Genetic diversity and population structure of ridge gourd (*Luffa acutangula*) accessions in a Tropical Asian collection using DArTseq based SNP markers

Grimar Abdiel Perez[1], Pumipat Tongyoo[2,3], Julapark Chunwongse[2,3,4], Hans de Jong[2,3,5], Paweena Chuenwarin[2,3,4*]

[1]Agricultural Science Program, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand;

[2]Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand;

[3]Center of Excellence on Agricultural Biotechnology: (AG-BIO/PERDO-CHE), Bangkok, Thailand;

[4]Department of Horticulture, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand

[5]Laboratory of Genetics, Wageningen University and Research Centre (WUR), Wageningen, the Netherlands

* Corresponding author

E-mail: paweena.chue@ku.th (PC)

26

Abstract

This study explored a germplasm consisting of 112 *Luffa acutangula* (ridge gourd) accessions originating from Thailand, Vietnam, China, Philippines, Indonesia, USA, Bangladesh and Laos for an analysis of the population structure and underlying genetic diversity using 2,834 DArTseq based SNPs. STRUCTURE analysis (ΔK at K=6) allowed us to group the accessions into six populations that corresponded well with the unrooted phylogenetic tree and Principle coordinate analyses. The phylogenetic tree differentiated the accessions into groups according to their passport geographic origin. STRUCTURE also revealed collection locations of the accessions for each cluster, demonstrating that genotypes of the ridge gourds were admixed at places nearby their regions of origin. AMOVA based on STRUCTURE clustering showed a higher variation within a population (87.17%) than AMOVA groups based on country of origin (84.94%). We also observed significant results in which a higher degree of genetic differentiation between populations according to the Phi statistics based on STRUCTURE clustering (0.128) rather than by country of germplasm collection (0.079). The clusters generated by STRUCTURE assigned characteristic fruit shape and length to each cluster grouping. The ample genetic diversity found in the *L. acutangula* germplasm can be utilized in ridge gourd breeding programs to help meet the demands of consumers (reduced bitterness) and aid farmers through breeding lines resistant to plant viruses.

Introduction

Luffa acutangula, commonly known as ridge gourd, angled loofah, or Chinese okra, is a domesticated vegetable of the Cucurbitaceae that originated from India [1-4]. Immature fruits are used as a vegetable which can be cooked or fried. In South-East Asia, the sweet juiciness, spongy texture, and mild-bitter taste are favorite characteristics [2]. In contrast, its use in traditional medicine is especially prevalent in Asia and middle America [5, 6] and India for its treatment of jaundice and urinary bladder stones [7, 8]. Extensive regional and cultural selection of specific cultivars have contributed to a wide variety of uses and value necessitating germplasm conservation to capture its enormous genetic variation. Most of the genetic diversity was established by describing the morphological variation of the accessions [9]. Still, such a limited approach is subject to environmental conditions as well as the developmental stages of the plant [10]. Instead, molecular markers provide more stable sources of information for carrying out genetic diversity studies.

Genetic diversity can be described as the measure of variation within a population. This variability is essential for future breeding schemes and represents the equilibrium between mutation and the loss of genetic variation [11]. The importance of population structure extends far beyond pure relatedness, in which plants with limited distribution have fewer habitats and smaller dense populations, which will lead to increased inbreeding and more structured populations [12]. If this structure is not correctly defined, it can cause loss of power in subsequent interpretations of the population under study. Another way of assessing genetic diversity which also serves to complement the results from population structure is by the use of phylogenetic modelling. This modelling helps to give an overview of the relative relatedness of each member of the population under study as well as ecological relatedness

[13]. It is always important to consider if results using a particular method reflects similar findings in other methods of genetic diversity assessment.

With the fast-paced innovations in molecular biology, several marker techniques have been developed for acquiring accurate and reliable information on population structure and genetic diversity of germplasms [14]. Molecular markers used for population structure and genetic diversity studies in ridge gourd include Inter Simple Sequence Repeat (ISSR) [15], Random Amplified Polymorphism Detection (RAPD) in *L. acutangula* [16], Simple Sequence Repeats (SSRs) [17], Inter Simple Sequence Repeats (ISSR), and Directed Amplification of Minisatellite DNA (DAMD) [18]. These techniques can be labor-intensive, costly and produce a low number of markers.

The present study aims to identify the population structure and genetic diversity of a *L. acutangula* germplasm in Thailand by using DArTseq based SNPs (https://www.diversityarrays.com). The study will provide essential information for genetic improvement programs in Thailand for meeting the demands of a diverse cultural cuisine in Thailand.

Materials and methods

Plant materials

We used in this study a *L. acutangula* germplasm comprising of 112 accessions from Thailand (91), Vietnam (4), Philippines (2), Indonesia (1), China (3), USA (1), Laos (3) and Bangladesh (7), conserved by the Tropical Vegetable Research Center (TVRC), Kasetsart University, Kamphaeng Saen Campus, and the World Vegetable Center, Taiwan.

97 ## Fruit trait evaluation

98 All *L. acutangula* accessions were evaluated for fruit traits such as fruit length by measuring
99 from stem-end to blossom-end and fruit shape adapted from the descriptors for sponge gourd
100 [19]. Accessions were grown under field conditions from August to December 2019 in which
101 five plants per plot were planted in a single bed with a 0.5 m space between the plants and 2
102 m spacing between beds. Fruits were harvested when they were young for consumption as
103 vegetable. Fruit shape was classified into either elongated slim, elliptical tapered, elongated
104 tapered, elongated elliptical tapered, tapered oblong or short elliptical tapered.
105

106 ## DNA extraction

107 Genomic DNA samples were extracted from 100 mg of pooled young leaves tissue of 2-
108 weeks-old seedlings from 20 plants per accession using a modified CTAB
109 (cetyltrimethylammonium bromide) method [20]. Precipitated DNA was resuspended in TE
110 buffer (10 mM Tris-HCl; 1 mM EDTA, pH 8.0) containing 2 µg/mL RNase. DNA
111 quality was evaluated by electrophoresis on a 1% agarose gel and was quantified with a
112 NanoDrop 2000c spectrophotometer V 1.6.0. The DNA concentration was adjusted to 50
113 ng/µL for DArTseq GBS analysis.
114

115 ## Genotyping of accessions of *L. acutangula* using DArTseq

116 The genomic DNA samples were sent to Diversity Arrays Technology Pty. Ltd., Canberra,
117 Australia, for DArTseq genotype-based sequencing [21]. To this end, DNA was digested
118 using *Pst*I-*Mse*I restriction enzymes as described by Kilian [22]. The digested fragments were
119 then ligated to adapters and amplified by PCR [23], followed by sequencing on Illumina
120 Hiseq2000. The single read sequencing was run for 77 cycles, and sequences generated were

121 handled by DArT analytical pipelines (Diversity Arrays Technology, Australia). In the
122 primary pipeline, poor-quality sequences were filtered from the FASTQ files by applying
123 rigorous selection criteria to the barcode region [24]. Identified sequences per
124 barcode/sample were used for marker calling. These files were then used in the secondary
125 pipeline for DArT P/L's proprietary SNP calling algorithms (DArTsoftseq).
126

127 **Population structure and data analysis**

128 DArTseq based SNPs were filtered using a call rate of 80% with a co-dominant marker
129 polymorphism information content (PIC) greater than 0.125. After filtering, 2,834 SNPs were
130 used for data analysis. The population structure of the 112 *L. acutangula* accessions was
131 determined using STRUCTURE version 2.3.4 [25]. Ten repeats were performed for each
132 number of hypothetical subpopulations (*K*) which were set from 1 to 10. The parameters used
133 consisted of an admixture model, a burning period of 50,000 steps, and 100,000 Markov
134 Chain Monte Carlo (MCMC). The STRUCTURE results were further analyzed using the R
135 package, POPHELPER version 2.3.0 [26]. The optimum number of *K* was calculated using the
136 Evanno method [27].
137 We constructed the phylogenetic trees with the weighted neighbor-joining method [28]
138 and visualized the data with DARWIN software version 6.0.021 [29]. Principal Coordinates
139 Analysis (PCoA), expected heterozygosity (H_E), observed heterozygosity (H_O), and pairwise
140 F_{ST} were calculated using ADEGENET version 2.1.1 [30] in the R statistical environment.
141 Analysis of Molecular Variance (AMOVA) and Shannon-Weiner Diversity index were
142 calculated in POPPR version 2.8.3 in R [31]. Jaccard distance was calculated in R software by
143 using ADE4 version 1.7-13 package [32]. One-way ANOVA test, box plots and Tukey HSD
144 were carried out using standard R statistics.
145

146 # Results

147 ## Population structure analysis

148 STRUCTURE analysis of the 2,834 SNPs revealed the highest value of *ΔK* at *K*=6 (Fig 1).
149 This value indicated a total of six informative subpopulations found across all *L. acutangula*
150 accessions. Each subpopulation was expressed as Cluster 1-6, representing 7%, 55%, 13%,
151 8%, 4% and 13% of the total number of accessions, respectively. Cluster 1 consisted of seven
152 accessions from Thailand and one from China (Fig 2). All 62 accessions in Cluster 2 were
153 from Thailand. Cluster 3 was composed of eleven accessions from Thailand and three from
154 Laos. Cluster 4 was made of two, four, two and one accession from Thailand, Vietnam,
155 Philippines and Indonesia, respectively. Cluster 5 consisted of two accessions from Thailand,
156 two from China and one from the USA. Cluster 6 contained seven accessions from Thailand
157 and seven from Bangladesh.

158 **Fig 1. Population structure of 112 *Luffa acutangula* accessions using DArTseq-based**
159 **SNP markers inferred by STRUCTURE program.** a) Number of subpopulations indicated
160 by the highest *ΔK*, b) Proportion of clustering of individuals to six subpopulations.

161 **Fig 2. Weighted neighbor-joining dendrogram of 112 *Luffa acutangula* accessions based**
162 **on DArTseq SNPs.** Branch colors correspond to clustering by STRUCTURE analysis.
163 Symbols indicate accession's country of origin.

164 The unrooted phylogenetic tree was able to differentiate the 112 *L. acutangula* accessions
165 into six clades that were consistent with the *ΔK* at *K*=6 (Fig 3). We further confirmed the
166 distinction of the six groups with the Principal Coordinate Analysis (PCoA) in which the
167 accessions of cluster 1 overlapped with those of cluster 2 (Fig 4). Axes 1 and 2 combined
168 explained 15.3% of the total variance. Cluster 1 and 2 had some accessions falling in two or

169 three quadrants suggesting a wide diversity in these two clusters, which were mostly
170 composed of Thailand accessions.
171 **Fig 3. Phylogenetic tree among the 112 *Luffa acutangula* accessions.** The colors of
172 branches illustrate accessions belonging to different clusters acquired from STRUCTURE
173 analysis. Six clades were identified as indicated by circle colors.
174 **Fig 4. Principal coordinate analysis (PCoA) of 112 *Luffa acutangula* accessions using**
175 **DArTseq-based SNPs.**
176

177 Genetic diversity

178 Heterozygosity can be measured by expected heterozygosity (H_E) and observed
179 heterozygosity (H_O). H_E gives information about the probability of an individual's portion of
180 heterozygosity for all analyzed loci, while H_O is the portion of heterozygous genes in the
181 analyzed population [33]. The expected heterozygosity (H_E) as measure of genetic diversity
182 ranged in the six clusters from 0.338 (Cluster 2) to 0.272 (Cluster 3); the remaining cluster
183 had values in between: Cluster 1 (0.274), Cluster 4 (0.295), Cluster 5 (0.287) and cluster 6
184 (0.334). The high H_E in Cluster 2 also corresponded with the highest Shannon & Weiner
185 diversity index (H) of 4.127. The observed heterozygosity (H_O)for the STRUCTURE
186 grouping was as follows: Cluster 1 (0.378), Cluster 2 (0.388), Cluster 3 (0.191), Cluster 4
187 (0.289), Cluster 5 (0.363) and Cluster 6 (0.410). When the observed heterozygosity was
188 found higher than the expected heterozygosity, a random mating system governs that
189 particular population [33]. This $H_O > H_E$ was observed in Cluster 1, Cluster 2, Cluster 5 and
190 Cluster 6 (Table 1). The observed heterozygosity (H_O) in Cluster 4 (0.289) was almost
191 similar in value to the H_E (0.295), which might indicate that crossing in the population was
192 almost unplanned [33]. Cluster 3 had the lowest H_E (0.272) and H_O (0.191) from all six
193 subpopulations, and given that H_O was lower than H_E indicated that Cluster 3 was potentially

194 an inbred population [33]. The mean SNP polymorphism information content (PIC) had a
195 range value from 0.125 to 0.375, with a mean of 0.288 (Table 1).

196 **Table 1. Genetic diversity among 112 *Luffa acutangula* accessions based on**
197 **STRUCTURE analysis.**

	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**	**Total**
N	8	62	14	9	5	14	112
H	2.079	4.127	2.639	2.197	1.609	2.639	4.718
H_O	0.378	0.388	0.191	0.289	0.363	0.410	0.357
H_E	0.274	0.338	0.272	0.295	0.287	0.334	0.359
PIC		Min_PIC=0.125		Max_PIC=0.375		Mean_PIC=0.288	

198 N, number of accessions; H, Shannon & Weiner diversity index; H_O, observed
199 heterozygosity; H_E, expected heterozygosity; PIC, polymorphism information content.

200

201 Pairwise genetic differentiation was highest between Cluster 3 and Cluster 4 (F_{ST}=0.262),
202 while the lowest was observed between Cluster 1 and Cluster 2 (F_{ST} = 0.089), and also
203 between Cluster 2 and Cluster 3 (F_{ST} = 0.086) (**Error! Reference source not found.**).
204 AMOVA based on STRUCTURE clustering showed a higher variation within a population
205 (87.17%) than between populations (12.83%). However, AMOVA on accessions grouped by
206 country of origin showed a slightly lower within a population variation (84.94%) and a
207 somewhat higher between population variation (15.06%). Phi statistics between
208 STRUCTURE clusters (0.128) indicate a higher degree of differentiation in comparison to
209 groups generated by country of origin (0.079) (Table 3).

210 **Table 2. Pairwise F_{ST} (genetic differentiation) values among clusters identified by**
211 **STRUCTURE analysis.**

STRUCTURE Clustering	Cluster 2	Cluster 1	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 2						
Cluster 1	0.089					
Cluster 3	0.086	0.180				
Cluster 4	0.164	0.230	0.262			
Cluster 5	0.147	0.203	0.202	0.210		
Cluster 6	0.123	0.186	0.127	0.193	0.174	

Table 3. Analysis of molecular variance (AMOVA) of 112 *Luffa acutangula* accessions.

Grouping of accessions	Sample	Df	Mean Sq	Variance	%	Phi statistics (ϕ)	P-value
Structure clustering	Between populations	5	4312.218	123.098	12.83	0.128	0.001
	Between individuals within populations	106	734.459	-101.542	-10.59		
	Within individuals	112	937.543	937.543	97.75		
	Total	223	916.675	959.099	100.00		
By country	Between populations	7	2425.899	153.389	15.06	0.079	0.001
	Between individuals within populations	104	792.620	-72.461	-7.11		
	Within individuals	112	937.543	937.543	92.05		
	Total	223	916.675	1018.470	100.00		

Df, degrees of freedom; Sq, square.

Geographic distribution and genetic diversity

Cluster 1 and 2 represented the genotype characteristics found in Thailand, while Cluster 3 consisted of characteristics belonging to Laos as is clear from the STRUCTURE analysis (Fig 5). Accessions from Vietnam, Philippines and Indonesia were closely related with the majority of the shared genotypic characteristics for Vietnam, Cluster 4. Defining characteristics for Cluster 5 from the USA and 6 were originally from the accessions from Bangladesh and Thailand. The *L. acutangula* accessions from Thailand were much more diverse when compared to accessions from other countries (Fig 6). In contrast, the accessions from Laos and Vietnam were highly uniform and had low genotype proportions from accessions belonging to other countries. Both the Philippines and Indonesia had genotypic proportions belonging to accessions from Vietnam and Bangladesh. Still, they differed for the Philippines which contained genotypes from the USA, while Indonesia had from Thailand. This observation may also correlate to the high composition of accessions belonging to Thailand.

Fig 5. Genetic diversity proportion of 112 *Luffa acutangula* accessions based on country of origin. Branch colors indicate country of origin. Side bars correspond to proportion clustering from Fig 1.

Fig 6. Distribution of *Luffa acutangula* accessions and their corresponding countries. Colors are based on STRUCTURE analysis bars from Fig 1.

Association of fruit trait and STRUCTURE clustering

The *L. acutangula* germplasm of 112 accessions was diverse for fruit shape and fruit length. Six main types of fruit shapes were used as typical distinguishable phenotypes: elongated

238 slim, elliptical tapered, elongated tapered, elongated elliptical tapered, tapered oblong, and
239 short elliptical conical (Fig 7). One-way ANOVA test of fruit length showed highly
240 significant differences between STRUCTURE clusters (p-value = 2.91e-09) (S2 Appendix).
241 Tukey HSD p-values after adjustment for the multiple comparisons for fruit length showed
242 significant differences for Cluster 1 and Cluster 2, Cluster 1 and Cluster 3, Cluster 1 and 6,
243 Cluster 2 and Cluster 6, and Cluster 4 and Cluster 6 (S3Appendix). Other cluster
244 combinations did not show significant differences in fruit length according to Tukey HSD.
245 **Fig 7. *Luffa acutangula* fruit shape**. Each cluster shape is represented by an accession with
246 the least or no admixture. (Source: Personal collection credit to Mr. Anucha Wongpraneekul
247 and Grimar A. Perez)
248

249 Discussion

250 Population structure and diversity in *L. acutangula*

251 In this study, we have been able to demonstrate a consistent six population clustering of the
252 112 *L. acutangula* accessions. Whole-genome DArTseq based SNPs, geographic area of
253 collection and fruit characteristics were used to study a total of 112 *L. acutangula* accessions.
254 Analysis using STRUCTURE (Fig 1), weighted neighbor-joining method (Fig 3) and PCoA
255 (Fig 4) showed a consistent representation of a total of 6 subpopulations. This grouping
256 consistency was also observed at the level of individual accessions in both STRUCTURE
257 analysis as well as a phylogenetic dendrogram (Fig 2). The subpopulations Cluster 1- 6
258 showed distinct genotypic characteristics that can be put to use in genetic improvement
259 programs.
260 Geographically, accessions from Thailand made up the majority of Cluster 1 and Cluster 2.
261 Similar correlations between the population structure and geographic origin were observed in

262 *Capsicum* germplasm [34]. Cluster 1 members were collected from the central plain of
263 Thailand while cluster 2 accessions (n = 62 accessions) were scattered all-over in Thailand.
264 Cluster 3 was composed of accessions from northern Thailand and Laos. This grouping
265 observed in Cluster 3 much emphasizes the similarity in cuisine between the north part of
266 Thailand and Laos. The genotypic characteristics found in the Vietnam accessions (Cluster 4)
267 were found in accessions collected from Southern Thailand. Besides, the genetic proportion
268 of Vietnam accessions found in Cluster 4 was more closely shared with accessions from the
269 Philippines and Indonesia, suggesting a relationship between people from these countries
270 which shared an ancestral history or the possibility that there were immigrations among
271 mainland (Vietnam) and islands in southeast Asia (the Philippines and Indonesia) and also
272 due to the exchange of goods between these countries in the past [35, 36]. The observed
273 distribution of cluster 4 accessions to the southern part of Thailand point at spreading of
274 seeds along historic commerce ship routes from the Southern Thailand regions to countries
275 such as India and China [36, 37]. The primary representative of Cluster 5 genetic was an
276 accession from the USA, which had an almost unmixed nature based on STRUCTURE
277 analysis. Two accessions from China and two accessions from Thailand shared genotypic
278 proportions to this accession in Cluster 5. Cluster 6 was made up of seven accessions from
279 Thailand and seven from Bangladesh, which together was separated into two groups in the
280 dendrogram. Cluster 6 genotype characteristic might have been introduced from Nepal
281 region via Bangladesh [38], but lack of samples from Myanmar, makes it difficult to confirm
282 the trafficking of this material [36, 39].
283 Cluster 2 and Cluster 6 are two potential groups for breeding programs due to their high
284 diversity when compared across all six subpopulations. Both Cluster 2 and Cluster 6 had high
285 H_O (4.127 and 2.639, respectively) and H_E (0.338 and 0.334, respectively) which indicates a
286 high degree of diversity and substantial heterozygosity. Overall, Cluster 5 displayed the

287 lowest diversity and heterozygosity despite being made up of two accessions from Thailand,
288 two from China and one from the USA.
289 In a closer look at the genetics of ridge gourd in Thailand such as cluster 1 accessions, we
290 noted that the spreading of their specific SNPs occurred in accessions collected nearby the
291 area where the pure cluster 1 representative accessions had been collected (Fig 6). The same
292 was true for the genetic of cluster 3, which was represented by accessions from Laos. Still,
293 the other accessions in cluster 3 that were domesticated in Thailand were already mixed with
294 the genetics of cluster 2, which is the primary governing genotype of ridge gourd in Thailand.
295

296 Genetic differentiation among populations

297 Pairwise genetic differentiation (F_{ST}) was relatively low between Cluster 2 and Cluster 1
298 (0.089) which is consistent with the PCoA in which the accessions in cluster 1 were found
299 together with the accessions in cluster 2. Cluster 2 and Cluster 3 pairwise genetic
300 differentiation was low (0.086), which also is supported by the PCoA. These two clusters
301 were spread nearby each another, and in the STRUCTURE result, most of the admixed
302 genetics of accessions in Cluster 2 contains the genetics of Cluster 3. Similar observations
303 were seen between Cluster 2 and Cluster 5 (0.147), Cluster 2 and Cluster 6 (0.123), and
304 Cluster 3 and Cluster 6 (0.127) (**Error! Reference source not found.**). Genetic
305 differentiation of F_{ST} lower than 0.05 was defined as low, F_{ST} between 0.05 and 0.15 as
306 moderate, F_{ST} between 0.15 and 0.25 as high, and F_{ST} greater than 0.25 as very high [40, 41].
307 The low differentiation observed between Cluster 2 with the other four Clusters (Cluster 1, 3,
308 5 and 6) may reflect the Thailand only composition found in Cluster 2, as well as a
309 substantial diversity found within Cluster 2. Clusters with a pairwise genetic differentiation
310 between 0.150 to 0.210 (Cluster 2 and 4=0.164, Cluster 1 and 3=0.180, Cluster 1 and
311 6=0.186, Cluster 4 and 6=0.193, Cluster 5 and 6=0.174, Cluster 1 and 5=0.203, Cluster 3 and

5= 0.202, and Cluster 4 and 5=0.210) showed substantial differentiation between populations with minor admixture being exchanged across corresponding paired clusters as seen in the STRUCTURE analysis (Fig 1b) and PCoA. The highest pairwise genetic differentiation occurred between Cluster 1 and Cluster 4 (0.230), and Cluster 3 and Cluster 4 (0.262). These two pairs showed minimal to almost no admixture across each paired cluster as seen in the STRUCTURE analysis, and there was no overlap in the PCoA. Pairwise genetic differentiation agreed with AMOVA results (Table 3) which showed that the majority of variation (87.17%) occurred within populations while only 12.83% occurred between populations.

Country of germplasm collection and genetic diversity

The AMOVA based on grouping according to the country of origin, showed a similar trend to the AMOVA results of STRUCTURE. A higher within-population variation (84.94%) was observed when compared to between-population variance (15.06%). Despite this variance similarity between STRUCTURE and country grouping, between populations, Phi statistics based on STRUCTURE grouping displayed a higher degree of genetic differentiation (0.128) than populations generated by country of germplasm collection (0.079) (Table 3).

The within population genetic variation is visible in the phylogenetic tree with corresponding STRUCTURE clustering proportions. Accessions of highly pure genotypes that represent each subpopulation from STRUCTURE analysis are LA083 and LA046 (Cluster 1), LA032 and LA029 (Cluster 2), vio55943 and vio55749 (Cluster 3), LA087 (Cluster 4), vio46065 (Cluster 5), and LA080A (Cluster 6) (**Error! Reference source not found.**). The genetic proportions of these accessions can be observed throughout cluster 1 through cluster 6. This observation corresponds to the higher within population diversity found in the AMOVA analysis. The lower AMOVA genetic differentiation between

populations observed by country of germplasm collection may partly reflect a substantial similarity between countries in Southeast Asia as well as a high genetic exchange between countries. Besides, it might add a new factor to consider in population structure analysis than the usual geopolitical and biological borders [42].

Part of the South East Asian countries such as Thailand, Laos, Vietnam, Indonesia and Philippines are separated by various biological barriers and marked by geopolitical borders. Despite the proximity of Thailand to Laos and Vietnam, Thailand's genetic diversity includes a low proportion of genetic characteristics from Vietnam, when compared to the larger proportions from Laos (Fig 6). This observation that may be heavily influenced by cultural similarities when compared to biological barriers or geopolitical borders can also be seen in accessions from Indonesia which to a great extent contain genetic proportions from Vietnam, but none from Laos. However, the Philippines and Vietnam share a similar genetic characteristic in that both of their accessions contain proportions from Vietnam and Bangladesh.

Association of fruit traits and country of germplasm collection

Cluster 1 and cluster 4 primarily consisted of long fruits, and elongated slim and elliptical tapered fruit shapes, respectively (Fig 7, S1 Appendix). Cluster 1 and 4 accessions from Thailand, Vietnam, Indonesia, and the Philippines favor long fruits with a wide range of fruit shapes. Cluster 2 was by far the largest subpopulation that was primarily made up of accessions collected in Thailand. Cluster 2 was characterized by medium length fruits with an elliptical tapered fruit shape. However, accessions within Cluster 2 displayed a range of shapes that stemmed from the representative elliptical conical fruit shape. Accessions from Thailand and Laos were grouped in Cluster 3, had medium-length fruits with an elongated tapered fruit shape. This observation between Laos and Thailand represents the similarity in

culture, including food shared across both countries. Medium length fruits characterized by elongated elliptical tapered were found in Cluster 5, which consisted of accessions from Thailand, China and the USA. Peculiarly, the accession from the USA in this cluster was perhaps collected from Asia, the reason being that Luffa is native to Asia and not the Americas [1, 2]. Cluster 6 was equally composed of accessions from Thailand and Bangladesh. The fruits of Cluster 6 were short in length with a short elliptical tapered fruit shape. This observation in Cluster 6 may indicate a selection of accessions from Bangladesh for breeding purposes in Thailand, or it segregated after being cultivated for some time. The reason for this could be because the majority of accessions in Thailand are widespread in all six subpopulations with a medium to long fruit length. However, fruits from Bangladesh are the only fruits that have significant short fruit lengths found in one out of six subpopulations.

Conclusions

The study found significant genetic diversity within the six populations identified from the *L. acutangula* germplasm. Phenotypic information, such as the fruit length and fruit shape, positively followed suit in the variation observed across the six populations inferred by the genotypic data. At the same time, geographical provenance was reflected in the clustering analysis, which also showed a relationship to the phenotypic information. This ample diversity serves as a strong catalyst to utilize such germplasm for breeding programs which can be geared towards farmer and consumer preference.

Our future objective will be to identify accessions endowed with important agricultural traits such as reduced bitterness and resistance to plant viruses for farmers in Thailand and Asia. Such analysis will then be followed by association mapping to identify potential markers for these traits of interest.

Acknowledgements

Special thanks to the Tropical Vegetable Research Center and the World Vegetable Center for providing seeds, and to Mr. Anucha Wongpraneekul and Ms. Waraporn Sinsathapornpong for their help with field management of ridge gourd and photos of the fruits.

397 # References

398

399 1. Heiser CB, Schilling EE. Phylogeny and distribution of *Luffa* (Cucurbitaceae).
400 Biotropica. 1988:185-91.
401 2. Pessarakli M. Handbook of Cucurbits: growth, cultural practices, and physiology:
402 CRC Press; 2016.
403 3. Soladoye MO, Adebisi AA. *Luffa acutangula* (L.) Roxb. 2004. Available from:
404 http://www.prota4u.org/search.asp.
405 4. Filipowicz N, Schaefer H, Renner SS. Revisiting *Luffa* (Cucurbitaceae) 25 years after
406 C. Heiser: species boundaries and application of names tested with plastid and nuclear DNA
407 sequences. Systematic Botany. 2014;39(1):205-15.
408 5. Sastri B. The wealth of India, vol. 6. Council of Scientific and Industrial Research,
409 New Delhi. 1962:382-3.
410 6. Morton JF. Atlas of medicinal plants of middle America: Bahamas to
411 Yucatan.[Volume 1]: Thomas; 1981.
412 7. Samvatsar S, Diwanji V. Plant sources for the treatment of jaundice in the tribals of
413 Western Madhya Pradesh of India. Journal of Ethnopharmacology. 2000;73(1-2):313-6.
414 8. Katewa S, Chaudhary B, Jain A. Folk herbal medicines from tribal area of Rajasthan,
415 India. Journal of Ethnopharmacology. 2004;92(1):41-6.
416 9. Prakash K, Pandey A, Radhamani J, Bisht I. Morphological variability in cultivated
417 and wild species of *Luffa* (Cucurbitaceae) from India. Genetic resources and crop evolution.
418 2013;60(8):2319-29.

419 10. Zhang S, Hu J, Xu S. Developmental genetic analysis of fruit shape traits under
420 different environmental conditions in sponge gourd (*Luffa cylindrical* (L) Roem. Violales,
421 Cucurbitaceae). Genetics and Molecular Biology. 2008;31(3):708-10.
422 11. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, et al.
423 Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol.
424 2012;10(9):e1001388.
425 12. Holsinger DAFKE. Genetics and conservation of rare plants: Oxford University Press
426 on Demand; 1991.
427 13. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends
428 in genetics. 2008;24(3):133-41.
429 14. Bretting P, Widrlechner MP. Genetic markers and plant genetic resource management.
430 1995.
431 15. Prakash K, Pati K, Arya L, Pandey A, Verma M, Prakash K, et al. Population
432 structure and diversity in cultivated and wild *Luffa* species. Biochemical Systematics and
433 Ecology. 2014;(56):165-70.
434 16. Hoque S, Rabbani M. Assessment of genetic relationship among landraces of
435 Bangladeshi ridge gourd (*Luffa acutangula* Roxb.) using RAPD markers. Journal of
436 Scientific Research. 2009;1(3):615-23.
437 17. Pandey S, Ansari W, Choudhary B, Pandey M, Jena S, Singh A, et al. Microsatellite
438 analysis of genetic diversity and population structure of hermaphrodite ridge gourd (*Luffa*
439 *hermaphrodita*). 3 Biotech. 2018;8(1):17.
440 18. Misra S, Srivastava AK, Verma S, Pandey S, Bargali SS, Rana TS, et al. Phenetic and
441 genetic diversity in Indian *Luffa* (Cucurbitaceae) inferred from morphometric, ISSR and
442 DAMD markers. Genetic resources and crop evolution. 2017;64(5):995-1010.

443 19. Joshi BK, KC HB, Tiwari RK, Ghale M, Sthapit BR. Descriptors for sponge gourd
444 (*Luffa cylindrica* (L.) Roem.). 2004.
445 20. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh
446 leaf tissue. Phytochemical Bulletin. 1987;19:11-5.
447 21. Von Mark VC, Kilian A, Dierig DA. Development of DArT marker platforms and
448 genetic diversity assessment of the US collection of the new oilseed crop lesquerella and
449 related species. PLoS One. 2013;8(5):e64062.
450 22. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al. Diversity arrays
451 technology: a generic genome profiling technology on open platforms. Data production and
452 analysis in population genomics: Springer; 2012. p. 67-89.
453 23. Raman H, Raman R, Kilian A, Detering F, Carling J, Coombes N, et al. Genome-wide
454 delineation of natural variation for pod shatter resistance in *Brassica napus*. PLoS One.
455 2014;9(7):e101673.
456 24. Barilli E, Cobos MJ, Carrillo E, Kilian A, Carling J, Rubiales D. A high-density
457 integrated DArTseq SNP-based genetic map of *Pisum fulvum* and identification of QTLs
458 controlling rust resistance. Frontiers in Plant Science. 2018;9:167.
459 25. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using
460 multilocus genotype data. Genetics. 2000;155(2):945-59.
461 26. Francis RM. pophelper: an R package and web app to analyse and visualize
462 population structure. Molecular ecology resources. 2017;17(1):27-32.
463 27. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using
464 the software STRUCTURE: a simulation study. Molecular ecology. 2005;14(8):2611-20.
465 28. Bruno WJ, Socci ND, Halpern AL. Weighted neighbor joining: a likelihood-based
466 approach to distance-based phylogeny reconstruction. Molecular biology and evolution.
467 2000;17(1):189-97.

29. Perrier X, Jacquemoud-Collet JP. DARwin software. 6.0.021 ed2006.

30. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27(21):3070-1.

31. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ. 2014;2:e281.

32. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. Journal of statistical software. 2007;22(4):1-20.

33. Chesnokov YV, Artemyeva A. Evaluation of the measure of polymorphism information of genetic diversity. Сельскохозяйственная биология. 2015;(5 (eng)).

34. Lee H-Y, Ro N-Y, Jeong HJ, Kwon JK, Jo J, Ha Y, et al. Genetic diversity and population structure analysis to construct a core collection from a large *Capsicum* germplasm. BMC genetics. 2016;17(1):142.

35. Tumonggor MK, Karafet TM, Hallmark B, Lansing JS, Sudoyo H, Hammer MF, et al. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. Journal of human genetics. 2013;58(3):165-73.

36. Bellina Pryce B, Silapanth P. Weaving cultural identities on trans-Asiatic networks: Upper Thai-Malay Peninsula–an early socio-political landscape. Bulletin de l'École française d'Extrême-Orient. 2006:257-93.

37. Castillo C. Rice in Thailand: the archaeobotanical contribution. Rice. 2011;4(3):114.

38. Marr KL, Bhattarai NK, Xia Y-M. Allozymic, morphological, and phenological diversity in cultivated *Luffa acutangula* (Cucurbitaceae) from China, Laos, and Nepal, and Allozyme Divergence between *L. acutangula* and *L. aegyptiaca*. Economic Botany. 2005;59(2):154-65.

39. Castillo C, Fuller DQ. Still too fragmentary and dependent upon chance? Advances in the study of early Southeast Asian archaeobotany. Fifty years of archaeology in Southeast Asia: Essays in honour of Ian Glover. 2010:91-111.

40. Frankham R, Briscoe DA, Ballou JD. Introduction to conservation genetics: Cambridge university press; 2002.

41. Xu Y, Mai JW, Yu BI, Hu HX, Yuan L, Jashenko R, et al. Study on the genetic differentiation of geographic populations of *Calliptamus italicus* (Orthoptera: Acrididae) in sino-kazakh border areas based on mitochondrial COI and COII genes. Journal of economic entomology. 2019;112(4):1912-9.

42. Meirmans PG. Seven common mistakes in population genetics and how to avoid them. Molecular ecology. 2015;24(13):3223-31.

Supporting information

S1 Appendix. Box plots of fruit length (cm) by STRUCTURE clustering.

S2 Appendix. One-way ANOVA based on STRUCTURE grouping. ***P-value < 0.001.

S3 Appendix. Tukey HSD based on STRUCTURE grouping with a 95% family-wise confidence level. Diff, mean difference between two groups; Lwr, lower endpoint of the interval; Upr, upper endpoint of the interval; P adj, p-value after adjustment for the multiple comparisons.

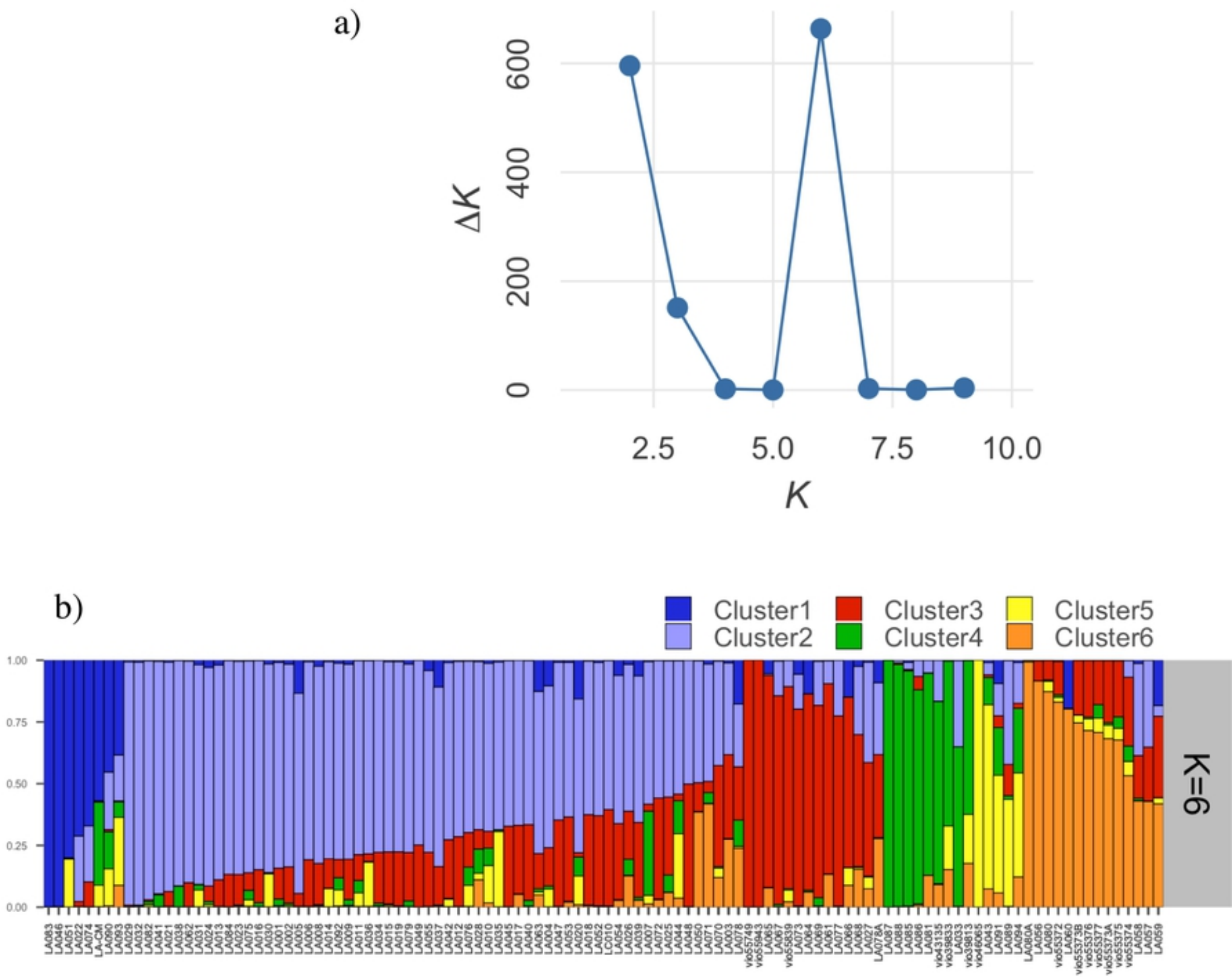
a)
ΔK
0
200
400
600
2.5
5.0
7.5
10.0
K
b)
Cluster1
Cluster2
Cluster3
Cluster4
Cluster5
Cluster6
1.00
0.75
0.50
0.25
0.00
K=6

Fig1

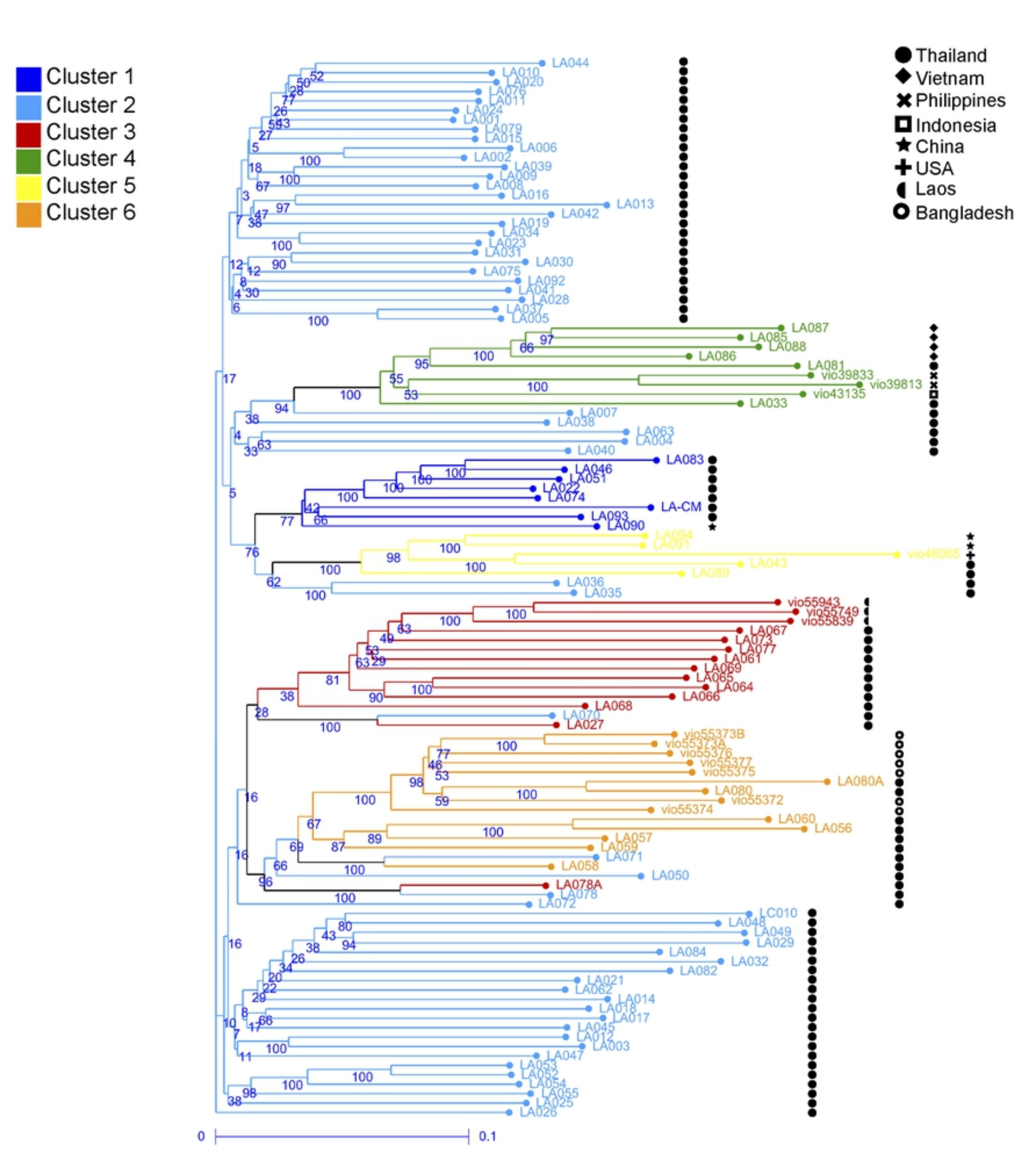
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Thailand
Vietnam
Philippines
Indonesia
China
USA
Laos
Bangladesh
LA044
LA010
LA020
LA076
LA011
LA024
LA001
LA079
LA015
LA006
LA002
LA039
LA009
LA008
LA016
LA013
LA042
LA019
LA034
LA023
LA031
LA030
LA075
LA092
LA041
LA028
LA037
LA005
LA087
LA085
LA088
LA086
LA081
vio39833
vio39813
vio43135
LA033
LA007
LA038
LA063
LA004
LA040
LA083
LA046
LA051
LA022
LA074
LA-CM
LA093
LA090
LA094
LA091
vio46065
LA043
LA089
LA036
LA035
vio55943
vio55749
vio55839
LA067
LA073
LA077
LA061
LA069
LA065
LA064
LA066
LA068
LA070
LA027
vio55373B
vio55373A
vio55376
vio55377
vio55375
LA080A
LA080
vio55372
vio55374
LA060
LA056
LA057
LA059
LA071
LA058
LA050
LA078A
LA078
LA072
LC010
LA048
LA049
LA029
LA084
LA032
LA082
LA021
LA062
LA014
LA018
LA017
LA045
LA012
LA003
LA047
LA053
LA052
LA054
LA055
LA025
LA026
0
0.1

Fig2

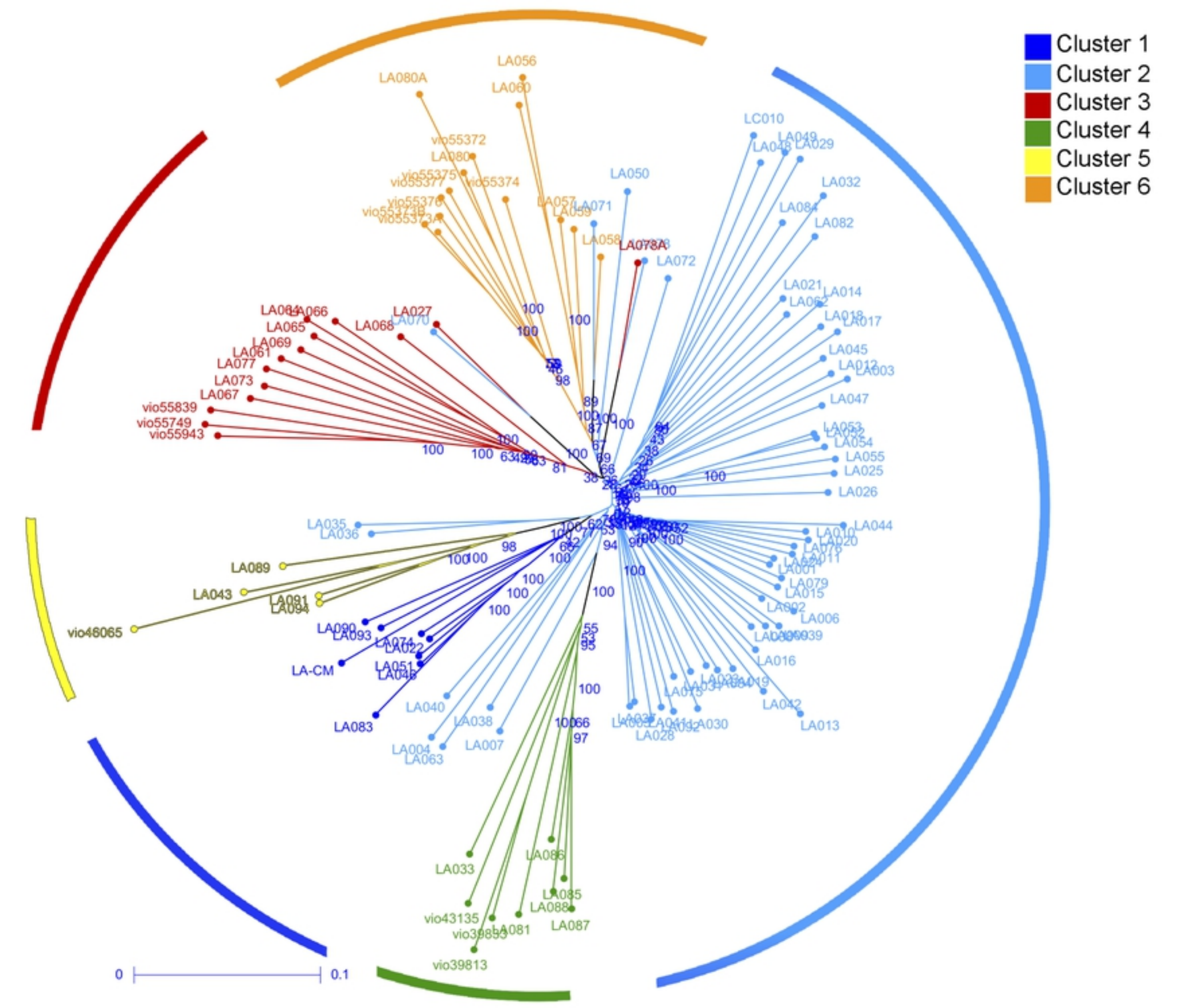
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
LA056
LA080A
LA060
vio55372
LA080
vio55375
vio55377
vio55374
vio55376
vio55373B
vio55373A
LA057
LA059
LA071
LA050
LA058
LA078A
LA072
LC010
LA049
LA048
LA029
LA032
LA084
LA082
LA021
LA014
LA062
LA018
LA017
LA045
LA012
LA003
LA047
LA053
LA052
LA054
LA055
LA025
LA026
LA044
LA010
LA020
LA076
LA011
LA024
LA001
LA079
LA015
LA002
LA006
LA039
LA016
LA019
LA042
LA013
LA030
LA028
LA064
LA066
LA065
LA069
LA061
LA077
LA073
LA067
vio55839
vio55749
vio55943
LA027
LA070
LA068
LA035
LA036
LA089
LA043
LA091
LA094
vio46065
LA090
LA093
LA074
LA022
LA051
LA046
LA-CM
LA083
LA040
LA038
LA007
LA004
LA063
LA033
LA086
LA085
LA088
LA087
vio43135
vio39853
LA081
vio39813
0
0.1

Fig3

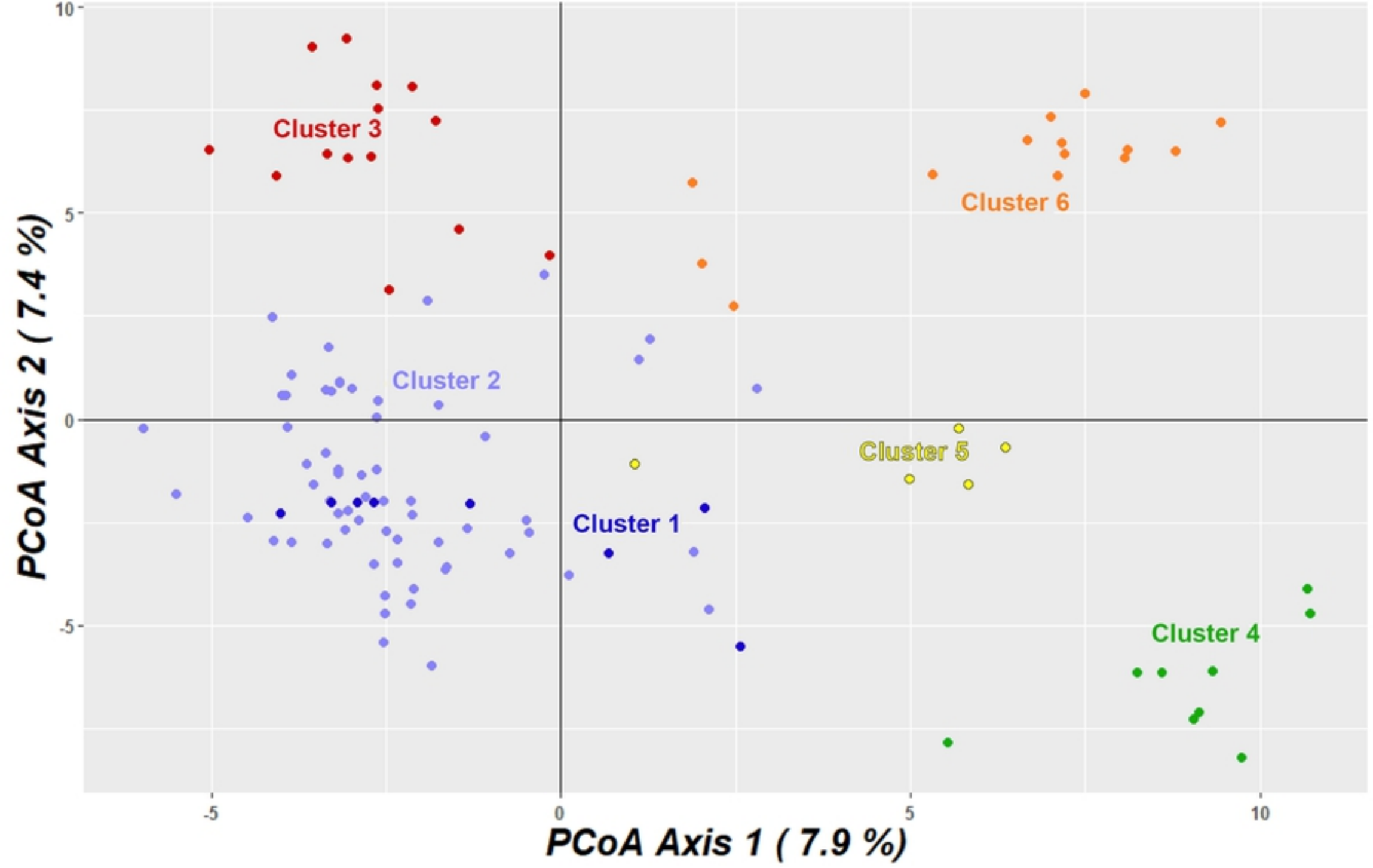
Cluster 3
Cluster 6
Cluster 2
Cluster 5
Cluster 1
Cluster 4
PCoA Axis 2 (7.4 %)
PCoA Axis 1 (7.9 %)
10
5
0
-5
-5
0
5
10

Fig4

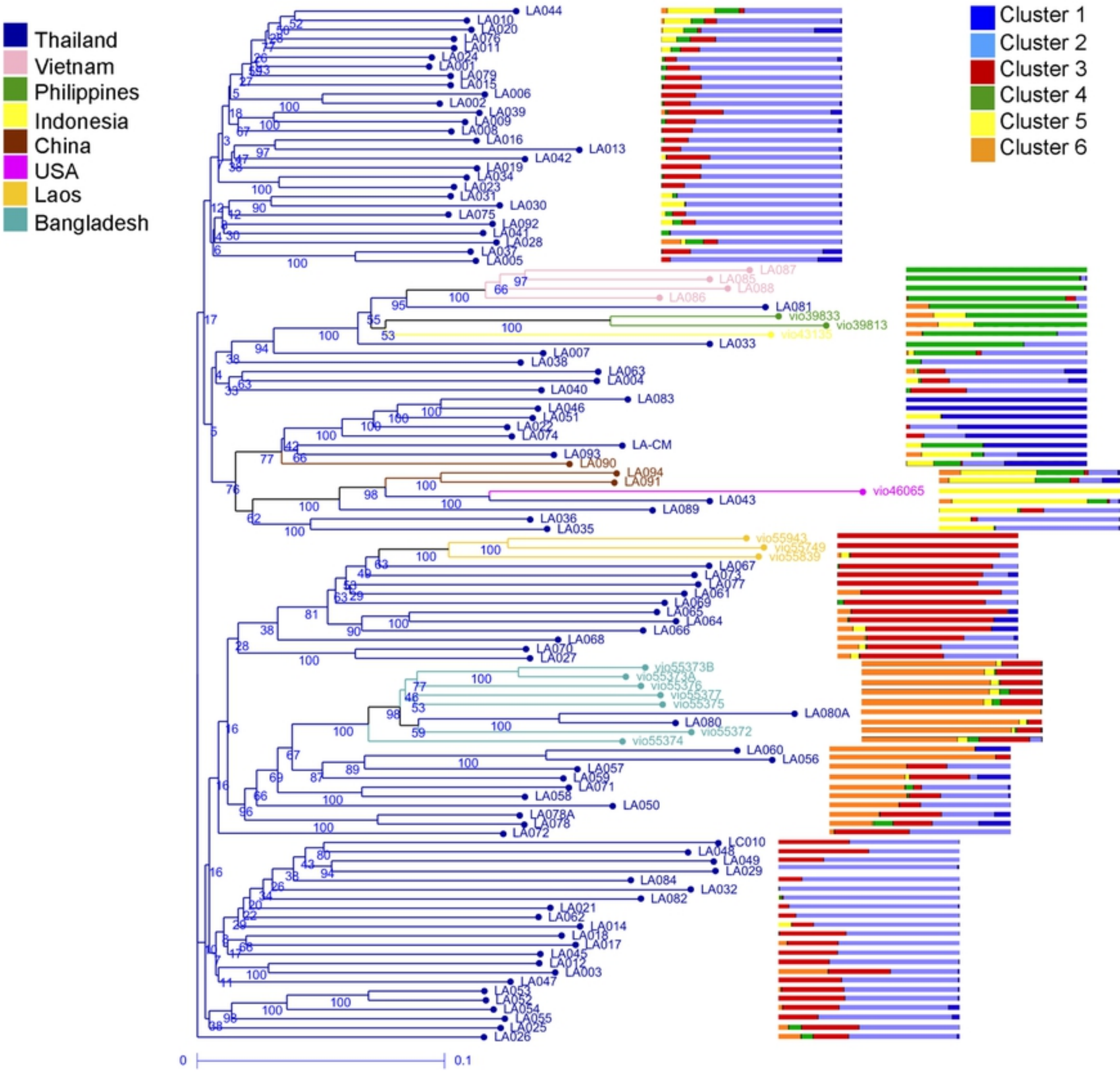
Thailand
Vietnam
Philippines
Indonesia
China
USA
Laos
Bangladesh
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
LA044
LA010
LA020
LA076
LA011
LA024
LA001
LA079
LA015
LA006
LA002
LA039
LA009
LA008
LA016
LA013
LA042
LA019
LA034
LA023
LA031
LA030
LA075
LA092
LA041
LA028
LA037
LA005
LA087
LA085
LA088
LA086
LA081
vio39833
vio39813
vio43135
LA033
LA007
LA038
LA063
LA004
LA040
LA083
LA046
LA051
LA022
LA074
LA-CM
LA093
LA090
LA094
LA091
vio46065
LA043
LA089
LA036
LA035
vio55943
vio55749
vio55839
LA067
LA073
LA077
LA061
LA069
LA065
LA064
LA066
LA068
LA070
LA027
vio55373B
vio55373A
vio55376
vio55377
vio55375
LA080A
LA080
vio55372
vio55374
LA060
LA056
LA057
LA059
LA071
LA058
LA050
LA078A
LA078
LA072
LC010
LA048
LA049
LA029
LA084
LA032
LA082
LA021
LA062
LA014
LA018
LA017
LA045
LA012
LA003
LA047
LA053
LA052
LA054
LA055
LA025
LA026
0
0.1

Fig5

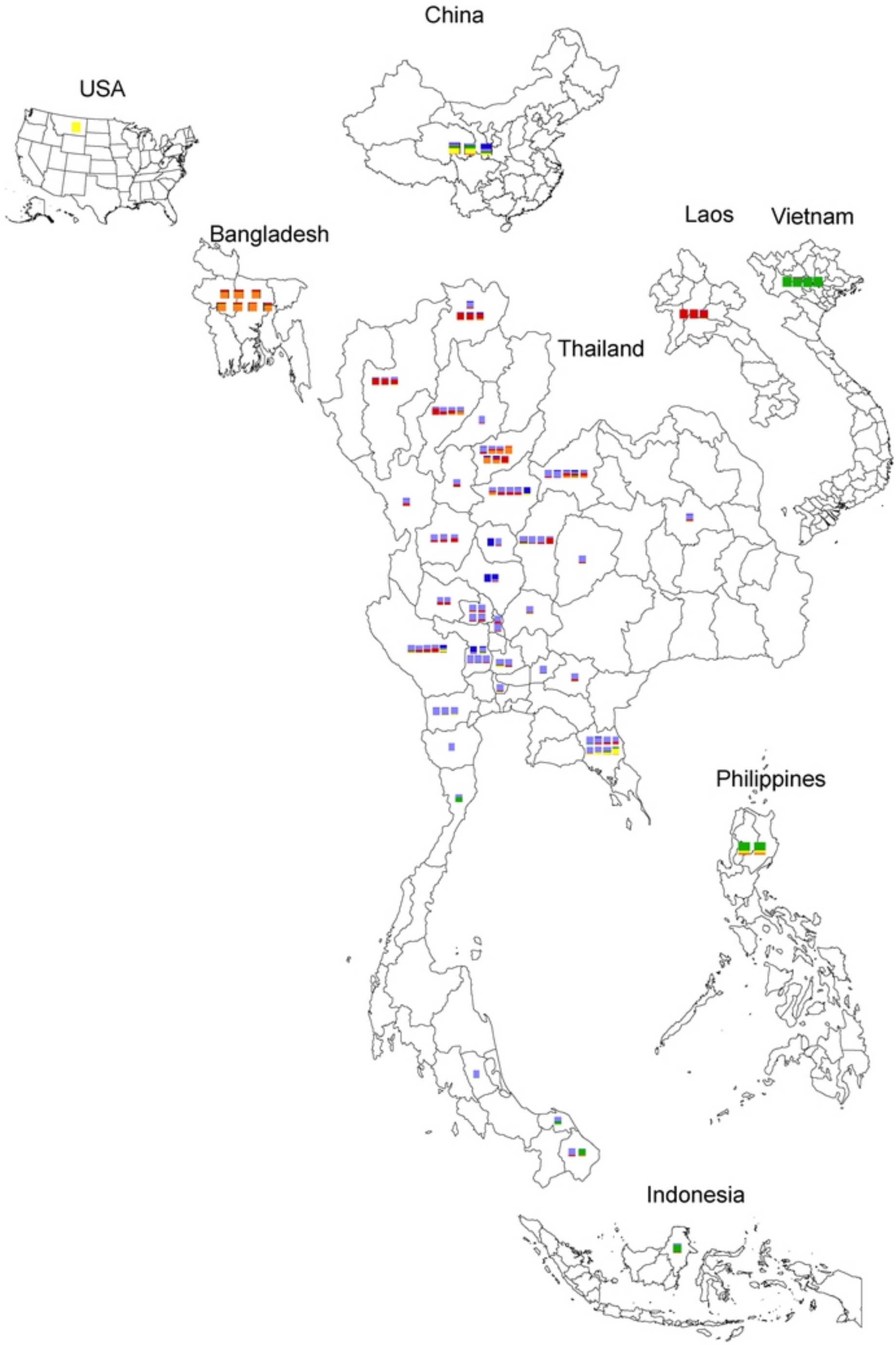
China
USA
Bangladesh
Laos
Vietnam
Thailand
Philippines
Indonesia

Fig6

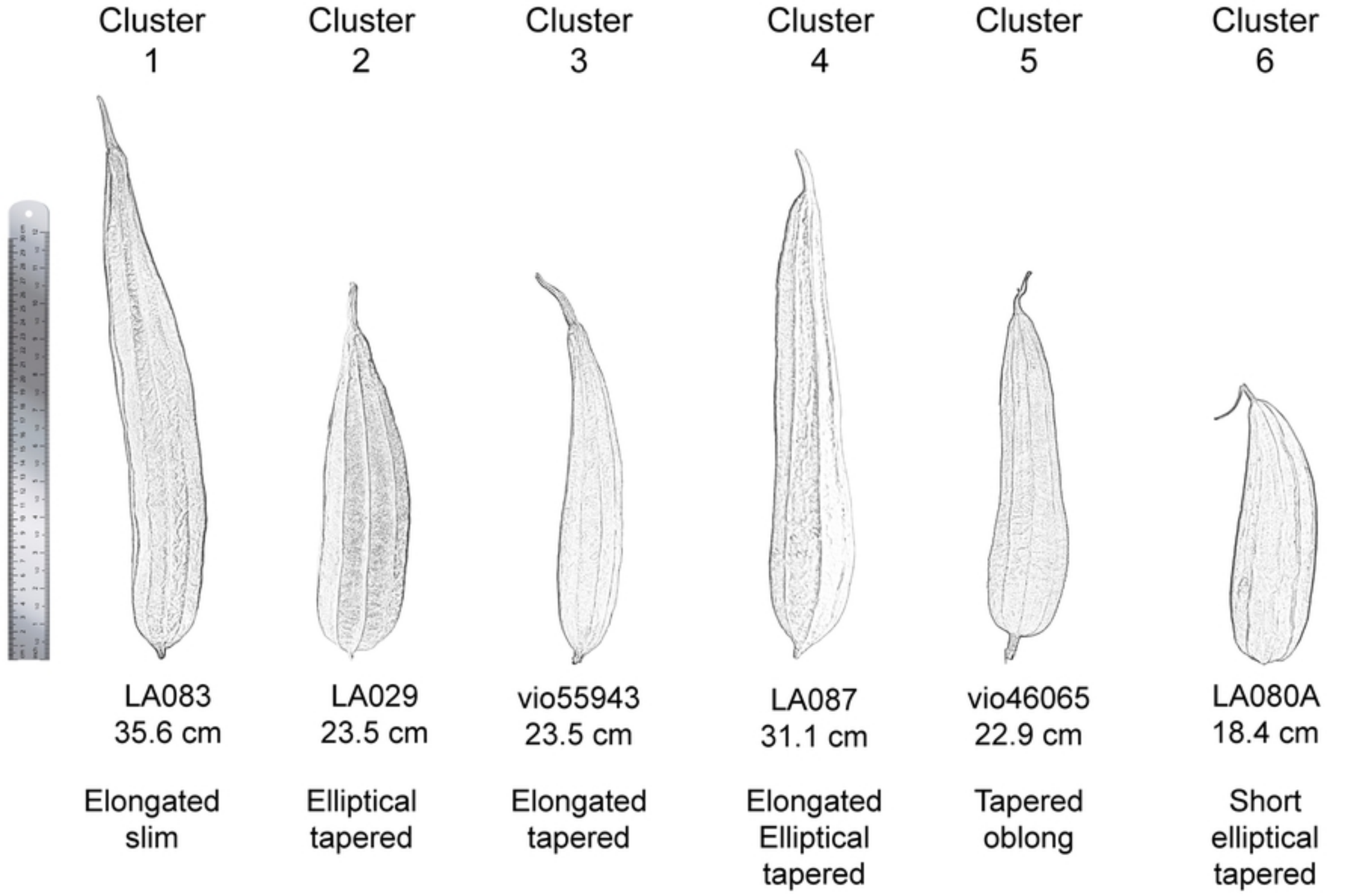
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
LA083
35.6 cm
LA029
23.5 cm
vio55943
23.5 cm
LA087
31.1 cm
vio46065
22.9 cm
LA080A
18.4 cm
Elongated slim
Elliptical tapered
Elongated tapered
Elongated Elliptical tapered
Tapered oblong
Short elliptical tapered

Fig7