Microsatellite Analysis of the Genetic Distance between 15 Potato (*Solanum tuberosum* L.) Genotypes

Tesfaye Abebe[1], C.D. Viljoen[2] and M.T. Labuschagne[2]
[1]Adet Agricultural Research Center, P.O.Box-08, Bahir Dar, Ethiopia; [2]Department of Plant Breeding, University of Orange Free State, P.O. Box-339, Bloemfontien 9300, South Africa

Abstract

Advanced breeding materials and commercial cultivars that possess a number of economically important attributes are good potential building blocks of a breeding program. Their efficient utilization, nevertheless, depends on the knowledge of the genetic distance existing between them. A total of 15 potato genotypes consisting of 11 advanced breeding clonal materials at multi-locational performance trial stage and four Ethiopian converted commercial cultivars of European origin were evaluated using the molecular marker called Simple Sequence Repeat (SSR) or microsatellites to determine the genetic distance between them. Three primers flanking to the gene associated with the potato wound-induced genes WIN1 and WIN2 and starch synthase were used to detect for variation in repeat length. The three primer pairs did reveal polymorphism, detecting between four and six alleles at each locus. Of these, two gave rise to complex band patterns, suggesting that multiple polymorphic loci were being amplified using a single primer pair. Consequently, 74 clearly scorable bands were recorded upon which a diversity index (H), a genetic distance matrix and cluster analysis were performed. The heterozygosity values ranged from 0.63 to 0.82. Similarly, the genetic distance matrix ranged between 0.25 through 0.89. The lowest genetic distance was observed for the genotypes CIP-387676.24 and Tolcha whilst the maximal distance was between KP-90134.5 and CIP-387676.24. Phenetic analysis of the derived information allowed the construction of a dendrogram depicting the genetic distance between the studied genotypes. Consequently, two main clusters and three singletons were formed. The potential of microsatellite markers for genetic distance analysis and parental material selection is discussed in the light of these results.

Introduction

The improvement of any crop species is facilitated by the introduction of novel genes for agronomic traits such as yield, as well as traits such as pest resistance and abiotic stress tolerance (Senior et al., 1998). These genes are most commonly gleaned from the primary gene pool of the crop species of interest. Advanced breeding materials that possess many economically important attributes are among this primary gene pool and can be used in the improvement of a crop species. An efficient utilization of these genetic resources, however, depends on the knowledge of the genetic distance between them. Information on the genetic distance among lines selected in a breeding program helps to avoid genetic erosion and thereby improve both maintenance of available genetic diversity and management of genetic resources (Goodman, 1994).

A combination of morphological and agronomic traits has been used and continues to be used to measure, describe and classify the genetic diversity of wild and cultivated plants and distinctly identify cultivars (Hawkes, 1994). The effect of environmental and management practices on morphology (Smith and Smith, 1992) and the few or low level of polymorphism (Tatineni et al., 1996) of morphological markers limited the use of numerical taxonomy to measure the genetic diversity and relationships in a reliable manner. DNA polymorphism assay for marker assisted plant breeding, genetic mapping, germplasm management, and investigation of genetic divergence has become an important tool in recent years. Various genetic marker systems and their related techniques are also available (Rus-Kortekaas et al., 1994). Although genetic markers are ideal for measuring divergence between two genotypes, the existing procedures have several limitations. For example, isozymes are available only for a limited number of loci and are influenced by plant development (Desborough and Peloquin, 1968) while that of restriction fragment length polymorphisms (RFLPs) require relatively large samples and take several days to examine (Gebhardt et al., 1989).

Similarly, random amplified polymorphic DNA (RAPD) analysis lacks the desired stringency (Demeke et al., 1993). As a result, these

procedures have been used only on a limited basis. Alternatively, the DNA marker system

called Simple Sequence Repeats (SSRs) or microsatellite are now in common use and become established as the marker system of choice in the studies of human and mammalian genomes (Murray et al., 1994; Dietrich et al., 1994). SSRs are short sequences of one to five base pairs (bp) repeated in tandem with an overall length in the order of tens of base pairs that are uniformly distributed in all eukaryotic genomes and are frequently polymorphic, with a variable number of repetitive elements (Hamada et al., 1982; Tautz and Renz, 1984). These repeat DNA sequences are found in transcribed as well as non-transcribed sequences of eukaryotic genome (Braaten et al., 1988) often localized around centromers and at the telomers of chromosomes (Russell, 1986; Wintro et al., 1992). They might play a role in chromosome pairing, meiotic recombination or speciation and gene mutation (Ros and Dolittle, 1983; Schmid, 1996).

SSRs have been observed in barley and Chickpea (Weising et al., 1991), grapes (Thomas et al., 1993), rice (Zhao and Kochert, 1992), soybean (Morgante and Olivieri, 1993), tomato (Ras-Kortokaas et al., 1994), and trees (Condit and Hubbell, 1991). Veilleux et al (1995) recently used SSRs to characterize amber-derived plants of a diploid potato clones. Variability analysis of SSRs marker system on soybean (Akkaya et al., 1992), rice (Wu and Tanksley, 1993), and lettuce (Bell and Ecker, 1994) indicated its ten fold more variability than

RFLPs. Westman and Kresovich (1999) reported that SSRs polymorphism is consistent with the variation patterns revealed by the species breeding history. Senior et al (1998) also reported the similarity of the genetic divergence pattern revealed by the SSR polymorphism and that of known pedigrees. Smulders et al (1997) have also demonstrated the use of SSRs in distinguishing cultivars that are genetically closely related to each other. Roder et al (1995) were also reported the detection of significantly more variations with the SSR marker than RFLPs markers in an intraspecific wheat genotypes with, on average, 4.6 different alleles per microsatellite. Hence, PCR amplified microsatellite loci may provide a good alternative marker system to reliably detect the genetic divergence in any species (Smulders et al., 1997; Chavarriaga-Aguirre et al., 1999).

Microsatellites or SSRs loci can be readily amplified by the PCR using primers that are complementary to the regions flanking the repeat (Akkaya et al., 1992). The resulting PCR products are then resolved electrophoretically in denaturing polyacrylamide gels (Wu et al., 1994). Consequently, the microsatellite or SSRs marker system was employed in this particular study to investigate the genetic distance and useful characters among four commercial cultivars and 11 advanced potato breeding materials at multi-location performance trial.

Materials and Methods

Genetic materials

A total of 15 potato genotypes consisting of 11 advanced tetraploid breeding clonal materials at multi-locational performance trial stage and four Ethiopian converted commercial cultivars of European origin (Table 1).

Table 1: List of potato genotypes used in the Study

Treatment number	Genotype
1	CIP-387879.8
2	MENAGESHA
3	CIP-388790.3
4	KP-93116.5
5	GUMER
6	KP-93143.5
7	CIP-388423.13
8	WOCHOA
9	CIP-387879.20
10	KP-93134.5
11	KP-93134.2
12	KP-93126.12
13	TOLCHA
14	CIP-387879.24
15	AWASH

DNA Extraction

DNA was extracted from fresh young tissues based on the modified monocot protocol of Edwards *et al.* (1991). Young plant leaf tissue collected from one plant of each clone was used to isolate DNA for fingerprinting analysis. Tissue was frozen with liquid nitrogen and ground into a fine powder using a mortar and pestle. Ten ml of extraction buffer (1 M Tris-HCl pH 8; 0.25 M EDTA and 20% SDS) was added to each sample and then incubated at 65 °c for half an hour with inversion every ten minutes, thereafter, 1 ml Cetyl triethyl ammonium bromide (CTAB) (10% w/v) and 2 ml 5 M NaCl were added to the homogenate and incubated for a further hour with gentle agitation every 10 minutes. Chloroform-isoamyl alcohol (24:1) extractions were performed at room temperature with centrifugation at 10 000 rpm for 15 minutes to remove proteins. After centrifugation the DNA was precipitated by the addition of 2 volumes (v/v) of ethanol and stored at 4 °c. The precipitated DNA was spooled and washed with 70% ethanol to remove residual salt before redissolving in sterile double distilled water and stored at 4 °C. DNA concentration was estimated using UV spectrophotometery. Then the DNA solution was diluted to a working concentration of 100 ng μ L^{-1} and stored at 4 °C. Finally the integrity and concentration of the DNA was confirmed by agarose electrophoresis and visualized with ethidium bromide under UV light.

SSR characterization

PCR was carried in a total reaction volume of 25 µL containing 2.5 µL 10x PCR buffer (20 mM Tris HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$), 1 unit of Taq DNA polymerase (Gibco), 10-pica mole (0.5 µM) of each primer (Table 2) and 20 ng template DNA. Amplifications were performed in Hybaid thermal cycler using the following parameters: one cycle of 4 min denaturation at 95 °c, 40 sec annealing at 55 °c and 2 min extension at 72 °c; 35 cycles of 20 sec denaturation at 95 °c, 40 sec annealing at 55 °c, and 2 min extension at 72 °c followed by a final extension at 72 °c for 4 min. Amplicons were separated in a vertical 10% PAGE gel in 1xTAE buffer at 70 V for 8 hours. After electrophoresis, the gel was stained with ethidium bromide and visualized by UV illumination. The DNA profiles were documented and scored manually by assigning a value of 1 for band presence and 0 for band absence. The scores of band presence or absence were used to calculate a pairwise genetic distance matrix using the Euclidean distance formula of Excoffier *et al.* (1992).

$$E = n\,[1-(2n_{xy}/2n)]$$

Where, $2n_{xy}$ equals the number of shared bands and n equals the total number of banding positions. Finally a phenogram summarising the genetic distance between the genotypes studied was constructed based on the most commonly used clustering algorithm of unweighted pair-group method using an arithmetic average (UPGMA) cluster analysis of the genetic distance matrix by Number Cruncher Statistical System for Windows (NCSS 2000). In total, n (n-1)/2 or 105 pairwise comparisons were made of the SSR DNA products obtained from the 15 genotypes. The level of polymorphism ofeach locus was also computed using the genetic diversity index formula of Nei (1978) and Shaghai Maroof *et al.* (1994),

$PIC=1-\sum p_i^{\,2}$, where PIC is polymorphic information content and p_i is the frequency of the i^{th} SSR allele.

Results and Discussion

The three microsatellite primer pairs (Table 2) have amplified between four and six alleles at each of the polymorphic loci, giving rise to heterozygosity values of between 0.63 to 0.82. Average values for number of alleles and hetrozygosity were 5 and 0.71, respectively. In total 74 clearly scorable bands were recorded. The number of amplified products observed ranged between zero to 10 at the (TGAAA)2(ATA)6 locus, zero to eight at the (TCAC)m and

zero to 12 at (TCAC)m.(CTT)n loci for the individual genotypes. This procedure is very discriminating having produced unique DNA profiles for all of the 15 genotypes examined. Diversity values of 0.63, 0.67 and 0.82 were computed for the (TGAAA)2 (ATA)6, (TCAC)m .(CTT)n and (TCAC)m SSRs, respectively, among the genotype tested.

Phenetic analysis of the relationships between the examined genotypes based on the microsatellite results, shared band analysis in NCSS (2000) was used to produce dendrogram shown in Figure 1 below. The reasons for the groupings obtained are not obvious from co-ancestry as it is difficult to establish the precise pedigree of all the examined genotypes. The estimated genetic distance between the genotypes based on the three microsatellite loci was 25-89% (Table 3). These distance levels are almost certainly the result of the high polymorphism revealed by microsatellites at each locus and the approach taken to analyse the data, which involved scoring each allele as a unique character. The latter is particularly relevant since in autotetraploid outbreeder each individual could contain between one and four different alleles at any one locus. Hence, there may be up to 8 different alleles within the gene pool studied. In conclusion, two main clusters and three singletons were formed. The lowest genetic distance, 0.25, was observed between the commercial cultivar Tolcha and the advanced breeding clone CIP-387676.24. Conversely, the maximum genetic distance of 0.89 was observed between the advanced genotypes KP-90134.5 and CIP-387676.24. Wochecha, CIP-386423.13, KP-90143.5, CIP-387676.20 and Menagesha were clustered under the first cluster while KP-90138.12, CIP-387676.24, Tolcha, KP-90108.5, Genet, CIP-389701.3 and CIP-387676.8 grouped under the second cluster. KP-90134.5, KP-90134.2 and Awash were distantly separated from the rest genotypes. Hence, a clear distance is observed between the genotypes in a breeding program. This in turn shows the relevance of genetic distance studies for the purpose of marker assisted plant breeding, genetic mapping, germplasm management, investigation of genetic divergence, and parental line selection.

References

Akkaya, M.S., Bhagwat, A.A. and Cregan, P.B.1992. Length polymorphisms simple sequence repeat DNA in soyabean. *Genetics 132:1131-1139.*

Beckman, J.S. and Weber, J.L.1992. Survey of human and rat microsatellites. *enomics12:627-631.*

Bruford, D.C., J.R. Thomas, R.D. Little, K.R. Dickson, I. Goldberg, D. Schlessinger, A. Ciccodicola, and M. d'Urso. 1988. *Nucleic Acids Res. 16:865-881.*

Chavarriaga-Aguirre, P., Maya, M.M., Tohm, J., Duque, M.C., Iglesias, C., Bonierbale, .W.,Kresovich, S. and Kochert, G. 1999. Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collectionand to assess the usefulness of DNA-based markers to maintain germplasm collections. *Mol. Breeding 5: 263-273.*

Demeke, T., Kawchuk, L. M. and Lynch, D.R. 1993. Identificaton of potato cultivars and clonal variants by random amplified polymorphic DNA analysis. *Am. Pot. J. 70: 561-570.*

Desborough, S. and S. J. Peloquin. 1968. Potato variety identification by use of electrophoretic patterns of tuber proteins and enzymes. *Am Potato J 45:220-229.*

Dietrich, W.F., J.C. Miller, R.G. Steen, M. Merchant, D. Damron, R. Nahf, A. Gross, D.C. Joyce, M. Wessel, R.D. Dredge, A. Marquis, L.D. Stein, N. Goodman, D.C. Page, and

E.S. Lander. 1994. A genetic map of the mouse with 4006 simple sequence length polymorphism. *Nature Gener. 7:220-225.*

Edwards, K., Johnstone, C. and

Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics 131: 479 - 491.*

Gebhardt, C., Blomendahl, C., Schachtschabel, U., Debener, T., Salamini, F., and Ritter, E.1989. Identification of 2n breeding lines and 4n varieties of potato (*Solanum tuberosum subsp. tuberosum*) with RFLP fingerprints. *Theor. Appl. Genet. 78:16-22.*

Goodman, M.M.1994. Maize.p.192-202. In N.W. Simmonds (ed.) Evolution of Crop Plants. Longman Group, Ltd., Essex, England.

Hamada, H., M.G. Petrino, and T. Kakunaga. 1982. A novel repeat element with Z-DNA forming potential is widely found in evolutionary diverse eukaryotic genomes. *Proc Natl Acad Sci USA 79:6465-6469.*

Hawkes, J.G. 1994. Origins, Species and Cytology of potatoes. In: Bradshaw, J.E. and Mackay, G.R. (eds.). *Potato Genetics. CAB International. Wallingford, and U.K.* Pp 3-42.

Morgante, M. and A.M. Olivieri. 1993. PCR-amplified microsatellites as markers in plant genetics. *Plant J 3:175-182.*

Murray, J.C., K.H. Buetow, J.L. Weber, S. Ludwigsen, T. Scherpbier-Heddema, F. Manion, J. Quillen, V.C. Sheffield, S. Sunden, G.M. Duyk, J. Weissenbach, G. Gyapay, C. Dib, J. Morrissette, G.M. Lathrop, A. Vignal, R. White, N. Matsunami, S. Gerken, R. Melis, H. Albertsen, R. Plaetke, S. Odelberg, D. Ward, J. Dausset, O. Cohen, and H. Cann. 1994. Comprehensive human linkage map with centimorgan density. *Science 265:2049-2054.*

Roder, M.S., Plaschke, J., Konig, S.U., Borner, A., Sorrells, M.E. Tanksley, S.D. and Ganal, M.W.1995. Abundance, variability and chromosomal location of microsatellites in wheat. *Mol. Gen. Genet. 246:327-333.*

Rose, M.R. and Doliite, W.F. 1983. Molecular biological mechanisms of speciation. *Science* 220. 157-162.

Russell, P.J. 1986. Genetics. Little, Brown and Company, Toronto, Canada. pp. 34- 63.

Rus-Kortekaas, W., M.J.M. Smulders, P. Arens, and B. Vosman. 1994. Direct comparision of levels of genetic variation in tomato detected by a GACA-containing microsatellite probe and by random amplified polymorphic DNA. *Genome 37:375-381.*

Schmid, C.L.1996. Alu: Structure, origin, evolution significance, and function of one-tenth of human DNA. *Progress in Nucl.Acid Res. And Mol. Biol. 53: 283-319.*

Senior, M.L.,Murphy, J.P.,Goodman, M.M., and Stuber,C.W.1998. Utility of SSRs for determination of genetic similarities and relationships in maize using agarose gel system. *Crop Sci. 38:1088-1098.*

Shughai Marcof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q. and Allard, W. 1994. Extraordinarily polymorphic microsatellite DNA in barely: species diversity, chromosomal locations, and population dynamics. *Proc Natl Acad Sci 91: 5466-5470*

Smith, J.C. S. and Smith, O. S. 1992. Fingerprinting crop varieties. *Adv. Agron. 47:85-140.*

Smulders, M.J.M., Bredemeijer, G., Rus-Kortekaas, W., Arens, P., and Vosman, B. 1997. Use of short microsatellites from data base sequences to generate polymorphisms among lycopersicon esculentum cultivars and accessions of other lycopersicon species. *Theor. Appl. Genet. 97:264-272.*

Tatineni,V.,Cantrell,R.G.,and Davis,D.D.1996. Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. *Crop Sci. 36:186-192.*

Thompson, C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR- analysis. *Nucl. Acids Res. 19: 1349.*

Tautz, D. and M. Renz. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res 12:4127-4138.*

Thomas, M.R., S. Matsumoto, P. Cain, and N.S. Scott. 1993. Repetitive DNA of grapevine:classes present and amplified polymorphic DNA and simple sequence repeats. *Genome 38:1153-1162.*

Weising, K., B. Beyermann, J. Ramser, and G. Kahl. 1991. Plant DNA fingerprinting with radioactive and digoxygenated oligonucleotide probes complementary to simple repetitive DNA sequences. *Electrophoresis 12:159-169.*

Weising, K., F. Weigand, A.J. Driesel, G. Kahl, H. Zischler, and J.T. Epplen. 1989. Polymorphic simple GATA/GACA repeats in plant genomes. *Nucleic Acids Res 17: 10128.*

Westman, A.L. and Kresovich, S.1999. Simple sequence repeat (SSR)-based marker variation in *Brassica nigra* genebank accessions and weed populations. *Euphytica 109:85-92.*

sequences suitable for cultivar identification. *Theor Appl Genet 86:173-180.*

Veilleux, R.E., L.Y. Shen, and M.M. Paz. 1995. Analysis of the genetic composition of anther derived potato by random

Wintero, A.K., Fredholm, M., and Thomsen,P.D. 1992.Variable (dG-dT)n.(dC-dA)n sequences in the porcine genome. *Genomics 12:607-609.*

Wu, K.S., R. Jones, L. Danneberger, and P.A. Scolnik. 1994. Detection of microsatellite polymorphisms without cloning. *Nucl. Acids Res 22(15): 3257- 3258.*

Wu, K.S. and Tanksley, S.D. 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol. Gen. Genet. 241: 225-235.*

Zhao, X. and G. Kochert. 1992. Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.). *Mol Gen Genet 231: 353-359.*

Table 2. Microsatellite primer sequences used, number of alleles, range of allele size, and the gene diversity obtained in this study

Name	Associated gene	Simple sequence Repeat	Primer sequence (5'-3')	Number of alleles detected	Gene diversity	Range of allele size (bp)	Total number of alleles scored
STWIN 12 G	Potato wound-induced genes WIN1 and WIN 2	(TGAAA)2(ATA)6	TGTTGATTGTGGTGATAA TGTTGGACGTGACTTGTA	4	0.63	<100 – 400 bp	19
STS1+2	Starch synthase	(TCAC)m	TCTCTTGACACGTGTCACTGAAAC TCACCGATTACAGTAGCGCAAGAGA	6	0.82	<100 – 500 bp	31
STS1+3	Starch synthase	(TCAC)m.(CTT)n	TCTCTTGACACGTGTCACTGAAAC TTGCCATGTGATGTGTGGTCTACAA	5	0.67	400 – 700 bp	24

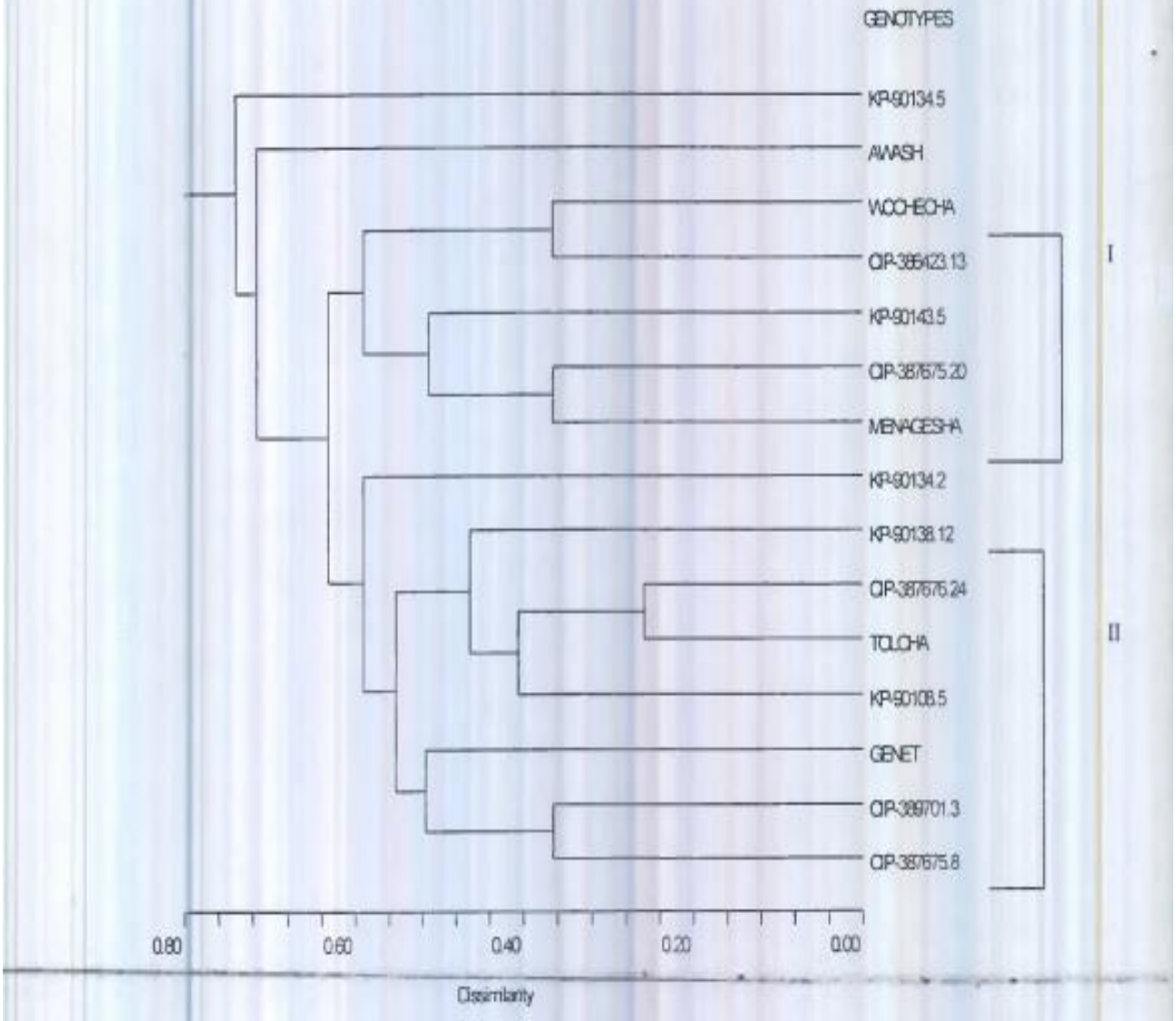

Fig 1. Dendrogram constructed with Unweighted Pair Group Method Using Arithmetic Averages (UPGMA) clustering algorithm from the pair wise matrix of genetic distance among 15 potato genotypes.