dunuwille¹, G.A.D. Perera², C.K. Weebadde³, D.S.A. Wijesundara⁴ and S.D.S.S. Sooriyapathirana^{1*}

¹Department of Molecular Biology and Biotechnology, University of Peradeniya, Peradeniya, Sri Lanka.

²Department of Botany, University of Peradeniya, Peradeniya, Sri Lanka.

³Departments of Plant, Soil and Microbial Sciences and Horticulture, College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI, USA.

⁴Department of Botanic Gardens, Royal Botanic Gardens, Peradeniya, Sri Lanka.

Accepted October 26, 2014

ABSTRACT

Phyllanthus emblica L. (V. Nelli) is an important constituent of indigenous medicine and a commercially important fresh fruit species. The genomic diversity of the *P. emblica* germplasm in Sri Lanka has not been studied in detail. Therefore, the present study was conducted to molecularly characterize the *P. emblica* germplasm in Sri Lanka. Young tender leaves from 66 trees of *P. emblica* were collected from various parts of the country for DNA assessment. Genomic DNA was extracted using DNeasy Plant mini kit. Six microsatellite (SSR) markers and one SCAR marker developed for *P. emblica* were used in the PCR amplification. The PCR products were size separated by using 6% denaturing polyacrylamide gel electrophoresis. A binary scoring matrix was constructed on presence and absence of alleles in each single locus and a dendrogram based on genetic similarities was constructed. For each allele, allele frequency, heterozygosity and polymorphic information content were calculated. Cluster analysis was conducted based on marker data and association analysis was carried out to detect putative QTL associated with drupe traits. The dendrogram based on seven DNA marker alleles, showed two major clusters at 14.67% molecular similarity coefficient. A total of 51 alleles were detected in the sampled *P. emblica* germplasm with the six SSR markers and the SCAR marker used. Out of the six SSR markers, four were polymorphic, one was monomorphic and the other one did not produce clear bands. Heterozygosity of the tested markers ranged from 0.44 to 0.89. Allele at 176 bp of the marker *Phyll112* showed a significant association ($P < 0.001$ and $P < 0.0001$) with commercially important traits such as drupe height, width, weight and mesocarp thickness. The *P. emblica* germplasm has a higher variation in Sri Lanka and this diversity can be used to develop genetically superior varieties in the future and also to dissect underlying molecular genetic mechanisms of drupe development.

Keywords: Drupe traits, genomic diversity, Indian gooseberry, Quantitative Trait Loci (QTL)

INTRODUCTION

Phyllanthus emblica is a medicinally important fruit crop species. It is a member of the family Phyllanthaceae (formerly, Euphorbiaceae) and occurs naturally in South Eastern Asia (Dassanayake and Fosberg, 1988). All parts of the tree are utilized in indigenous medicine, yet drupe plays the central role as a remedy to a number of ailments (Krishnaveni and Mirunalini, 2011; Daisy and Rajathi, 2009). Inscriptions indicate that *P. emblica* was used in Sri Lankan indigenous medicine as far back as 350 AD. In ancient literature, it is referred to as “*Ambuli*” (Compendium of Medicinal Plants, 2002). *P. emblica* has a high economic value because of its medicinal and nutritive richness. Therefore, it is used to produce cosmetics, commercial beverages, confectionaries and medicinally important value

added products. Fruit itself has a high demand as fresh fruit in the market. Thus, the mature drupe size and the bitterness of *P. emblica* are the most important traits for commercial exploitation. In modern era where there is a boost in ‘global herbal market’, *P. emblica* has a high potential as a cash crop in Sri Lanka. In India alone 10,000 tons of *P. emblica* drupes are commercially utilized per annum (Sharma *et al.*, 2008a).

P. emblica germplasm in Sri Lanka exhibits vast diversity in drupe morphology (Pushpakumara *et al.*, 2007 and Mawalagedera *et al.*, 2014). Some trees bear small bitter drupes while others produce large and less bitter drupes. The fresh fruit market prefers the larger and less bitter drupes. Therefore, there is a grower driven selection towards the large and less bitter drupe bearing trees. However, small and bitter drupe types are immensely important to

*Corresponding author’s email: sunethss09@gmail.com

maintain the *P. emblica* diversity and mostly to preserve its medicinal value. The grower driven selection leads to uprooting of the small drupe bearing trees causing genetic erosion, threatening the use of it in indigenous medicine.

The diversity of drupe traits has been examined on the basis of drupe morphological traits of 66 trees observed in various parts of Sri Lanka (Mawalagedera *et al.*, 2014). Identifying genetic loci controlling drupe size variation would serve as an aid for selecting trees for marker assisted breeding of superior varieties (preferably those with large and bitterer drupe bearing trees), which could ease the tedious processing of drupes in indigenous medicine. Assuming drupe size is a quantitative trait, as in case of the most fruit crops, dissecting QTL associated with the trait could also help in identifying trees capable of producing large drupes at the seedling stage through DNA marker screening. This will undoubtedly help growers to obtain a yield with desired drupe size that could capture both local and global fresh fruit markets. Furthermore, it will assist in effective conservation of germplasm by identifying a core collection for *P. emblica* germplasm in Sri Lanka avoiding duplicates and ensuring that the medicinally important portion of the germplasm is conserved. However, to do so, the diversity of *P. emblica* germplasm in Sri Lanka has to be molecularly characterized, which to our knowledge has not yet been conducted in the past.

The current study was planned with the objectives of assessing the DNA based genetic diversity, comparing molecular vs. morphological diversity, identifying candidate QTL associated with drupe traits and making inferences on the ploidy level of *P. emblica*.

MATERIALS AND METHODS

Sample collection

Young tender leaves were collected from 66 trees of *Phyllanthus emblica* from various parts of Kandy, Kurunegala and Anuradhapura districts of Sri Lanka. Leaf samples were stored in -80°C until the extraction of DNA.

DNA extraction

DNeasy Plant mini kit (Qiagen, Valencia, California, USA) was used to extract DNA from 0.1 g of finely powdered leaf in liquid nitrogen following manufacturer's instructions. Genomic DNA was visualized by 1% agarose gel

electrophoresis and quantified by UV spectroscopy (Cintra10e, GBC, Australia). Absorbance was measured at 260 nm to determine the DNA concentration (Sambrook and Russell, 2001).

Polymerase Chain Reaction (PCR)

PCR for diversity analysis

Six SSR markers (Pandey and Changtragoon, 2012) and one SCAR marker (Dnyaneshwar *et al.*, 2006) developed for *P. emblica* were used for PCR amplification (Table 1). Initial screening for successful PCR amplification was carried out with 12 randomly selected DNA samples from *P. emblica*. DNA amplification was performed in 15- μl reactions containing 1 \times GoTaq[®] Green Master Mix (Promega Corporation, Madison, Wisconsin, USA), 0.5 μM each of forward and reverse primer and 1 μl of DNA template (60ng/ μl). Amplifications were carried out in Thermal Cycler (Takara, Otsu, Shiga, Japan) using the PCR cycle: Initial denaturation: 5 minutes at 94°C ; 35 cycles of 45 seconds at 94°C , 1.30 minutes at annealing temperature (see Table 1 for annealing temperatures of each pair of primer), 2 minutes at 72°C ; and a final extension step of 10 minutes at 72°C . PCR products were visualized in ethidium bromide stained 1% agarose gel after electrophoresis. After confirming PCR amplification, all the successful DNA samples were screened by using all the seven pairs of primers in the PCR.

PCR for DNA barcoding

The *matK* forward and reverse primers were used for PCR amplification of four DNA samples of individuals belonging to four different clusters based on the morphological similarity dendrogram explained in Mawalagedera *et al.*, (2014). Amplifications were carried out in Thermal Cycler (Takara, Otsu, Shiga, Japan) using the PCR cycle: Initial denaturation: 10 minutes at 94°C ; 35 cycles of 20 seconds at 94°C , 20 seconds at annealing temperature 48°C , 2 minutes at 72°C ; and a final extension step of 10 minutes at 72°C . PCR products were visualized in ethidium bromide stained 1% agarose gel after electrophoresis.

DNA sequencing

The PCR amplicons of *matK* of four DNA samples of individuals belonging to four different clusters explained in Mawalagedera *et al.*, (2014) were sequenced using Sanger sequencing ABI 3500 series Genetic Analyzer (Applied Biosystems[®]).

Table 1: Sequences and annealing temperatures of the DNA markers

Locus	Forward and Reverse primer sequences (5'→3')	Annealing temperature (°C)
<i>Phyll7</i>	CGGGAAAGAGAAACGAAATG GCATCAGGTGGACTTCTTGG	67
<i>Phyll13</i>	AAGATCCGGCTTTAAAACCTTTG GCTAGCACTCTTCCCTTCTTGC	65
<i>Phyll31</i>	AACTGGTGACTCCCCCTTTACTC TCCTTGGCTGAATTTTGGAG	50
<i>Phyll53</i>	CTTTCTCCAGCCACCAAATG GTTGGTGGGTTTTCAACCTG	52
<i>Phyll68</i>	CAGGGACATTACACGGACAAC CAGCCTAAGACAACCTCTCATTTACC	53
<i>Phyll112</i>	TCGCTTTTATTTTCTTCAGTTCC AAACCCACTGAGCATGAACC	50
SCAR	CAGATCTCGTGTA AAAAGCGTTG TGCAGTGAATTCCAAGTGTTC	55

(Dnyaneshwar *et al.*, 2006; Pandey and Changtragoon, 2012)

Denaturing Polyacrylamide Gel Electrophoresis

The PCR products of the SSR markers of the DNA samples of 66 *P. emblica* trees were then size separated by using 6% denaturing polyacrylamide gel electrophoresis.

Data analysis

Allele Diversity

Different alleles for each marker were visually identified based on the length of the fragment (bp) and a binary (1= allele present, 0= allele absent) scoring matrix was built for all the markers amplified in the 66 *P. emblica* trees. For each allele, allele frequency (Pi), heterozygosity (H) and polymorphic information content (PIC) were calculated. The H (Botstein *et al.*, 1980) and PIC (Shete *et al.*, 2000) was calculated using the following formula:

$$H = 1 - \sum_{i=1}^n P_i^2$$

$$PIC = 1 - \sum_{i=1}^n P_i^2 - \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n P_i^2 P_j^2 \right]$$

n = Number of alleles

P_i =Frequency of i^{th} allele

Using the allele frequencies, unique alleles (UA), rare alleles (RA) and frequent alleles (FA) were identified.

Cluster analysis

Cluster analysis was performed for 66 trees using all the marker allele data in the binomial data matrix having McQuitty linkage, Euclidean distance using the statistical package Minitab 14.

The 66 trees were given the same tree ID as used in the dendrogram based on drupe morphometric traits explained in Mawalagedera *et al.*, (2014).

Association mapping for QTL analysis

To identify association between marker genotype and trait value, a single marker analysis was conducted. By using the software MapQTL® (Van Ooijen, 2009) marker alleles that were significantly associated with drupe morphometric traits were identified based on Kruskal–Wallis statistics (Kruskal and Wallis, 1952) for the unbalanced natural population.

DNA sequence analysis

Automatic multiple sequence alignment was carried out using the software MEGA 5.2.2 (Tamura *et al.*, 2011) and Clustal Omega free software packages.

RESULTS

Allele composition for the seven DNA markers in *P. emblica*

The PCR performed with DNA extracted from 66 *P. emblica* trees using seven species-specific pair of primers amplified a total of 51 alleles (Table 2). Out of the six microsatellite markers, four exhibited polymorphism, one was monomorphic while the other did not show successful amplification. The SCAR marker was also monomorphic among the entire germplasm tested. Heterozygosity of the tested markers ranged from 0.89 to 0.44. The microsatellite marker *Phyll53* showed the lowest heterozygosity while *Phyll112* showed the highest heterozygosity (Figure 1). Polymorphic information content (PIC) was also high (0.88) in *Phyll112* making it the most

informative marker for the assessment of genetic diversity in *P. emblica* germplasm in Sri Lanka. From those 51 alleles detected, 26 were rare and found in less than 10% of the germplasm, 18 were frequent but seven were unique alleles, found only in a single individual (Table 2).

Cluster analysis for DNA marker alleles

The dendrogram based on seven DNA marker alleles, showed two major clusters at 14.67% molecular similarity co-efficient (Figure 2). Cluster two included 70 % of the germplasm assayed, while cluster one included 30 % only. Cluster two showed higher number of subdivisions than cluster one. The two clusters did not show any biasness or specificity in relation to sampling location.

This dendrogram based on microsatellite DNA marker alleles did not align with the dendrogram constructed by Mawalagedera *et al.*, (2014), using drupe morphometric traits, where four clusters were detected based on drupe traits; height, width, weight and mesocarp thickness. The trees showed a random distribution regardless of their drupe morphometric traits. Yet a high diversity was visible in the dendrogram with no two trees being one hundred percent similar.

The comparison of presence and absence of alleles of the polymorphic microsatellite markers by observing the dendrogram based on drupe morphometric data (Mawalagedera *et al.*, 2014) revealed that, the cluster three bearing the smallest drupes contained four alleles (366 bp, 365 bp, 364

bp and 363 bp) for the marker *Phyll* 68 were unique to that cluster.

Association mapping for QTLs associated with drupe traits

Single marker QTL analysis was conducted to detect marker alleles that are significantly associated with the commercially important drupe traits. The analyses revealed that out of 51 alleles, 15 were significantly associated (Appendix 1) in determining drupe weight, width, height and mesocarp thickness.

Of these markers, allele 176 bp of *Phyll*112, (Table 3) showed the most significant association ($P<0.0001$) with commercially important traits; drupe height, width, weight and mesocarp thickness. Furthermore, 363 bp allele of *Phyll*68 also showed a significant association ($P<0.001$) with drupe height. In contrast, the stone traits were significantly associated with a different set of marker alleles than the drupe traits including the 150 bp allele of *Phyll*31, and the 368 bp, 174 bp and 172 bp alleles of *Phyll*68.

Inferences on ploidy level of *P. emblica*

When the most informative DNA markers, *Phyll*31 and *Phyll*112 were considered they yielded more than two alleles per single locus. For all the four informative markers the number of alleles produced for a single individual ranged from one to nine. Thus, these markers resulted in multiple alleles for the *P. emblica* germplasm in Sri Lanka implying the presence of polyploidy.

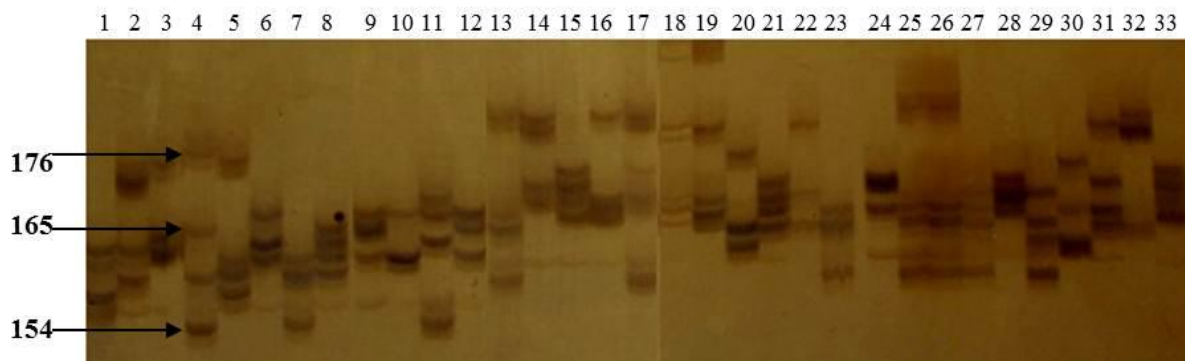


Figure 1: Polymorphic alleles of *Phyll*112 for 32 (Nos. 1-32) trees of *P. emblica*

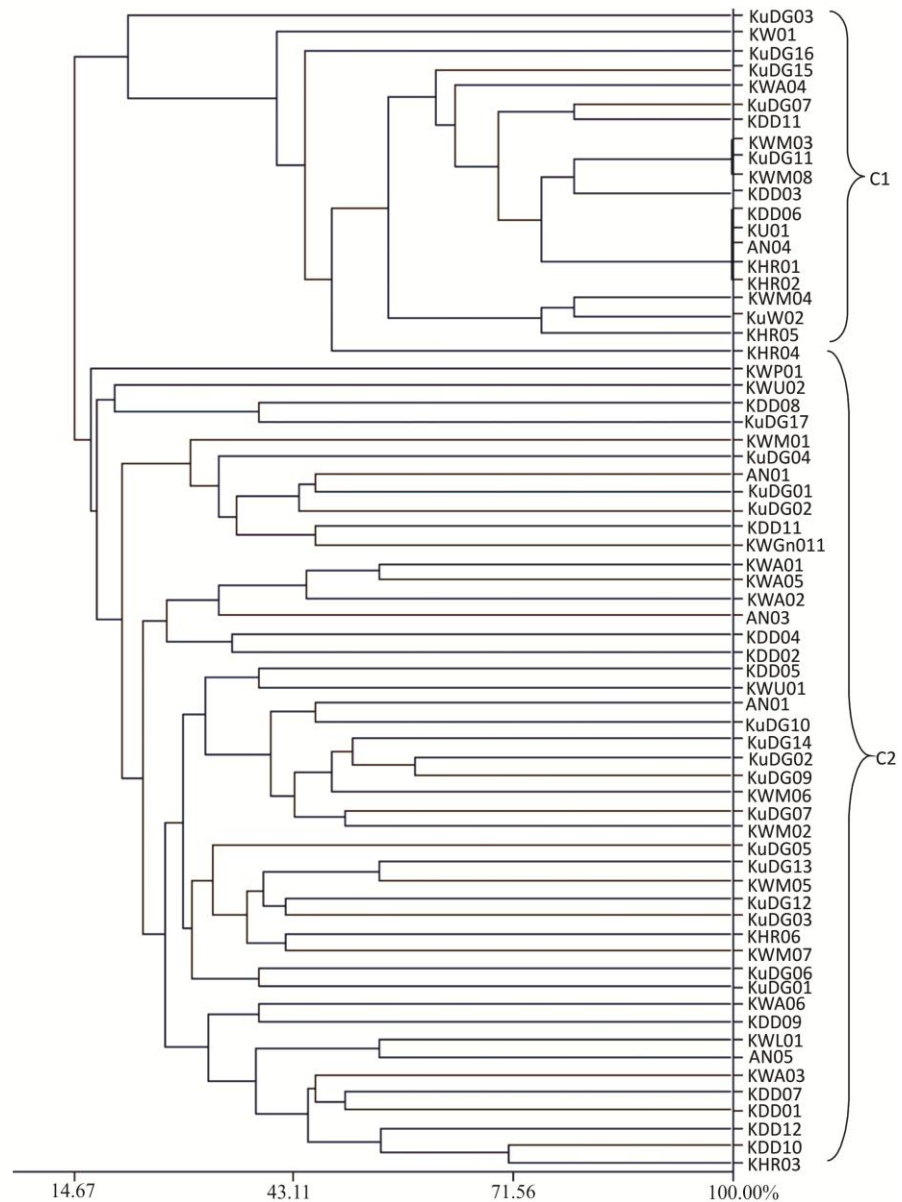


Figure 2: Dendrogram representing molecular similarity of 66 *Phyllanthus emblica* trees from different locations. The tree was clustered based on seven species-specific marker alleles. Two clusters were identified at 14.67% of molecular similarity coefficient. The two clusters are indicated as C1 and C2 using two right brackets. KHR-Kandy hilly Rikillagaskada, KWGn-Kandy wet Gannoruwa, KWM-Kandy wet Meewathura, KWG-Kandy wet Gatambe, KWA-Kandy wet Alupathakada, KWL- Kandy wet Lewella, KWP-Kandy wet Pitakanda, KWU-Kandy wet Uduwela, KDD- Kandy dry Digana, KuDG-Kurunegala dry Ganewatte, KuW-Kurunegala wet, An- Anuradhapura.

DNA barcoding

The DNA sequences of *matK* of four DNA samples of individuals belonging to four different clusters based on fruit size explained in

Mawalagedera *et al.*, (2014) showed no polymorphism, indicating fruit size differences are not caused by species level differences (Appendix 2).

Table 2: Composition of alleles in the sampled *P. emblica* germplasm

Locus	Alleles/ band (bp)	Abundance of allele	<i>Pi</i>	<i>H</i>	<i>PIC</i>			
<i>Phyll7</i>	175	FA	1.00	0.00	0.00			
<i>Phyll31</i>	168	RA	0.07	0.87	0.86			
	167	FA	0.15					
	166	FA	0.03					
	165	FA	0.19					
	164	FA	0.17					
	163	FA	0.12					
	162	RA	0.04					
	156	FA	0.10					
	150	FA	0.10					
	149	RA	0.02					
	<i>Phyll53</i>	150	FA			0.67	0.44	0.39
		149	FA			0.33		
	<i>Phyll68</i>	168	FA			0.16	0.88	0.87
170		FA	0.17					
171		FA	0.16					
172		RA	0.01					
173		RA	0.01					
174		RA	0.01					
175		RA	0.01					
363		RA	0.01					
364		UA	0.00					
365		RA	0.01					
366		UA	0.00					
367		RA	0.02					
368		UA	0.00					
369		RA	0.02					
370		RA	0.02					
371		RA	0.09					
372		RA	0.09					
374		FA	0.15					
375		RA	0.03					
<i>Phyll112</i>		150	RA	0.04	0.89	0.88		
	153	FA	0.11					
	154	RA	0.04					
	156	RA	0.09					
	158	FA	0.12					
	165	RA	0.08					
	170	RA	0.05					
	172	RA	0.06					
	172	RA	0.08					
	176	RA	0.02					
	178	RA	0.02					
	179	RA	0.01					
	180	RA	0.02					
	210	UA	0.00					
	220	UA	0.00					
	225	UA	0.00					
	250	UA	0.00					
SCAR	300	FA	0.22	0.00	0.00			
	343	FA	1.00					

Pi: Allele frequency, *H*: Heterozygosity, *PIC*: Polymorphic Information Content, FA: Frequent allele, RA: Rare allele and UA: Unique allele

Table 3: Marker alleles that are significantly associated with drupe quantitative traits.

Locus	Allele/Bands (bp)	K*	Mean a	Mean b
Drupe width (mm)				
<i>Phyll112</i>	176	10.60***	24.42	19.59
	178	8.28***	22.97	19.73
	180	8.53***	23.20	19.71
Drupe height (mm)				
<i>Phyll112</i>	176	7.89****	20.78	17.32
<i>Phyll68</i>	363	6.92***	13.46	17.83
Drupe weight (g)				
<i>Phyll112</i>	176	10.46****	7.63	4.56
	178	8.02****	6.38	4.69
	180	8.15****	6.57	4.67
Mesocarp thickness (mm)				
<i>Phyll112</i>	176	11.64*****	8.19	5.30
	178	8.80****	7.31	5.39
	180	9.34****	7.47	5.37
Stone height (mm)				
<i>Phyll31</i>	150	6.77***	8.53	9.64
Stone weight (g)				
<i>Phyll68</i>	368	8.62****	0.48	0.66
	172	8.12****	0.46	0.60
	174	7.07***	0.49	0.66

*** $P < 0.001$, **** $P < 0.0001$, ***** $P < 0.00001$; K: Kruskal Wallis Index, Mean a: mean at presence of the allele and Mean b: mean at absence of the allele

DISCUSSION

The molecular characterization of the *P. emblica* germplasm revealed that there is high genomic diversity within the species. Though phenotypic traits are commonly used to characterize this diversity, changes in the phenotype due to environmental factors alone can mislead diversity analyses (Ganopoulos *et al.*, 2011). Hence molecular genotyping provides a more accurate insight for identifying and establishing the diversity structure of *P. emblica* germplasm. Molecular markers such as microsatellite markers, Random Amplified Polymorphic DNA (RAPD), Restriction Fragment Length Polymorphic DNA (RFLP), Amplified Fragment Length Polymorphism (AFLP) and Single Nucleotide Polymorphism (SNP) are among the most widely used DNA markers to characterize germplasm at the DNA level. Microsatellite markers for *P.*

emblica were developed with the aim of using them in diversity and population genetic studies (Pandey and Changtragoon, 2012). Genetic diversity studies on sweet cherry (Ganopoulos *et al.*, 2011), peach (Fernandez *et al.*, 2012), wild rice varieties (Feng *et al.*, 2006) and bamboo (Sharma *et al.*, 2008b) are some examples of using SSR markers for genetic diversity assessment in plants.

The molecular diversity analysis of *P. emblica* required intact and quality genomic DNA. The *P. emblica* leaves are rich in phenolic compounds and other secondary metabolites which reduce pH during CTAB protocol of DNA extraction, degrading and yielding low quality DNA (Nagarajan *et al.*, 2011). The early attempts to extract DNA from *P. emblica* leaves using CTAB method too had less success in the present study. Therefore, DNA extraction was carried out using DNeasy Plant mini kit (Qiagen, Valencia, California, USA). For resolving the PCR products

of the amplified *P. emblica* DNA, 6% denaturing polyacrylamide gel electrophoresis was used because it provides a high resolution for DNA of 40 bp – 400 bp, while a 1% agarose gel could only resolve 500bp – 5000bp DNA (Barril and Nates, 2012). The actual DNA product size of the microsatellite markers fell between 100bp – 200bp (Pandey and Changtragoon, 2012). Furthermore, denaturing polyacrylamide gel electrophoresis provides a stable chemically cross-linked gel providing sharper bands even for the low molecular weight DNA (Sambrook and Russell, 2001).

The microsatellite markers (Pandey and Changtragoon, 2012) designed for the Indian *P. emblica* germplasm resulted in successful amplification of the Sri Lankan *P. emblica* genomic DNA. This shows that the *P. emblica* germplasm in these two locations have an evolutionary relatedness. Yet it does not provide evidence about the origin of this species. The microsatellite marker *Phyll53*, which was monomorphic for Indian germplasm (Pandey and Changtragoon, 2012) showed a heterozygosity of 0.44 in the Sri Lankan germplasm. The marker *Phyll7*, which was monomorphic for Sri Lankan *P. emblica* germplasm showed a heterozygosity of 0.75 in the Indian germplasm. The highest heterozygosity was observed with the *Phyll112* marker in both germplasm sources. But the observed heterozygosity was much higher ($H=0.89$) in the Sri Lankan germplasm than the Indian germplasm ($H=0.76$) suggesting that the Sri Lankan *P. emblica* germplasm is more genetically diverse.

The cluster analysis based on molecular data did not support the clustering based on drupe morphometric data (Mawalagedera *et al.*, 2014). It would have been possible to obtain a similar clustering if the DNA markers used were associated with drupe size determining genomic regions of the *P. emblica* genome. In such a study, RAPD markers have been identified in melon, where the markers were linked with the QTL determining fruit weight and width (Park *et al.*, 2005). Similarly in tomato, a locus, *fw.2.2* on the second chromosome was found to be responsible for controlling fruit size (Frary *et al.*, 2000). In the case of *P. emblica*, out of all the alleles, four alleles were unique to the smallest drupe-bearing cluster of the drupe morphological dendrogram. Hence, it highlights the unique genomic composition of the small drupe bearing *P. emblica* trees and the need for conserving this portion of the germplasm in Sri Lanka. If the culling of the small drupe bearing trees continues, it would cause the trees to bear larger fruits as in the case of tomato where the larger fruits were selected by growers over decades

causing the tomato fruit to become progressively larger (Tanksley, 2004), ultimately losing an important portion of the germplasm forever.

Through association mapping putative QTL associated with important drupe traits were detected. Association mapping was used to avoid the limitations of pedigree based QTL mapping (Khan and Korban, 2012) because this sampled *P. emblica* germplasm was a random natural population. Single marker analysis revealed that 15 alleles were significantly associated (Appendix 1) in determining drupe weight, width, height and mesocarp thickness. The 176 bp allele of *Phyll112* locus was significantly associated with all the four commercially important drupe traits (drupe; height, width, weight and mesocarp thickness). Based on this it is likely that the same genomic region might be controlling all the four phenotypically related drupe traits. Hence, this marker could be important in marker-assisted selection to determine the drupe size and flesh thickness.

For diploid species maximum of two alleles are found at a locus. But some individuals in the sampled germplasm exhibited as high as nine alleles in a single locus providing a hint of polyploidy. Previous studies indicated that the genus *Phyllanthus* is predominantly polyploid ($X = 13$) and that *P. emblica* might be octoploid (Webster and Ellis, 1962). Yet it has been found that microsatellite markers are evolutionary diverse. These microsatellite regions undergo mutations as DNA slippage and mutations that decrease the length of pure repetitive DNA sequences. Furthermore, it has been found that within-locus polymorphism is positively correlated with allele length due to contraction-mutations in the locus (Ovidiu and Hörandl, 2006). Therefore, it is premature to make inferences on ploidy level of *P. emblica* based on the number of alleles at a locus without cytogenetic analysis.

DNA barcode for *matK* was ineffective to distinguish between the four clusters based on drupe morphological similarity explained by Mawalagedera *et al.*, (2014). Since the barcode for *matK* resulted in 100% similarity it can be concluded that the variation in drupe size is not due to species level differences within the *P. emblica* germplasm in Sri Lanka. Yet more accurate insight can be obtained through detailed barcoding based on several barcoding genes.

In summary, these results showed that *P. emblica* germplasm in Sri Lanka exhibit a significantly high genomic diversity, which can be exploited to develop superior varieties using the putative QTLs. Further studies are required to assess the

cytogenetic features of *P. emblica* germplasm in Sri Lanka.

ACKNOWLEDGMENTS

Authors wish to acknowledge the University of Peradeniya (Research Grant No.: RG/2013/15/S) for financial assistance and the Michigan State University, East Lansing, Michigan, USA for their support.

REFERENCES

- Barril, P. and Nates, S. (2012). Introduction to agarose and polyacrylamide gel electrophoresis matrices with respect to their detection sensitivities. In: S. Magdeldin (eds), Gel Electrophoresis - Principles and Basics. Intech China
- Botstein, D., White, R. L., Skalnicky, M. H. and Davies, R. W. (1980). Construction of genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics* **32**: 314-331.
- Compendium of Medicinal plants, A Sri Lankan study, (2002). Department of Ayurveda, Sri Lanka.
- Daisy, P. and Rajathi. M. (2009). Hypoglycemic and hypolipidemic effects of *Phyllanthus* (Euphorbiaceae) fruits in alloxan-induced diabetes in rats. *Tropical Journal of Pharmaceutical Research* **1**: 35-39.
- Dassanayake, M.D. and Fosberg, F.R. (1988). A Revised Hand Book of the Flora of Ceylon. Vol VI, Amerind Publishing Company, New Delhi, India.
- Dnyaneshwar, W., Chavan P., Joshi K. and Patwardhan B. (2006). Development and application of RAPD-SCAR marker for identification of *Phyllanthus emblica* LINN. *Biological and Pharmaceutical Bulletin* **29**: 2313-2316.
- Feng, Z. Y., Liu, X. J., Zhang, Y. Z. and Ling, H. Q. (2006). Genetic diversity analysis of Tibetan wild barley using SSR markers. *Acta Genetica Sinica* **33**: 917- 928.
- Fernandez, M., Angel, B. A., Tyson, T., Carolina, F. F., Amit, D. and Nnadozie, O. (2012). Genetic diversity and relatedness of sweet cherry (*Prunus avium* L.) cultivars based on single nucleotide polymorphic markers. *Frontiers in Plant Science* **3**: 1-13.
- Frary, A., Nesbitt, T. C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B. and Tanksley, S. T. (2000). fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**: 85-88.
- Ganopoulos, I. V., Konstantinos, K., Ioannis, C., Irene, K. and Athanasios, S. T. (2011). Genetic diversity, structure and fruit trait associations in Greek sweet cherry cultivars using microsatellite based (ssr/issr) and morpho-physiological markers. *Euphytica* **181**: 237-251
- Khan, M. A. and Korban, S. S. (2012). Association mapping in forest trees and fruit crops. *Journal of Experimental Botany* **105**: 1-16.
- Krishnaveni, M. and Mirunalini, S. (2011). Amla the role of ayurvedic therapeutic herb in cancer. *Asian Journal of Pharmaceutical and Clinical Research* **4**: 1317.
- Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47**: 583-621.
- Mawalagedara, S.M.U.P., Perera, G.A.D. and Sooriyapathirana, S.D.S.S. (2014). Morphological characterization of drupes reveals a higher diversity of *Phyllanthus emblica* germplasm in Anuradhapura, Kandy and Kurunegala Districts of Sri Lanka. *Ceylon Journal of Science (Biological Science)* **43**: 125-135.
- Nagarajan, S., Mercy, S., Kavitha, M., Nair, R., Sethuraman, T., Alagar, P. and Ganesh, D. (2011). Improved protocol for isolation of genomic DNA from leaf tissues of *Phyllanthus emblica* Gaertn. *Iranian Journal of Biotechnology* **9**: 308- 313
- Ovidiu, P. and Hörandl, E. (2006). Evolution of hypervariable microsatellites in apomictic polyploid lineages of *Ranunculus carpaticola*: directional bias at dinucleotide loci. *Genetics* **174**: 387-398.
- Pandey, M. and Changtragoon, S. (2012). Isolation and characterization of microsatellites in a medicinal plant, *Phyllanthus emblica* (Euphorbiaceae). *American Journal of Botany* **99**: 468-469.
- Park, S. O., Kevin, M. C. and Zhoo, H. K. (2005). Mapping of QTL for fruit size and shape traits in *Ananas melon*. *Report-cucurbit Genetics Cooperative* **28**: 31.
- Pushpakumara, D.K.N.G. and Heenkenda, H.M.S. (2007). Nelli (Amla) *Phyllanthus emblica* L. In: Pushpakumara, D.K.N.G., Gunasena, H.P .M. and Singh, V .P. (eds) Underutilized fruit trees in Sri Lanka. Volume 1. World Agroforestry Centre, South Asia Office, New Delhi, India. Chapter 6: 180-221.
- Sambrook, J. and Russell, D. W. (2001). Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory Press, USA.
- Sharma, A., Shanker, C., Tyagi, L.K., Singh, M. and Rao, C.V. (2008a). Herbal medicine for market potential in india: An overview. *Academic Journal of Plant Sciences* **1**: 26-36.
- Sharma, R. K., Gupta, P., Sharma, V., Sood, A.,

- Mohapatra, T. and Ahuja, P. S. (2008b). Evaluation of rice and sugarcane SSR markers for phylogenetic and genetic diversity analyses in bamboo. *Genome* **51**: 91-103.
- Shete, S., Tiwari, H. and Elston, R. C. (2000). On estimating the heterozygosity and polymorphism information content value. *Theoretical Population Biology* **57**: 265-271.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**: 2731-2739.
- Tanksley, S. D. (2004). The genetic, developmental, and molecular basis of fruit size and shape variation in tomato. *The Plant Cell Online* **16**: 181-189.
- Van Ooijen, J. W. (2009). MapQTL® 6, Software for mapping of quantitative trait loci of experimental populations of diploid species. Kyazma B. V., Wageningen, Netherlands.
- Webster, G. L. and Ellis, J.R. (1962). Cytotaxonomic studies in the Euphorbiaceae, subtribe Phyllanthinae. *American Journal of Botany* **49**: 14-18.

Appendix 1: Putative QTLs significantly associated with drupe traits for DNA markers (a) *Phyll112*, (b) *Phyll113* and (c) *Phyll68*

(a)

Drupe Trait	Alleles/bands (bp) for DNA marker <i>Phyll112</i>										
	153	156	165	172	173	176	178	179	180	193	371
Drupe Width (mm)	-	-	-	-	-	K=10.60*** a=24.42 b=19.59	K=8.28*** a=22.97 b=19.73	K=4.19** a=23.08 b=19.88	K=8.53*** a=23.20 b=19.71	-	-
Drupe Height (mm)	K=2.78* a=16.88 b=18.39	-	-	-	-	K=7.89**** a=20.78 b=17.32	K=5.59** a=19.19 b=17.42	K=3.24* a=19.49 b=17.50	K=5.48** a=19.15 b=17.48	-	-
Drupe Weight (g)	K=3.02* a=4.19 b=5.19	-	-	-	-	K=10.46*** a=7.63 b=4.56	K=8.02**** a=3.68 b=4.69	K=3.82* a=6.47 b=4.77	K=8.15**** a=6.57 b=4.67	-	-
Mesocarp Thickness (mm)	K=3.15* a=5.10 b=6.03	-	-	K=3.02* a=6.19 b=5.35	-	K=11.64**** * a=8.19 b=5.30	K=8.80**** a=7.31 b=5.39	K=4.06** a=7.34 b=5.48	K=9.34**** a=7.47 b=5.37	-	-
Stone Width (mm)	-	-	-	-	-	-	-	-	-	-	-
Stone Height (mm)	-	-	-	-	-	-	-	-	-	-	-
Stone weight (g)	-	-	-	-	-	-	K=3.17* a=0.53* b=0.61	-	K=3.27* a=0.47 b=0.52	K=2.17* a=0.51 b=0.59	-
L*	-	-	K=5.39** a=-40.39 b=-34.19	K=2.72* a=-40.23 b=-35.33	K=5.39** a=-40.39 b=-34.19	K=5.39** a=-40.39 b=-34.19	K=2.81* a=-40.85 b=-36.11	-	K=2.72* a=-40.23 b=-35.53	-	-
a*	-	-	-	-	-	-	-	-	-	-	K=3.24* a=-5.08 b=-4.24
b*	-	-	K=3.00* a=11.15 b=16.72	-	K=3.00* a=11.15 b=16.72	-	-	-	-	-	-
Hue Angle	-	-	K=5.33** a=-22.89 b=-16.30	-	K=5.33** a=-22.89 b=-16.30	-	-	-	-	-	-

Cell Number	-	-	-	K=3.15* a=45.00 b=40.00	-	-	-	-	-	-	-
Cell Size	K=2.77* a=0.11 b=0.18	-	-	-	-	-	-	-	-	-	-
Bitterness	-	K=2.94* a=2.03 b=1.69	-	K=3.94** a=1.46 b=1.39	-	-	-	-	-	-	-

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, K: Kruskal Wallis Index, a: mean trait value at presence of the allele/band and b: mean at absence of the allele/band, '-' indicates that the particular allele/band is not significantly associated with the respective drupe trait.

(b)

Drupe Trait	Alleles/bands (bp) for DNA marker <i>Phyll31</i>							Alleles/bands (bp) for DNA marker <i>Phyll54</i>
	150	156	162	163	164	165	167	149
Drupe Width (mm)	-	K=3.95** a=18.63 b=20.88	K=3.26* a=21.06 b=19.84	-	-	-	-	-
Drupe Height (mm)	-	K=4.11** a=16.33 b=18.43	-	-	K= 3.49* a=16.85 b=18.84	-	-	-
Drupe Weight (g)	-	K=4.01** a=3.71 b=5.53	K=2.76* a=5.07 b=4.80	-	-	-	-	-
Mesocarp Thickness (mm)	-	K=6.57** a=4.71 b=6.08	-	-	-	-	-	-
Stone Width (mm)	-	-	-	-	-	-	-	-
Stone Height (mm)	K=6.77*** a=8.53 b=9.64	-	-	K=3.33* a=8.97 b=9.54	K=2.98* a=9.04 b=9.70	-	-	K=2.72* a=8.92; b=9.67

Stone weight (g)	K=5.96** a=0.46 b=0.61	-	-		K=4.60** a=0.52 b=0.63	K=3.92** a=0.52; b=0.65	-	-
L*	-	-	-	K=5.06** a=- 40.86 b=-35.15	-	-	-	-
a*	-	-	-	-	-	-	-	-
b*	-	-	-	-	-	-	-	-
Hue Angle	-	-	-	K=3.75* a=- 22.98 b=-17.36	K=3.75* a=- 21.71 b=-16.08			-
Cell Number	K=3.15* a=45.00 b=39.00	-	-	-	K=2.91* a=44.00 b=38.00	-	K=2.91* a=44.00 b=38.00	-
Cell Size	-	-	-	-	-	-	-	-
Bitterness	-	-	-	-	-	-	-	-

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, K: Kruskal Wallis Index, a: mean trait value at presence of the allele/band and b: mean at absence of the allele/band, '-' indicates that the particular allele/band is not significantly associated with the respective drupe trait.

(c)

Drupe Trait	Alleles/bands (bp) for DNA marker <i>Phyll68</i>														
	158	168	170	171	172	174	363	365	367	368	369	370	371	372	374
Drupe Width (mm)	-	-	-	-	-	-	K=6.14** a=15.54 b=20.42	K=3.36* a=15.82 b=20.16	-	-	-	-	-	-	-
Drupe Height (mm)	-	-	K=5.10** a=16.77 b=18.73	-	-	-	K=6.92*** a=13.46 b=17.83	K=4.23* * a=13.73 b=17.76	-	-	K=5.75** a=14.61 b=17.88	-	K= 3.29* a=16.35 b=18.19	-	-
Drupe Weight (g)	-	-	-	-	-	-	K=6.29** a=2.06 b=4.98	K=3.36* a=2.34 b=4.92	-	-	K=3.08* a=3.12 b=4.98	-	-	-	-
Mesocarp Thickness (mm)	-	-	-	-	-	-	K=6.45** a=3.43 b=5.67	K=2.77 a=3.82 b=5.62	K=3.30* a=4.20 b=5.68	-	-	-	K=2.73* a=5.03 b=5.80	-	-
Stone Width (mm)	-	-	-	-	-	-	-	K=3.43* a=7.78 b=9.05	-	-	K= 4.34** a=8.05 b=9.07	-	-	-	K=3.53* * a=8.59 b=9.41
Stone Height (mm)	-	K=3.14* a=8.39 b=9.42	-	K=2.75* a=8.74 b=9.51	-	K=5.95* * a=8.85 b=9.88	-	-	-	K=5.20** * a=8.80 b=9.85	K= 4.66** a=8.92 b=9.97	K=5.84** * a=8.85 b=9.92	-	-	-
Stone weight (g)	-	-	-	-	K=8.12** ** a=0.46 b=0.60	K=7.07* ** a=0.49 b=0.66	-	-	-	K=8.62**** * a=0.48 b=0.66	K=3.36* a=0.50 b=0.67	K=6.46** * a=0.49 b=0.67	-	-	-
L*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	K=2.72*a = 40.53 b=- 35.53
a*	K=3.03* a=-5.21 b=-4.30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
b*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hue Angle	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cell Number	-	-	-	-	-	-	-	-	-	-	-	-	K=3.94* *	K=2.91* * a=16.0	-

													a=46.00 b=41.00	0 b=41.0 0		
Cell Size	-	-	-	-	-	-	-	-	-	-	-	-	K=3.00 * a=0.11 b=0.18	-	-	-
Bitterness	-	-	-	-	-	-	-	-	-	-	-	-	K=3.54* a=2.83 b=1.79	-	-	-

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, K: Kruskal Wallis Index, a: mean trait value at presence of the allele/band and b: mean at absence of the allele/band, '-' indicates that the particular allele/band is not significantly associated with the respective drupe trait.

Appendix 2: The alignment of four DNA sequences of the *matK* PCR products belonging to *P. emblica* fruit size clusters explained at Mawalagedera et al., (2014). The package ClustalW was used to align the sequences.

```

1      CCAAATAGGTTGACTAATAGGATGTCCTGCTGCGTTACAAAATTCGTTTTGGCCAATGATCGAATCAAAGGAATAATTGGAATCCTTGTATCAACTTTC 100
2      CCAAATAGGTTGACTAATAGGATGTCCTGCTGCGTTACAAAATTCGTTTTGGCCAATGATCGAATCAAAGGAATAATTGGAATCCTTGTATCAACTTTC
3      CCAAATAGGTTGACTAATAGGATGTCCTGCTGCGTTACAAAATTCGTTTTGGCCAATGATCGAATCAAAGGAATAATTGGAATCCTTGTATCAACTTTC
4      CCAAATAGGTTGACTAATAGGATGTCCTGCTGCGTTACAAAATTCGTTTTGGCCAATGATCGAATCAAAGGAATAATTGGAATCCTTGTATCAACTTTC

1      TTCATAGCATTATCTATTA AAAATGCATTTTCTAGCATTGACTCCGTACCCTGAAAGATTGAGTCGTAGACTTGAAAGATAGCCCAAAAAGTCGAGAG 200
2      TTCATAGCATTATCTATTA AAAATGCATTTTCTAGCATTGACTCCGTACCCTGAAAGATTGAGTCGTAGACTTGAAAGATAGCCCAAAAAGTCGAGAG
3      TTCATAGCATTATCTATTA AAAATGCATTTTCTAGCATTGACTCCGTACCCTGAAAGATTGAGTCGTAGACTTGAAAGATAGCCCAAAAAGTCGAGAG
4      TTCATAGCATTATCTATTA AAAATGCATTTTCTAGCATTGACTCCGTACCCTGAAAGATTGAGTCGTAGACTTGAAAGATAGCCCAAAAAGTCGAGAG

1      ACTGCTTCGATAATGGGTTTATACAGATCTTTGCTGGTTGAAACCACACAGAAAATGACATTGACATAAATTGACAAGGTAATATTTCCATTTTTTTAA 300
2      ACTGCTTCGATAATGGGTTTATACAGATCTTTGCTGGTTGAAACCACACAGAAAATGACATTGACATAAATTGACAAGGTAATATTTCCATTTTTTTAA
3      ACTGCTTCGATAATGGGTTTATACAGATCTTTGCTGGTTGAAACCACACAGAAAATGACATTGACATAAATTGACAAGGTAATATTTCCATTTTTTTAA
4      ACTGCTTCGATAATGGGTTTATACAGATCTTTGCTGGTTGAAACCACACAGAAAATGACATTGACATAAATTGACAAGGTAATATTTCCATTTTTTTAA

1      TAGAAGAGGGGTATCCTTTGAAGCCAAGATTGACTTTCCTTGATATCTAACATAATGCATGAAAGGATCCTTGAACAACCATAGAAATGGCCTGAAAATCA 400
2      TAGAAGAGGGGTATCCTTTGAAGCCAAGATTGACTTTCCTTGATATCTAACATAATGCATGAAAGGATCCTTGAACAACCATAGAAATGGCCTGAAAATCA
3      TAGAAGAGGGGTATCCTTTGAAGCCAAGATTGACTTTCCTTGATATCTAACATAATGCATGAAAGGATCCTTGAACAACCATAGAAATGGCCTGAAAATCA
4      TAGAAGAGGGGTATCCTTTGAAGCCAAGATTGACTTTCCTTGATATCTAACATAATGCATGAAAGGATCCTTGAACAACCATAGAAATGGCCTGAAAATCA

1      TTAGCAAAGACTTCTGCAAAATGTTCTATTTTTCCAAAAAATATATTCGCTCAAGAAGGACCCGAAAAAATGTTGATCGTAAATGAAAGGATTGCTTAC 500
2      TTAGCAAAGACTTCTGCAAAATGTTCTATTTTTCCAAAAAATATATTCGCTCAAGAAGGACCCGAAAAAATGTTGATCGTAAATGAAAGGATTGCTTAC
3      TTAGCAAAGACTTCTGCAAAATGTTCTATTTTTCCAAAAAATATATTCGCTCAAGAAGGACCCGAAAAAATGTTGATCGTAAATGAAAGGATTGCTTAC
4      TTAGCAAAGACTTCTGCAAAATGTTCTATTTTTCCAAAAAATATATTCGCTCAAGAAGGACCCGAAAAAATGTTGATCGTAAATGAAAGGATTGCTTAC

1      GGAGAAAAAAGAAAATGGATTTCATATTCATATACATGAGAATTATATAGAAACAGGAATAATCGTGGATTACTTTTTGCAAAAAAAGAAATCAAATTATT 600
2      GGAGAAAAAAGAAAATGGATTTCATATTCATATACATGAGAATTATATAGAAACAGGAATAATCGTGGATTACTTTTTGCAAAAAAAGAAATCAAATTATT
3      GGAGAAAAAAGAAAATGGATTTCATATTCATATACATGAGAATTATATAGAAACAGGAATAATCGTGGATTACTTTTTGCAAAAAAAGAAATCAAATTATT
4      GGAGAAAAAAGAAAATGGATTTCATATTCATATACATGAGAATTATATAGAAACAGGAATAATCGTGGATTACTTTTTGCAAAAAAAGAAATCAAATTATT

1      TGGAAAAATAAGACTGTTCCAATTCCAATACTCGTGAAGAAAAAACCGTAATAAATGCAAAGAAGAAGTATCTTTGACCCAGTAGCGAAGAGTTTGAAGT 700
2      TGGAAAAATAAGACTGTTCCAATTCCAATACTCGTGAAGAAAAAACCGTAATAAATGCAAAGAAGAAGTATCTTTGACCCAGTAGCGAAGAGTTTGAAGT
3      TGGAAAAATAAGACTGTTCCAATTCCAATACTCGTGAAGAAAAAACCGTAATAAATGCAAAGAAGAAGTATCTTTGACCCAGTAGCGAAGAGTTTGAAGT
4      TGGAAAAATAAGACTGTTCCAATTCCAATACTCGTGAAGAAAAAACCGTAATAAATGCAAAGAAGAAGTATCTTTGACCCAGTAGCGAAGAGTTTGAAGT

```