Understanding the genetic relationships between Indonesian bambara groundnut landraces and investigating their origins

E. S. Redjeki[1] W. K. Ho[2,3], N. Shah[4], O.O. Molosiwa[4], N. R. Ardiarini [5], Kuswanto[5] and Sean Mayes[2,4]

1. University of Muhammadiyah Gresik, Jl. Sumatera No. 101 GKB Gresik 61121, Jawa Timur, Indonesia
2. Crops For the Future, Jalan Broga, 43500 Semenyih, Selangor, Malaysia
3. School of Biosciences, University of Nottingham Malaysia, Jalan Broga, 43500 Semenyih, Selangor, Malaysia
4. Plant and Crop Sciences, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, UK
5. University of Brawijaya, Veteran Street, Malang, 65145, Indonesia

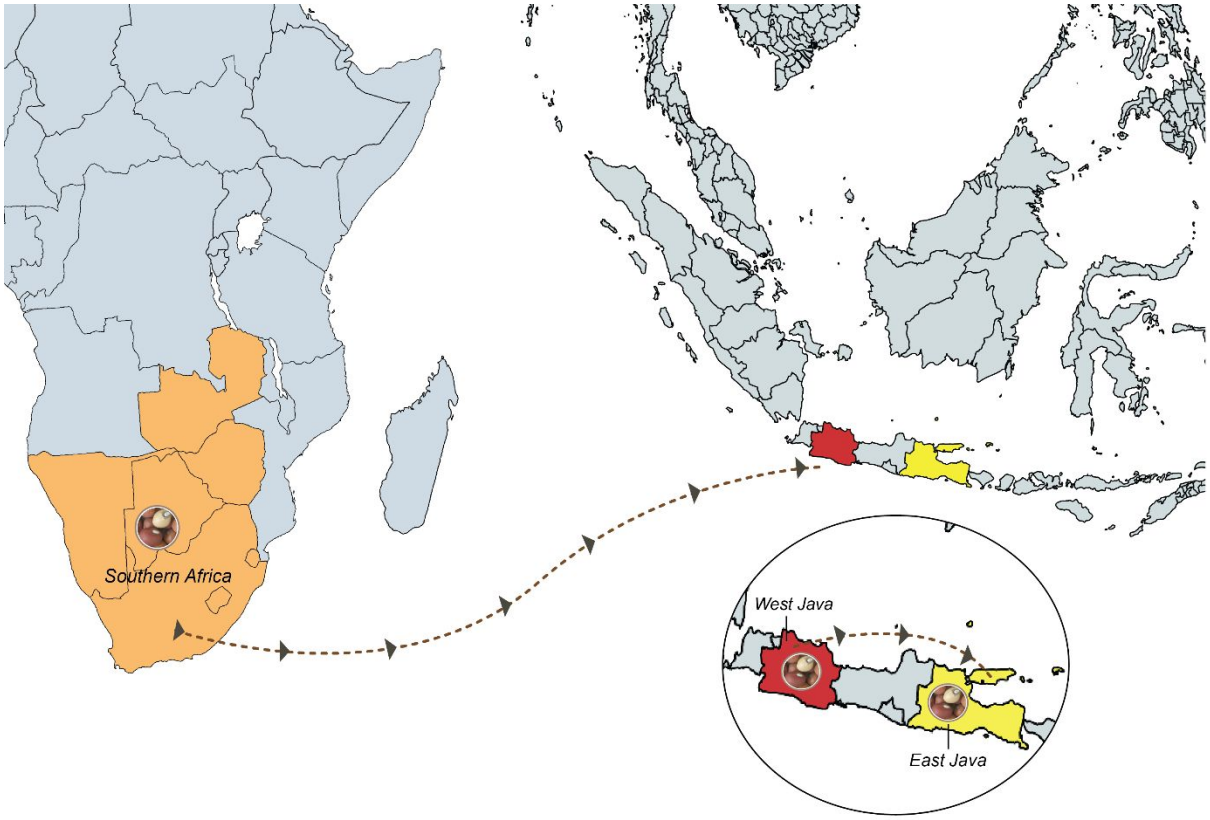
Corresponding author: sean.mayes@nottingham.ac.uk

Abstract

A total of 170 bambara groundnut (*Vigna subterranea*) accessions were evaluated using both Simple Sequence Repeats (SSR) and Single Nucleotide Polymorphism (SNP) markers generated using Genotyping-by-Sequencing (GbS) of which 56 accessions were collected from West and East Java. Principal Coordinate Analysis (PCoA), population structure and cluster analysis suggest that the East Java materials studied in this study could be as a result of the introduction of selected West Java materials. In addition, the current Indonesian accessions are likely to have been introduced from Southern Africa, undergoing a strong founder effect, with only a small fraction of the genetic variability within the species.

Keywords: *bambara groundnut, microsatellite, marker, SNP, genetic relationship, population structure*

Graphic abstract



World map diagram from www.mapchart.net

Introduction

Bambara groundnut (*Vigna subterranea (*L.) Verdc; $2n = 2x = 22$) belongs to the leguminous Fabaceae family with Burkina Faso, Cameroon, Democratic Republic of the Congo, Mali, Niger and Togo reported to be the main cultivation areas, producing approximately 180MT from 250,000 ha, annually (FAOSTAT, 2017). An extensive survey conducted in Zimbabwe across seven districts has revealed that the cultivation of bambara groundnut is highly district dependent and largely driven by the end use purpose, either as cash crop or for own consumption (Mubaiwa et al. 2018), while surveys in Ghana and Nigeria suggest that the crop is valued for its drought tolerance and food value, although suffers from a lack of improved varieties and being very labour intensive (Adwala et al. 2016a, 2016b; Olayide et al. 2018).

All of the bambara groundnut plant, including the leaf, stem, pod, seed, shell and 'offal', can be used for human consumption. In addition, it has been reported being used as herbal

55 medicine, as animal feed, a green fertilizer and a biopesticide (Mkandawire 2007, Daniel et al.
56 2016). The bambara groundnut seed is composed of 4.8% ash, 7.2% moisture, 47.0%
57 carbohydrate, 19.0% protein, 7.0% oil and 1.0% free fatty acid and compares well with other
58 legumes, although systematic approaches to comparing data on underutilised crops are limited
59 to-date (Okonkwo and Opara 2010, Halimi et al. 2019a, 2019b). Total calorific values for
60 bambara groundnut are reported to be higher than other pulses, such as pigeon pea, lentil and
61 cowpea; 367 kcal, 343 kcal, 354 kcal and 345 kcal, respectively (FAO, 1982). Comparatively,
62 bambara groundnut contains higher levels of lysine and methionine than other grain legumes,
63 making bambara groundnut an important staple in the diet in combating food and nutrient
64 security issues (Halimi et al. 2019a). Nevertheless, the lack of improved varieties has limited
65 its wider adoption to benefit subsistence farmers particularly those with marginal lands.
66
67 Begemann (1988) carried out detailed analyses of the seed-pattern diversity within a
68 large collection of bambara groundnut at the International Institute of Tropical Agriculture
69 (IITA), Nigeria. His conclusion strengthened the hypothesis that the centre of origin of
70 bambara groundnut is likely to be in the region of North-eastern Nigeria and Northern
71 Cameroon. A recent study comprising of 33 landraces (total sample number = 128) from 14
72 countries suggested that the gene flow of bambara groundnut germplasm was not only
73 influenced by geographic proximity but also the distribution was observed to contain a
74 linguistic component (Santos 2018).
75
76 Java, having a clear dry season in the middle and east is known as the main planting
77 area for bambara groundnut in Indonesia. Although extensive cultivation data is still lacking,
78 the production of bambara groundnut in 2007 from Sumedang district in West Java was
79 recorded to be 138 tonnes according to Widyasanti and colleagues (2019). Similarly,

information on the origin of bambara groundnut in Indonesia is scarce, with this species being native to Africa. One of the hypotheses was the crop was brought to Madagascar by the Arabians and subsequently spread to Brazil and Suriname in the early 17th century before being introduced to the Philippines and Indonesia (Adyi and Wahyudi 2018). Information on the origins of the original introduction(s) of germplasm into Indonesia are important for crop improvement and breeding programs in Indonesia, in order to widen the genetic base and also to introduce new traits of value to farmers. This is one area where molecular genetic tools could help to reveal the likely source of the introduction of bambara groundnut to Indonesia. Our previous study has shown that seeds derived from a single plant are essentially inbred suggesting that selecting from a single plant is an effective method to develop near-homozygous pre-breeding lines in this strongly inbreeding species (Molosiwa et al. 2015). Understanding the ancestral origin of germplasm and the genetic base conserved *in situ* by the bambara groundnut farmers in Indonesia would facilitate the development of a structured breeding programme. This would also shed light on how this crop has adapted to local humid growing conditions.

MATERIALS AND METHODS

Plant materials and DNA extraction

The plant materials consisted of 12 accessions from East Java, 44 from West Java, 16 from East Africa, 30 from Central Africa, 24 from Southern Africa and 44 from West Africa (Table S1) were planted in the climate-controlled glasshouse located at the Sutton Bonington Campus of University of Nottingham, UK. DNA was extracted from young leaflets using the GenElute Plant Genomic DNA kit (Sigma Aldrich) according to the manufacturer's instructions (Basu *et al*., 2007; Molosiwa, 2015). The DNA quality and quantity were evaluated under UV light on 1% Tris-borate-EDTA (TBE) agarose gel stained with ethidium bromide.

105

106 ***SSR genotyping***

107 After quantification, the DNA samples were diluted to approximately 10 ng/μL. A total of 11
108 codominant markers developed by Molosiwa et al. (2015) were used to assess the variation of
109 Indonesian materials (Table S2). The allele sizes were scored after the fragments were
110 separated using the CEQ^{TM} 8000 Genetic Analysis System (Beckman Coulter) with a 400bp
111 internal standard. Visual investigation of the allele pattern combined with the automated
112 scoring software were used to interpret the capillary electrophoresis results.

113 With the inclusion of data from samples reported by Molosiwa et al (2015), the allelic
114 sizes of 11 SSR markers were scored from a total of 170 accessions.

115

116 ***SNP genotyping***

117 SNP variation of samples (Table S1) were supplied by Diversity Array Technologies Pty Ltd,
118 Canberra, Australia (www.diversityarrays.com) using DArTseq^{TM} genotype-by-sequencing
119 method and a *Pst*I-*Taq*I genome complexity reduction method. Markers with minor allele
120 frequencies > 0.01 were considered as polymorphic. Population structure analysis was
121 performed using fastSTRUCTURE (Raj et al. 2014) and the 'chooseK' function was used to
122 suggest the optimal K value range.

123

124 ***Genetic diversity analysis***

125 The genetic measures of both types of markers including number of alleles per locus (N_a),
126 number of effective alleles (N_e), level of expected (H_e) and observed (H_o) heterozygosity and
127 fixation index (F, inbreeding coefficient) were computed using GenAIEx v6.5 (Peakall and
128 Smouse 2012) while SSR marker information was generated using PowerMarker v3.25 (Liu
129 and Muse 2005). GenAIEx v6.5 was also used for Principal Coordinate Analysis (PCoA) and

Analysis of Molecular Variance (AMOVA) with 999 permutations to assess the differentiation among subpopulations.

Hierarchical clustering analysis was carried out by using both Neighbour-Joining (NJ) with 10,000 bootstraps value and UPGMA methods calculated from 10,000 bootstraps of the 'Simple matching' Dissimilarity Index in DARwin v6 (http://darwin.cirad.fr/darwin) (Perrier et al. 2003).

RESULT AND DISCUSSION

In order to gain a better understanding of the genetic relatedness of Indonesian cultivars with those cultivated in African countries, bambara groundnut accessions collected from East Java, West Java, East Africa, West Africa, Central Africa and Southern Africa (Table S1) were evaluated by SSR (n = 170) and SNP markers (n = 168), respectively with 85 common single seed descent derived accessions (Table S1).

A total of 99 alleles with an average of nine alleles per locus were identified by 11 SSR markers (Table 1). The number of alleles (N_a) observed at each locus varied from four to 16 with the PIC values ranging from 0.11 to 0.83, and an average of 0.58 (Table S3). Eight of these had a PIC value of more than 0.5, and so are considered to be highly informative. Nevertheless, from the low bootstrap values for the nodes of the NJ tree (Figure 1c) as well as the relatively low levels of molecular variation explained in the PCoA plot (25.6% of the total variance explained by first two components, Figure 2a), the current set of SSR markers is not sufficiently informative to clearly distinguish the Central and West African accessions, although samples from East and West Java and also Southern and East Africa did cluster.

In terms of DArT Seq SNP markers, a total of 3,148 SNPs were obtained after filtering with minor allele frequency (MAF) > 0.01 and with no missing data across samples. The majority of the SNP markers (34.8%) fall into the high H_e index category and are prevalently are [CT] and [AG] types (Figure 3a & 3b). Among these, there was a total of 649 SNPs that could be considered as 'rare alleles' as their MAF values less than 5%.

The genetic diversity within individuals revealed by the genotypes evaluated in this study by both types of markers was low (Table 1). It was consistent across both types of makers with the mean observed heterozygosity (H_o) far lower than the average expected heterozygosity (H_e), reflecting the cleistogamous nature of bambara groundnut. Low observed heterozygosity from these markers (0.012±0.005 from SSR, 0.011±0.001 from SNP) suggested that seed from a single plant are likely to represent an unselected cultivar (without trait selection) and that a single round of seed collection from a single plant would (on average) be sufficient to achieve homozygosity in pre-breeding materials, consistent with our previous observation using SSR and dominant DArT markers (Molosiwa et al. 2015).

From Bayesian clustering analysis using SNP markers, three major clusters could be observed with a second peak at K = 5 (Figure 4 & S1). When K = 3, the subpopulation clustering coincided largely with their geographical origins; Q1: 76 accessions (45.2%) mainly from Central and Western Africa, Q2: 36 accessions largely consisted of accessions from Southern and East Africa and Q3: solely 56 Indonesian accessions (Table 2). However, when compared with the PCoA plot (Figure 2b), some of the accessions collected from Central African countries were at a distance from the West African group by the second principal component which explained 10.3% of the molecular variability. Interestingly, at K = 4, Q1 was not subdivided into two clusters as observed from the PCoA, instead Q2 was sub-divided into

two; I: 23 accessions predominantly from East and Southern Africa whilst II: 12 accessions with 11 from Southern Africa. Accessions from Southern Africa were seen to be relatively clustered even in Cluster I and Cluster II. The phylogenetic NJ tree with most of the nodes having a bootstrap value of more than 70% supported this grouping (Figure 1a & 1b). At K = 5, nine accessions (5.4%) were classified into an admixture (Q ≤ 70), three from Central Africa and the others from West Africa. Q1: 20 accessions (11.9%) with a majority from Central African, Q2: 22 accessions (13.1%) predominantly West African samples (*n* = 15), Q3: 25 accessions (14.9%), similar to Q2, 21 from West Africa and four from Central Africa, Q4: 36 accessions (21.4%) consisting of accessions primarily from Southern Africa (*n* = 23), Q5: 12 East Java and 44 West Java accessions (33.3%). This is in good correspondence with the UPGMA tree (Figure S2). Figure 5 summarises the total variance explained by two first coordinates when K = 3 and K = 5. There was no sample having a clear membership with any new cluster when K > 5. At K = 5, the distribution of accessions from Nigeria into Q2 and Q3 might suggests the existence of greater genetic diversity within the populations close to the centre origin. Furthermore, given that the fixation index was lowest in Q2 (when K = 5), higher genetic variability could be found in these accessions of which the majority are collected from Nigeria.

In brief, there are three major sub-populations that could be observed from the bambara groundnut accessions evaluated in this study. Overall, the resulting subpopulations and genetic clusters were mainly correlated to the geographic origins of the collection sites for the samples, suggesting that region-specific selection and potentially a founder effect have had a major role in influencing the diversity of bambara groundnut germplasm, with partially limited gene flow being observed between locations. Nevertheless, the influences from dietary habit and a cultural role for bambara groundnut should not be underestimated. For example, a survey

204 conducted in Zimbabwe has revealed that although peanut is the cash crop, in some districts
205 the cultivation areas of bambara groundnut are comparable or have exceeded the amount of
206 land allocated for growing peanut (Mubaiwa et al. 2018). Moreover, Santos (2018) detected a
207 linguistic signal in the distribution of bambara groundnut.
208
209 Both NJ and UPGMA dendograms (Figure 1a, 1b, 5a & 5b) also suggested that the
210 most likely origin of recent Indonesian materials is from Southern Africa. This is in good
211 correspondence with the previous report of Molosiwa et al. (2015) even though those authors
212 sampled a limited number of Indonesian lines (four out of 123 accessions). The Dutch shipping
213 routes between 1750 – 1800 could be speculated to be one of the plausible bambara groundnut
214 introduction routes to Java (Burn-Murdoch 2012). In addition, the analysis provides evidence
215 that the narrow genetic base of current East Java materials could result from the introduction
216 of limited West Java materials to East Java. This preliminary observation could be further
217 confirmed with the use of a wider germplasm set collected from the East Java cultivation
218 regions. The genetic base of Indonesian accessions could potentially be widened through the
219 introduction of genetic variation from another cluster; accessions grouped in Q2.I at K = 4.
220 Four Southern African accessions sharing the highest similarity with the Indonesian groups
221 and these were collected from Zambia, where the climate can be broadly classified into humid
222 subtropical or semi-arid steppe in different ecoregions.
223
224 There is moderately strong differentiation between the subpopulations regardless of
225 whether sub-clustered into three or five groups (F_{ST} = 0.251 and 0.259 respectively, Table 3),
226 indicating the groups are genetic distinct. The majority of the genetic variance occurred within
227 populations and accounted for 70 to 72% of the total variation, whereas only 25 to 26% was
228 attributed to the difference between subpopulations (Figure 5). This suggest that substantial

229 genetic variability may be accessed from within the same clusters, perhaps minimising the
230 disruption of adaptive complexes already in place. Cultivated at a small or subsistence scale
231 for centuries without strong selection pressure from the farmers, beyond matching the local
232 agroecosystem, landraces may well contain many allelic variants which have not experienced
233 strong selection. Both high subpopulation inbreeding coefficients, F_{IS}, and F_{IT} values, also
234 indicate that the lines making up these groups are inbred lines, consistent with the self-
235 pollination mechanism of bambara groundnut. Subpopulation Q2 and Q4 are the most diverged
236 groups, which could be partially contributed to by geographical barriers limiting material
237 exchange (Table S4).
238
239 Integration with agronomic and phenotypic data following molecular characterisation
240 would allow the informed development of crop improvement breeding programmes,
241 particularly with the availability of reference genome despite it being currently fragmented
242 (Chang et al. 2019). Application in genome-wide association mapping (GWAS) would identify
243 quantitative trait loci (QTL) or causal genes governing the traits of interest. Germplasm within
244 the same subpopulation identified in this study, particularly those collected from humid
245 subtropical regions, should be characterised in the field trials, if the goal is to improve bambara
246 groundnut in Indonesia.
247
248 **Conclusion**
249 The genetic clusters postulated in this study have shed light on the potential origin of bambara
250 groundnut cultivars in Indonesia from Southern Africa countries. Although the genetic base of
251 bambara groundnut in Indonesian is generally narrow, an understanding of the diversity of
252 bambara groundnut conserved *in situ* facilitates future breeding efforts towards development
253 of new cultivars with a wider genetic base or to mine favourable alleles from traits of interest.

Acknowledgment

This research was conducted with financial support from the EU FP6 INCO-DEV "BAMLINK" project and an ITPGRFA R3 W3 Benefit Sharing Fund Project (P26). The authors would like to thank Rossuraya Abdullah and Nur Zaffan Ariffin in assisting in the graphical design.

References

Adhi, R. K. and Wahyudi, S. 2018. Pertumbuhan and hasil kacang bogor (*Vigna subterranea* (L.) Verdc.) varietas local Lembang di Kalimantan Selatan. ZIRAA'AH 43: 192-197.

Adwala, W., Donkoh, S. A., Nyarko, G., O'Reilly, P. and Mayes, S. 2016a. Use patterns and perceptions about the attributes of bambara groundnut (*Vigna subterranea* (L.) Verdc.) in Northern Ghana. Ghana J Sci Technol Dev (GJSTD) 4: 56-71.

Adzawla, W., Donkoh, S.A., Nyarko, G., O'Reilly, P., Olayide Olawale, E., Mayes, S., Feldman, A. and Azman Halimi, R. 2016b. Adoption of bambara groundnut production and its effects on farmers' welfare in northern Ghana. Afr J Agric Res 11:583–594.

Basu, S., Roberts, J.A., Azam-Ali, S.N. and Mayes, S. 2007. Development of microsatellite markers for bambara groundnut (*Vigna subterranea* L. Verdc.) – an underutilized African legume crop species. Molecular Ecology Notes, 7: 1326 – 1328.

Begemann, F. (1988) Ecogeographic differentiation of Bambara groundnut (*Vigna subterranea*) in the collection of the International Institute of Tropical Agriculture (IITA)., PhD Dissertation, Technical University of Munich, Germany.

Burn-Murdoch J. 2012. 18th Century shpping mapped using 21st Century technology. The Guardian. Retrieved from https://www.theguardian.com/news/datablog/2012/apr/13/shipping-routes-history-map.

279 Yue Chang, Huan Liu, Min Liu, Xuezhu Liao, Sunil Kumar Sahu, Yuan Fu, Bo Song, Shifeng
280 Cheng, Robert Kariba, Samuel Muthemba, Prasad S Hendre, Sean Mayes, Wai Kuan Ho, Anna
281 E J Yssel, Presidor Kendabie, Sibo Wang, Linzhou Li, Alice Muchugi, Ramni Jamnadass,
282 Haorong Lu, Shufeng Peng, Allen Van Deynze, Anthony Simons, Howard Yana-Shapiro, Yves
283 Van de Peer, Xun Xu, Huanming Yang, Jian Wang, Xin Liu. 2019. The draft genomes of five
284 agriculturally important African orphan crops, GigaScience: 8, giy152,
285 https://doi.org/10.1093/gigascience/giy152.
286 Daniel, C., Wiafe, O.D. and Boye E.B.O. 2016. Restoration of biodiversity using Voandzeia
287 subterranean (bambara benes) [online]. The Quarry Life Award. Available from
288 https://www.quarrylifeaward.nl/node/29721 [accessed 31 Jul 2019].
289 FAO. 1982. Legumes in human nutrition. FAO food & nutrition paper No. 20, Food and
290 Agriculture Organisation of the United Nations, Rome, Italy.
291 FAO, 2017. FAOSTAT statistical database. Available from
292 http://www.fao.org/faostat/en/#data/QC [assessed 1 Dec 2017].
293 Halimi, A.R., Mayes, S., Barkla, B. and King, G. 2019a. The potential of the underutilized
294 pulase bambara groundnut (*Vigna subterranea* (L.) Verdc.) for nutritional food security. J Food
295 Compos Anal, 77: 47-59.
296 Halimi, A.R., Barkla, B.J. Andrés-Hernandéz, L., Mayes S. and King, G.J. 2019b. Bridging
297 the Food Security Gap: an information-led approach to
298 connect dietary nutrition, food composition and crop production J Sci Food and Agric
299 https://doi.org/10.1002/jsfa.10157.
300 Liu, K. and Muse, S.V. 2005. PowerMarker: An integrated analysis environment for genetic
301 marker Data. Journal Bioinformatics, 21 (9): 2128-2129.
302 Mkandawire, C.H. 2007. Review of bambara groundnut (*Vigna subterranea* (L.) Verdc.)
303 production in sub-Saharan Africa. Agricultural Journal 2(4): 464-470.

304 Molosiwa, O.O., Aliyu, S., Stadler, F., Mayes, K., Massawe, F., Kilian, A. and Mayes, S. 2015.
305 SSR marker development, genetic diversity and population structure analysis of Bambara
306 groundnut [*Vigna subterranea* (L.) Verdc.] landraces. Genet Resour Crop Evol, 62: 1225-43.
307 Mubaiwa J., Fogliano, V., Chidewe, C., Bakker, E.J., Linnemann, A.R. 2018. Utilization of
308 bambara groundnut (*Vigna subterranea* (L.) Verdc.) for sustainable food and nutrition security
309 in semi-arid regions of Zimbabwe. PLoS One 13(10): e0204817.
310 Okonkwo, S. and Opara, M. 2010. The analysis of bambara nut (*Voandzeia subterranea* (L.)
311 *thouars*) for sustainability in Africa. Research Journal of Applied Sciences, 5(6): 394-396.
312 Olawale, E., Donkoh, S.A., Ansah, I.G.K., Adzawla W., O'Reilly P.J., Mayes, S., Feldman A.,
313 Azman Halimi, R., Nyarkom, G., Ilori, C.O. and Alabi, T. 2019. Assessing socioeconomic
314 factors influencing production and commercialization of bambara groundnut as an indigenous
315 climate resilient crop in Nigeria. *In* Handbook of Climate Change Resilience. *Edited by* W.
316 Leal Filho W. Springer, pp 1-19. https://doi.org/10.1007/978-3-319-71025-9_158-1.
317 Peakall, R. and Smouse, P.E. 2012. GenAlEx 6.5: Genetic analysis in excel. Population genetic
318 software for teaching and research — An update. Bioinformatics, 28: 2537-2539.
319 Perrier, X., Flori, A. and Bonnot, F. 2003. Data analysis methods. In: Hamon, P., Seguin, M.,
320 Perrier, X., Glaszmann, J. C. Ed., Genetic diversity of cultivated tropical plants. Enfield,
321 Science Publishers. Montpellier. pp 43 - 76.
322 Raj, A., Stephens, M. and Pritchard, J.K. 2014 fastSTRUCTURE: Variational interference of
323 population structure in large SNP data sets. Genetics 197: 573-589.
324 Santos, R. O. 2018. Geo-genetics patterns in bambara groundnut: Investigating the role of
325 geography in the distribution of genetic vaiation. Ph.D thesis, School of Geography, University
326 of Nottingham, U. K.

Widyasanti, A., Silvianur, Zain, S. 2019. Pengaruh perlakuan blanching dan level daya pengeringan microwave terhadap karakteristik tepung kacang bogor (*Vigna subterranea* (L.) Verdcourt). Jurnal Teknologi Pertanian Andalas: 23: 80-90.

331 Table 1. Summary statistics of genetic variation at 11 SSR loci and 3,148 SNP loci across the entire germplasm collection with K=3 and K=5 for
332 the GbS data.

Markers	Sample	N	Na	Ne	I	H_e	H_o	F	Polymorphism
SSR	Total	170	9.000±1.314	3.535±0.555	1.386±0.209	0.606±0.079	0.012±0.005	0.984±0.006	0.58*
SNP	Total	168	2.000±0.000	1.479±0.006	0.439±0.004	0.286±0.003	0.011±0.001	0.953±0.003	0.21**
					K = 3,				
	Q1	76	1.923±0.005	1.228±0.004	0.280±0.003	0.163±0.002	0.010±0.001	0.940±0.004	-
	Q2	36	1.934±0.004	1.461±0.006	0.431±0.004	0.280±0.003	0.011±0.001	0.951±0.004	-
	Q3	56	1.931±0.005	1.507±0.006	0.446±0.004	0.295±0.003	0.012±0.001	0.949±0.004	-
					K = 5,				
	Q1	20	1.152±0.006	1.069±0.004	0.063±0.003	0.041±0.009	0.009±0.001	0.600±0.010	-
	Q2	21	1.119±0.006	1.046±0.003	0.044±0.003	0.028±0.002	0.010±0.001	0.468±0.010	-
	Q3	25	1.845±0.006	1.305±0.005	0.330±0.004	0.204±0.003	0.009±0.001	0.947±0.004	-
	Q4	36	1.938±0.004	1.488±0.006	0.446±0.004	0.293±0.003	0.011±0.001	0.948±0.004	-
	Q5	56	1.921±0.005	1.511±0.006	0.449±0.004	0.298±0.003	0.011±0.001	0.951±0.004	-

333 mean values except N; *: polymorphic information content (PIC), **: minor allele frequency (MAF)

334 N: sample size; Na: number of different alleles; Ne: effective number of alleles; H_e: expected heterozygosity; H_o: observed heterozygosity; F:
335 fixation index (inbreeding coefficient).

336

337 Table 2. Number and percentage of bambara groundnut accessions assigned into clusters among six regions

Region	**K = 3**			**K = 4**				**K = 5**					
	Q1	Q2	Q3	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q5	Admixture
Central Africa	29	1	0	29	0	1	0	16	6	4	1	0	3
West Africa	43	1	0	43	0	1	0	1	15	21	1	0	6
East Africa	3	11	0	3	1	10	0	3	0	0	11	0	0
Southern Africa	1	23	0	1	11	11	0	0	1	0	23	0	0
East Java	0	0	12	0	0	0	12	0	0	0	0	12	0
West Java	0	0	44	0	0	0	44	0	0	0	0	44	0
%	45.2	21.4	33.3	45.2	7.1	13.7	33.3	11.9	13.1	14.9	21.4	33.3	5.4

338

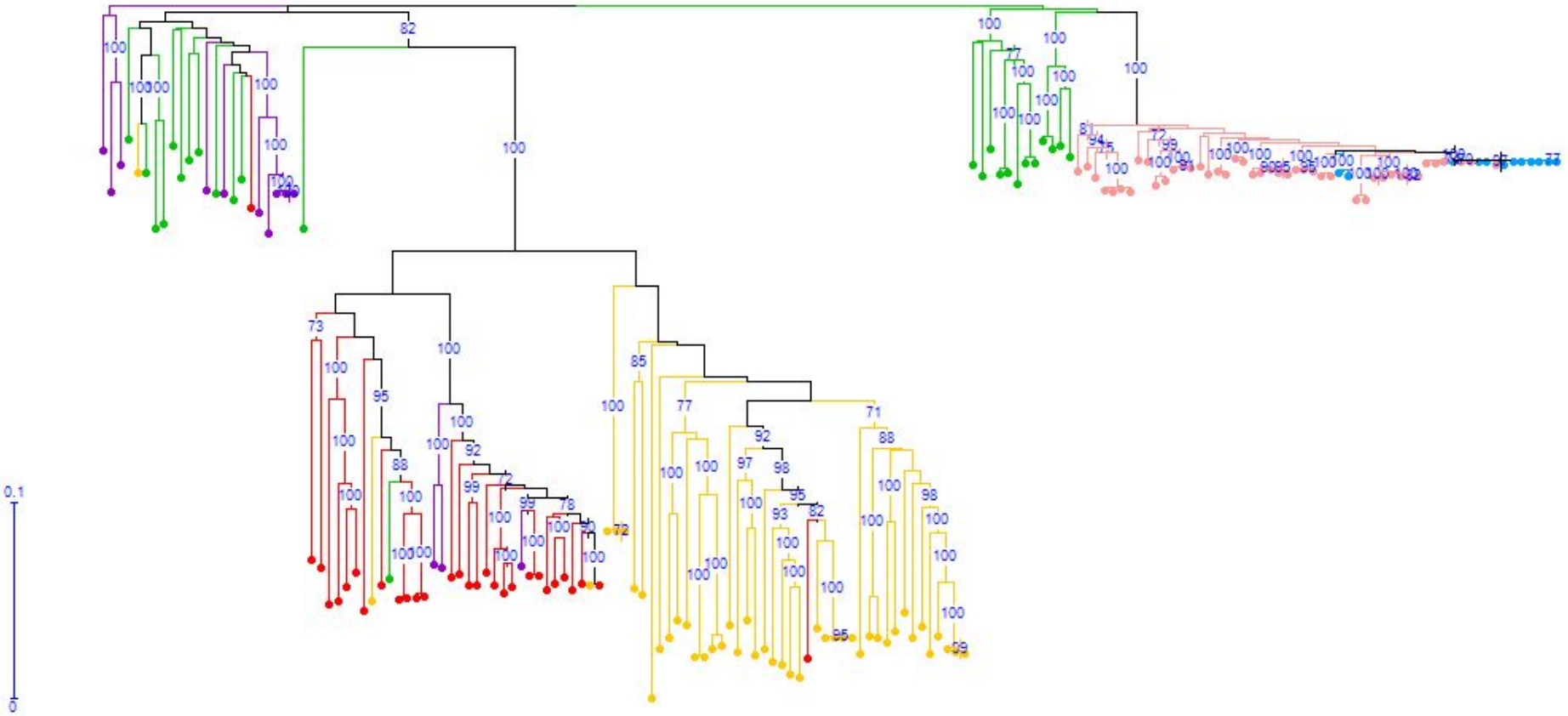
339

340

341 Table 3. Analysis of molecular variance among and within *V. subterranea* populations for K = 3 and K = 5.

Variation source	df	SS	MS	Estimated variance	%	F-statistics	*p*-value	Nm
When K = 3,								
Among groups	2	28238.63	14119.32	124.88	25	F_{ST} = 0.251	0.001	-
Among individuals	165	120067.98	727.69	355.38	71	F_{IS} = 0.955	0.001	-
Within individuals	168	2842.50	16.92	16.92	3	F_{IT} = 0.966	0.001	-
Total variation	335	151149.12		497.18	100	-	-	0.745
When K = 5,								
Among groups	4	31965.02	7991.25	121.24	26	F_{ST} = 0.259	0.001	-
Among individuals	153	103461.06	676.22	329.81	71	F_{IS}= 0.952	0.001	-
Within individuals	158	2623.50	16.60	16.60	4	F_{IT} = 0.964	0.001	-
Total variation	315	138049.58			100	-	-	0.714

343 a)



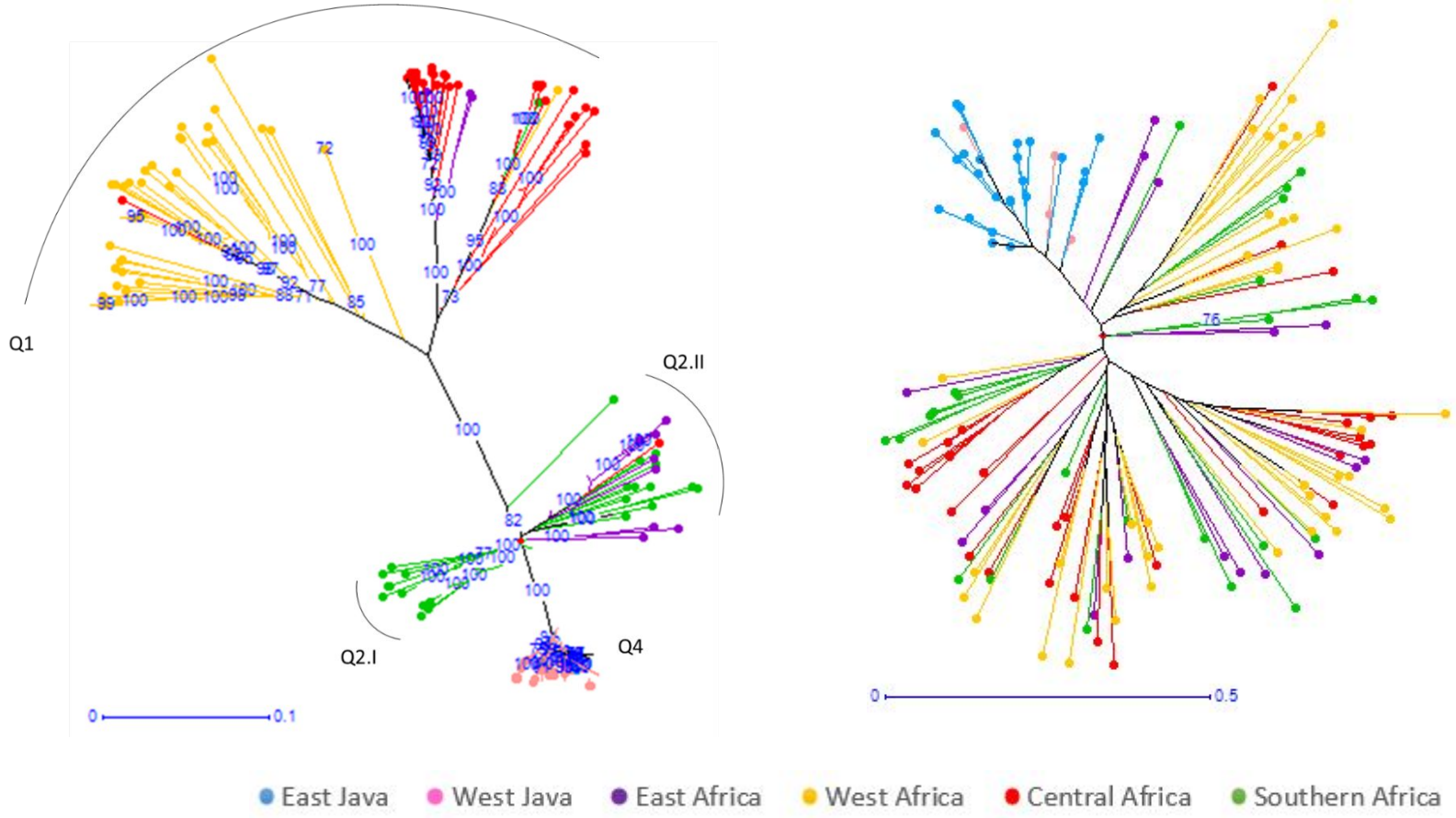
344

345

346

347

348 b) c)



349

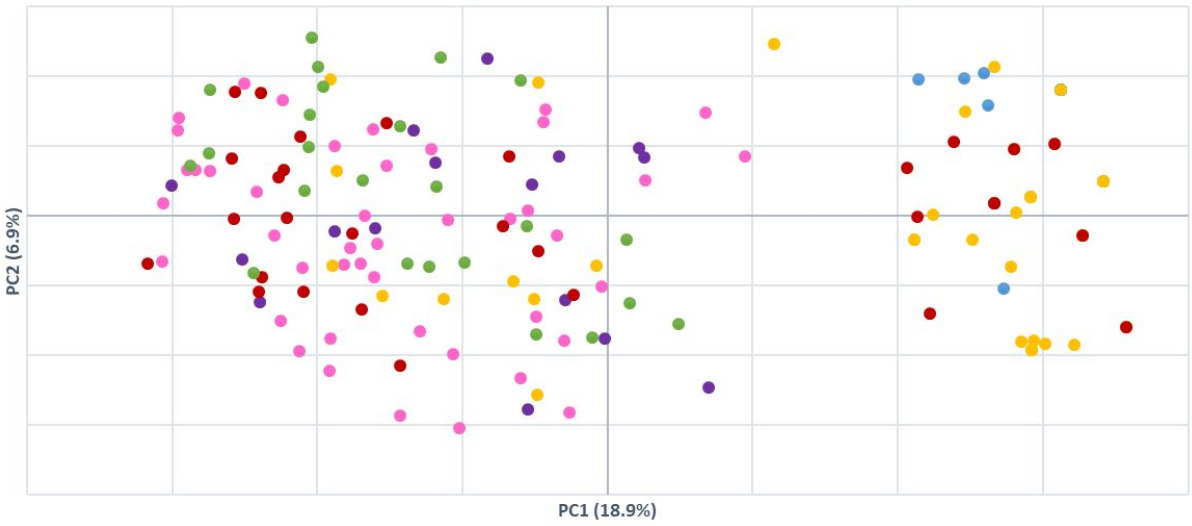
350

351 Figure 1. Neighbour-Joining tree of bambara groundnut accessions in this study based on a) & b) SNP markers and c) SSR markers, colours reflect

352 geographical origin and values in branches indicate bootstrap threshold ≥ 70. The clusters from SNP markers correlate with population structure

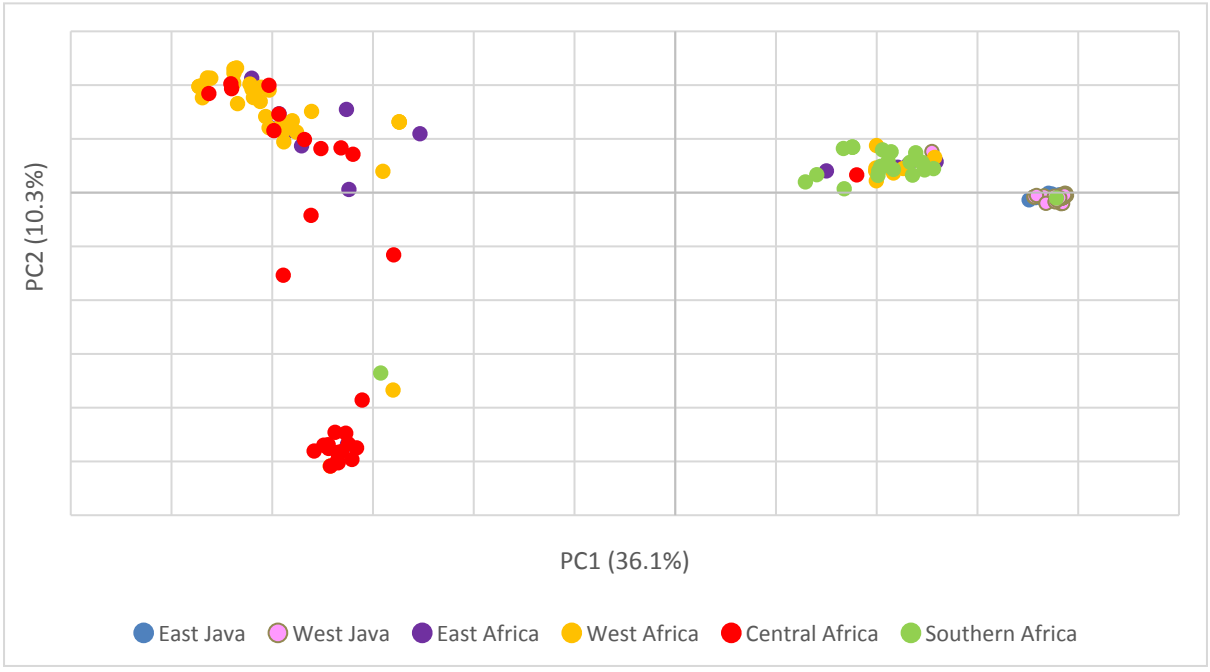
353 when K = 4.

354 a)



355

356 b)



357

358 Figure 2 Principal coordinates analysis (PCoA) based on genetic distance derived from a) 11

359 SSR markers and b) 3,148 SNP markers showing different clustering patterns.

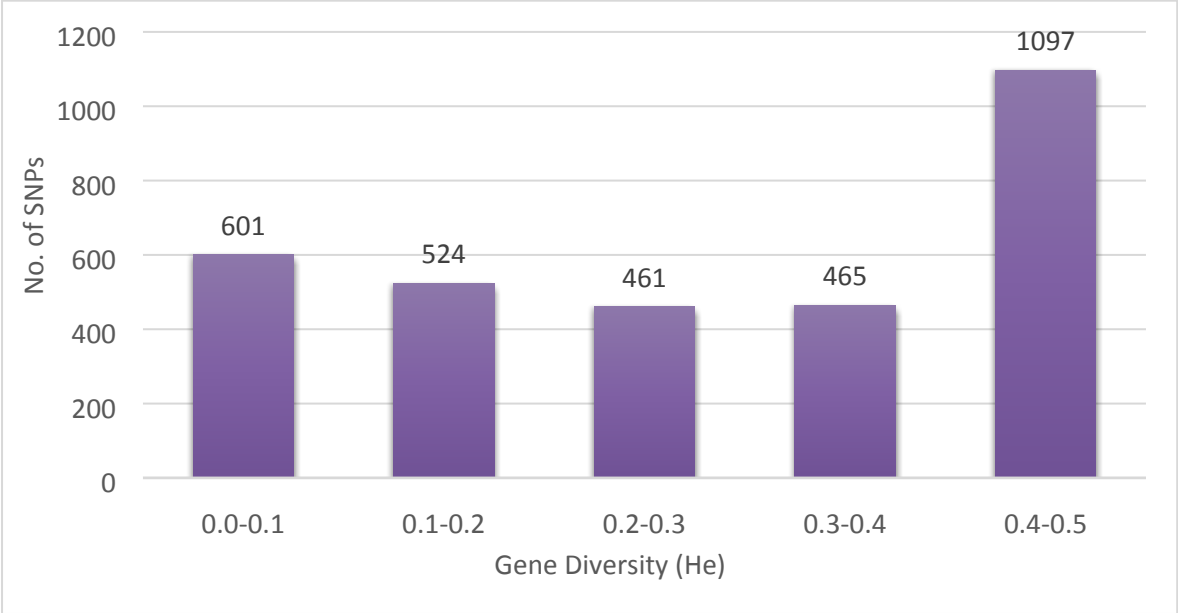
360

361

362

363

a)



b)

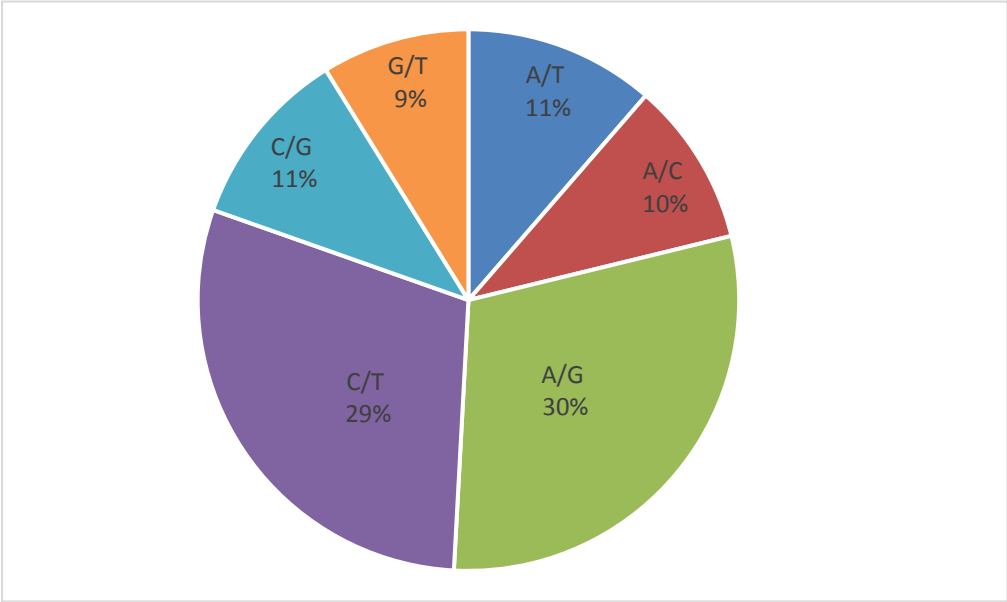
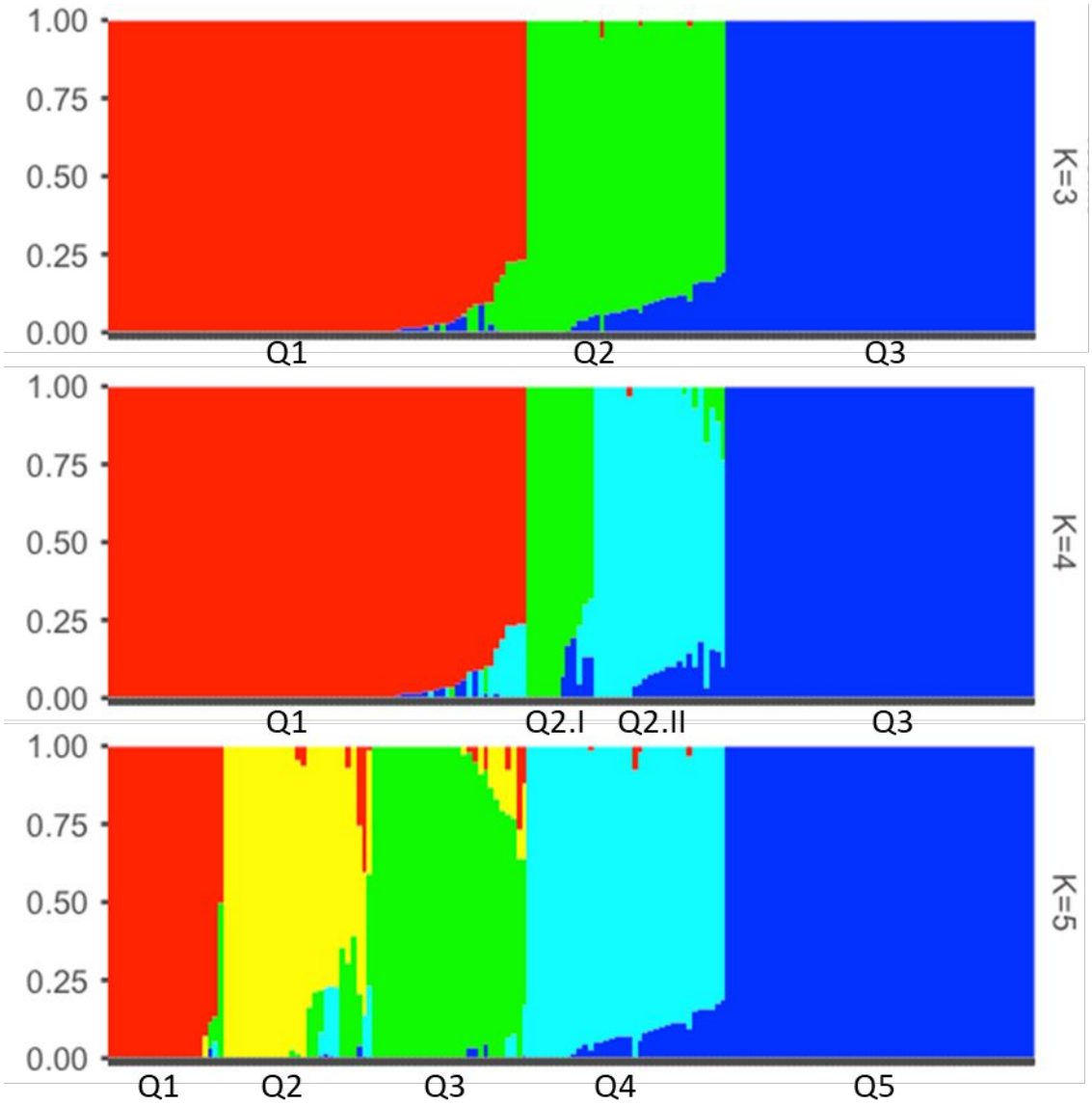


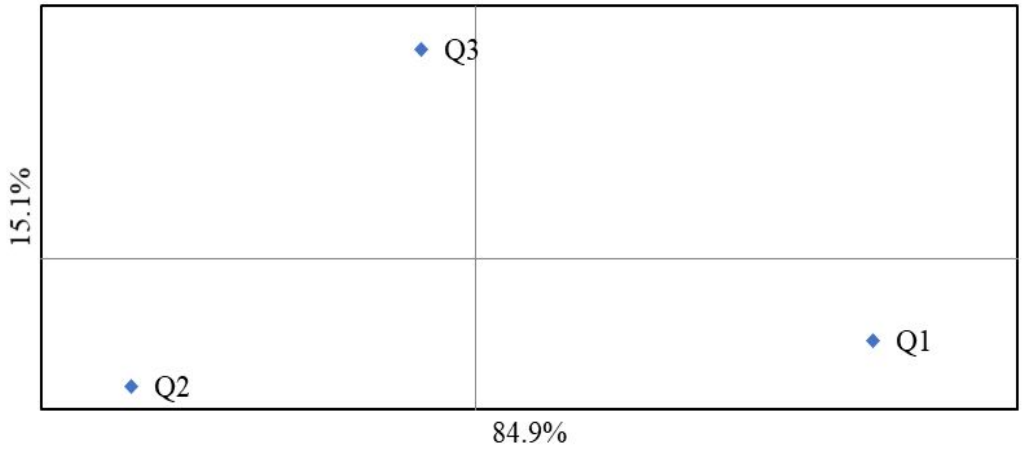
Figure 3. Distribution of a) genetic diversity (expected heterozygosity; H_e) b) SNP type for 3,148 SNP markers in the 168 *V. subterranea* accessions.

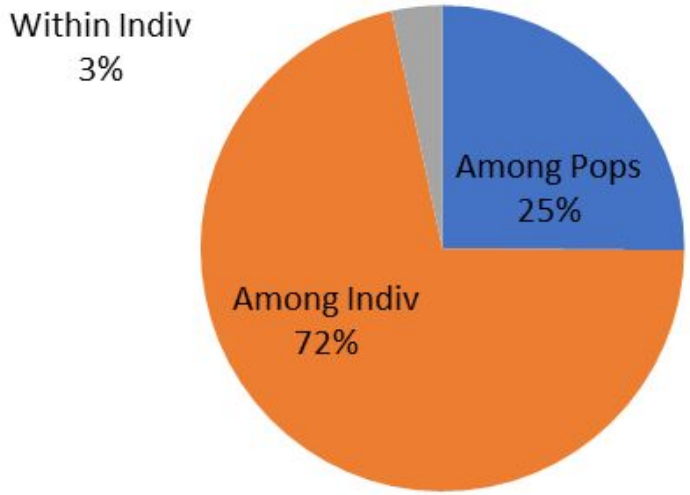


373

374 Figure 4. Clustering analysis showing K values from 3 to 5, with strong geographic signals.

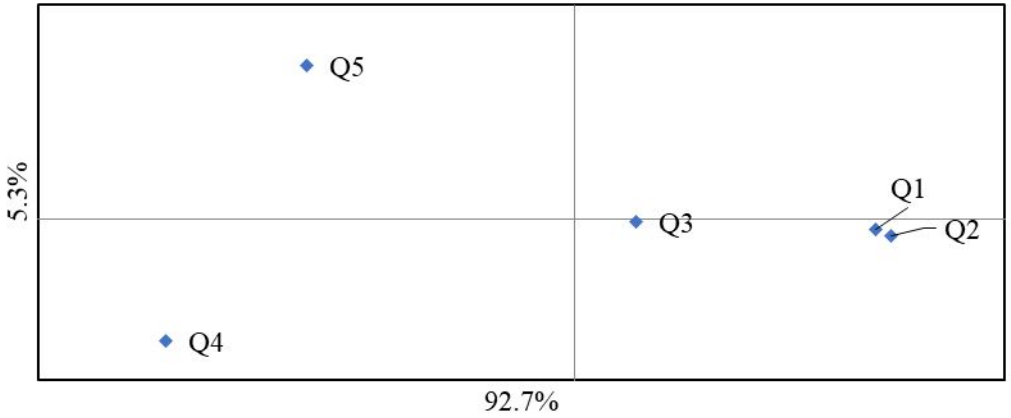
375 a) When K = 3

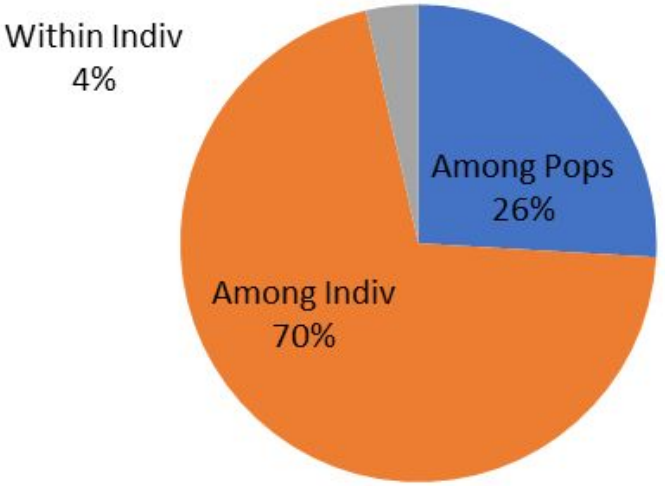




376

377 b) When K = 5,





378

379 Figure 5. PCoA of the subpopulations when a) K = 3 and b) K = 5, along with the AMOVA analysis explaining the total variance found among
380 and within groups.
381
382
383