CHAPTER IV
RESULTS AND DISCUSSION

Analysis of genetic diversity is very important in plant breeding programs and Molecular markers offer an approach to unveil the genetic diversity among genotypes based on nucleic acid polymorphisms. In this study, SRAP and EST-SSR markers were simultaneously used to investigate genetic diversity of 47 *H. tuberosus* genotypes

4.1 Genetic diversity of *H. tuberosus* genotypes using SRAP markers

A total of 9 different combinations of primers were employed using 3 forward and 3 reverse primers for amplification of DNA fragments from a diverse collection of 47 *H. tuberosus* genotype (Fig. 4.1-4.9).

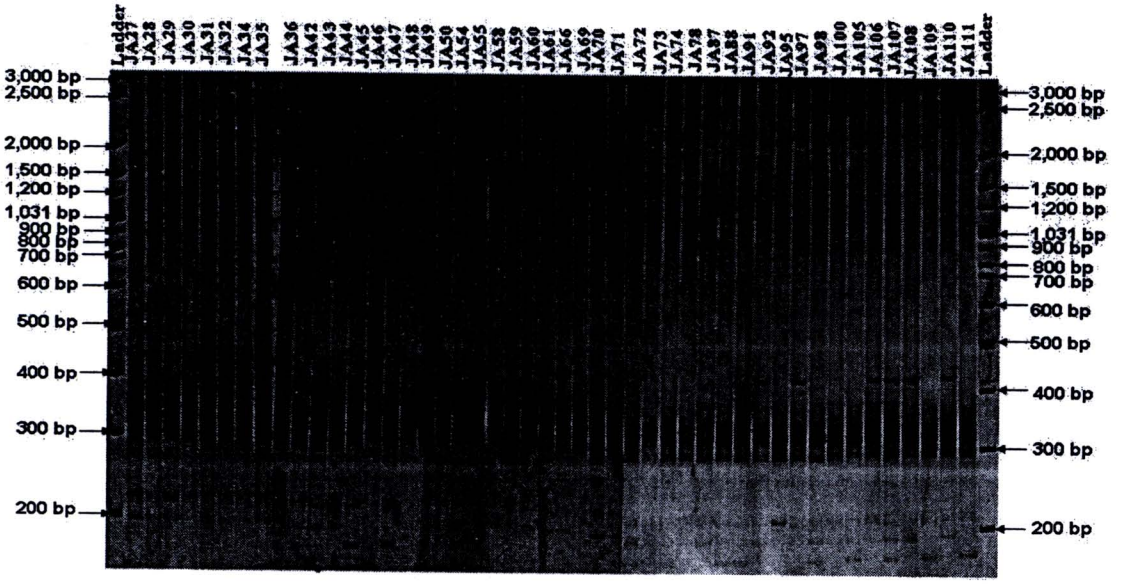


Figure 4.1 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME2-EM5 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

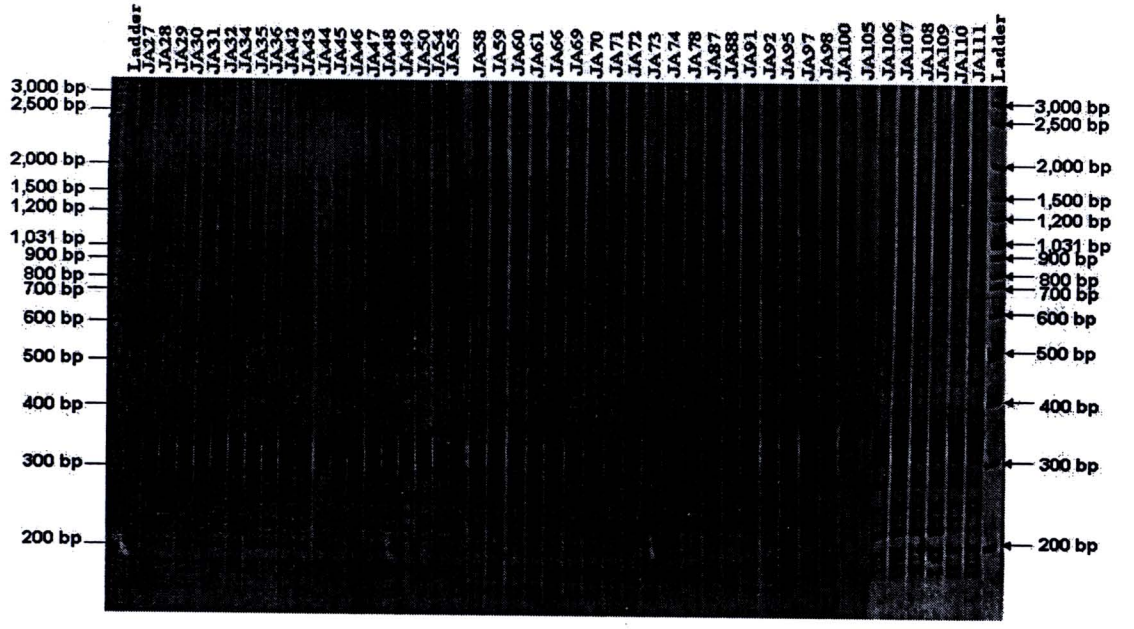


Figure 4.2 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME2-EM6 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

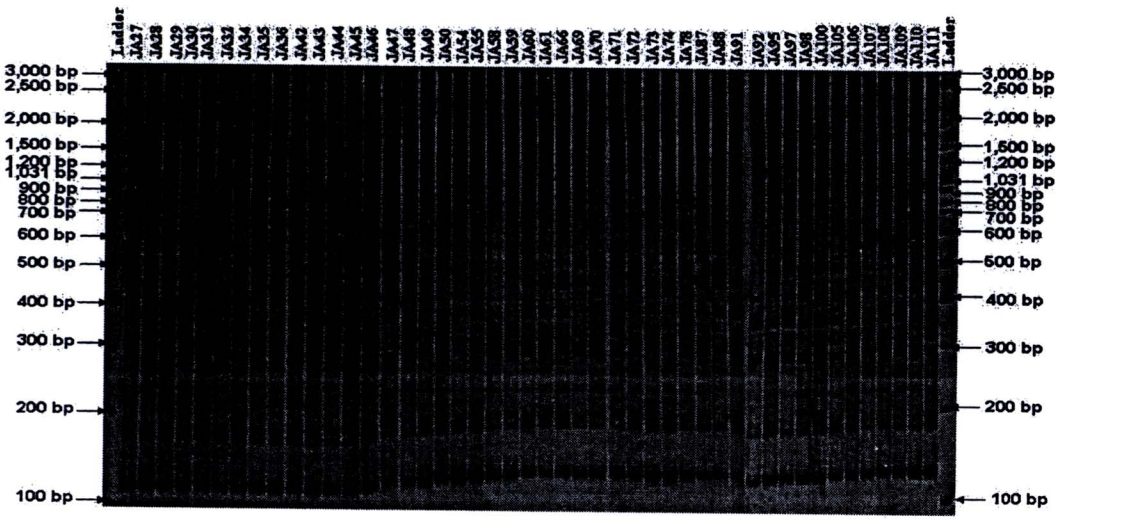


Figure 4.3 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME2-EM8 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

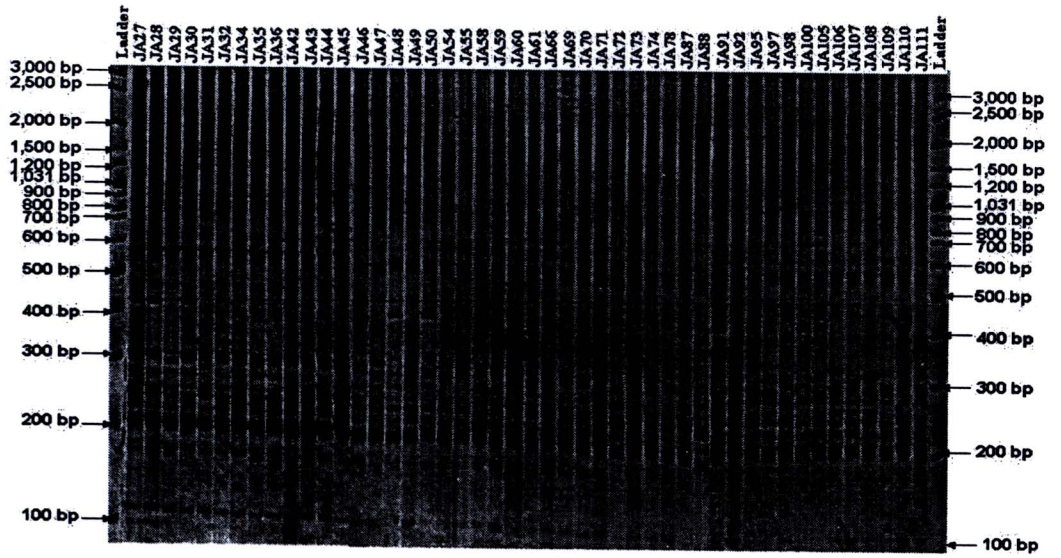


Figure 4.4 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME5-EM5 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

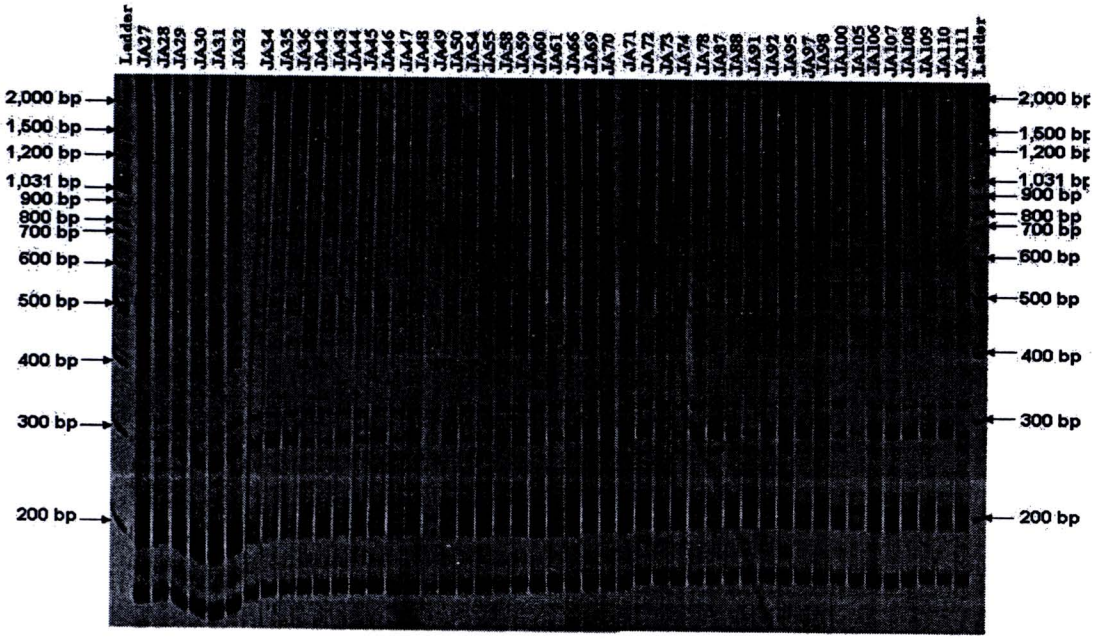


Figure 4.5 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME5-EM6 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

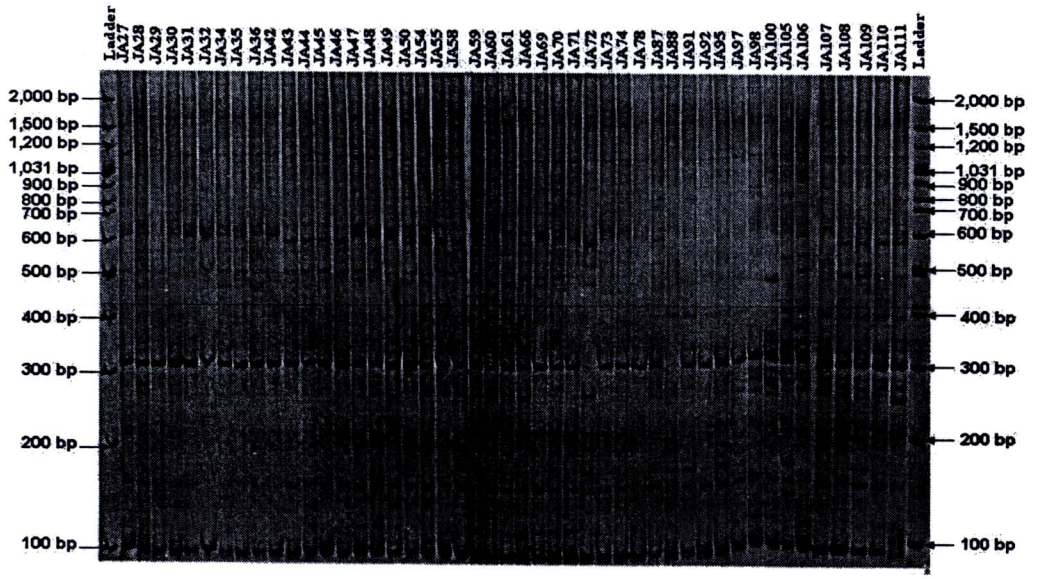


Figure 4.6 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME5-EM8 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

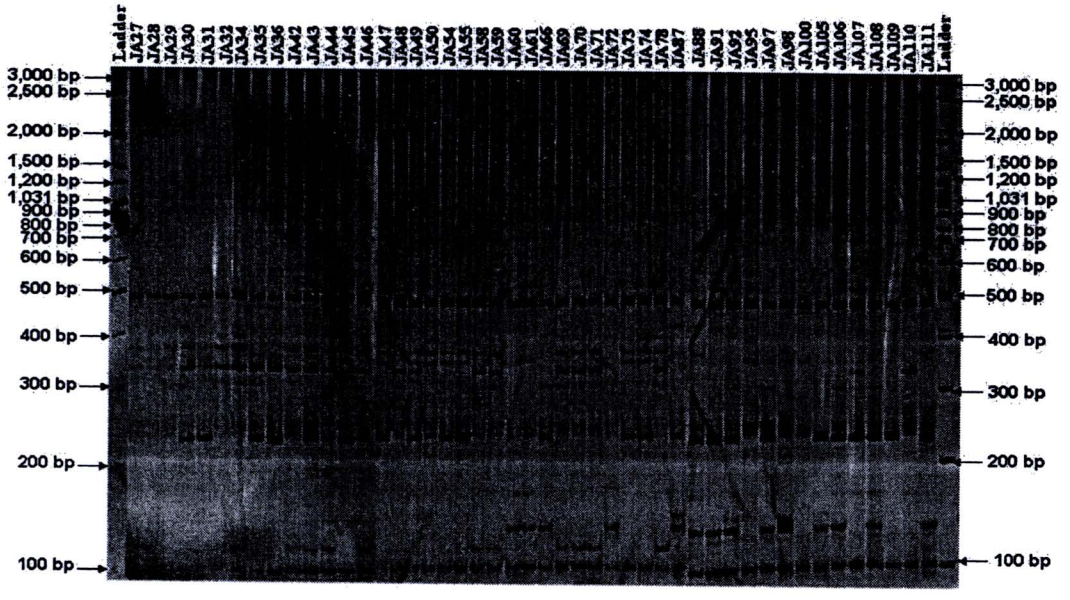


Figure 4.7 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME7-EM5 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

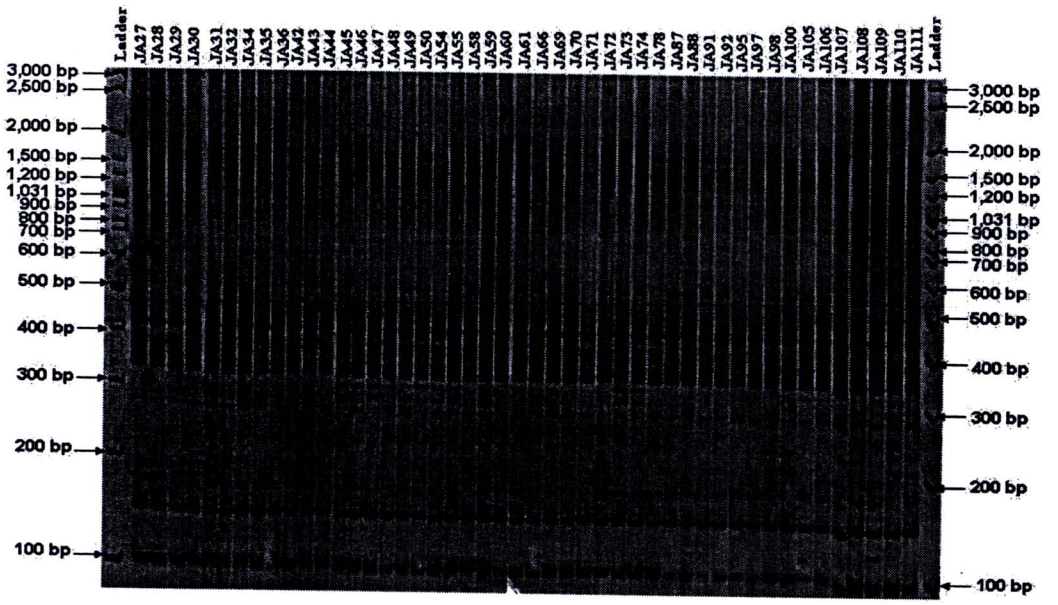


Figure 4.8 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME7-EM6 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

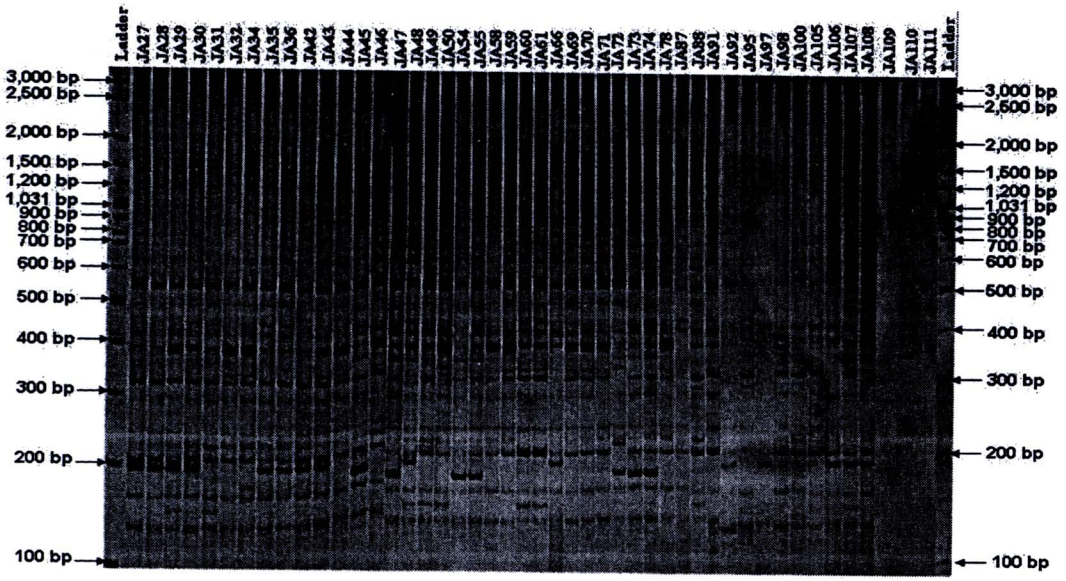


Figure 4.9 The extent of SRAP polymorphism observed among 47 *H. tuberosus* genotypes, revealed by ME7-EM8 primers on 10% polyacrylamide gel; Ladder: 100 bp DNA ladder plus, Vivantis

Table 4.1 Parameters of 47 *H. tuberosus* genotypes amplified with 9 SRAP primer combinations

Primer	Total number of bands	Number of polymorphic bands (PPB %)	PIC	Mean band per primer	Mean band frequency per primer	Size of PCR fragment (bp)
ME2-EM5	19	19 (100%)	0.94	11.38	0.60	195-2,000
ME2-EM6	37	36 (97.3%)	0.96	18.74	0.51	180-2,000
ME2-EM8	20	18 (90.0%)	0.95	15.91	0.80	120-1,031
ME5-EM5	24	19 (79.2%)	0.95	17.19	0.72	200-2,000
ME5-EM6	20	16 (80.0%)	0.95	16.55	0.83	120-2,000
ME5-EM8	33	31 (94.0%)	0.96	19.57	0.59	95-1,750
ME7-EM5	12	10 (83.0%)	0.90	6.51	0.54	100-700
ME7-EM6	26	14 (54.0%)	0.96	18.62	0.72	100-1,750
ME7-EM8	28	27 (96.0%)	0.95	16.23	0.58	130-620

Note: PPB = Percent of polymorphic bands; PIC Polymorphic information content

A total of 219 distinct bands were observed, among which 190 (86.8%) were polymorphic bands. The size of scored bands ranged from 95 to 2,000 bp. The number of scored loci amplified per primer combination varied from 12 to 37 with an average of 24.3. However, there was no specific band for any *H. tuberosus* genotypes observed from this study. The number of scored polymorphic bands per primer combination varied from 10 to 36 with an average of 21.1. Percentage polymorphism ranged from 54.0% (ME7-EM6), to a maximum of 100.0% (ME2-EM5), with an average of 86.0%. The various polymorphism of using SRAP primer combination have been found; with the same using ME7-EM6 primer combination; Ferriol et al., 2003 have reported 71.4% and 57.1% polymorphism in the *Cucurbita pepo* and *Cucurbita ovifera,* respectively while Zeafizadeh and Goliev 2009 have found 77.5% polymorphism in *Triticum durum.*

The Polymorphic information content (PIC) ranged from 0.90 to 0.96 with an average 0.95, revealed that all primer combinations had a high polymorphic DNA fragment of 47 *H. tuberosus* genotypes . Mean band per primer and mean band frequency per primer represented an locus efficiency for indicating the genetic difference were ranged from 6.51 to 18.74 and 0.51 to 0.83, respectively (Table 4.1).

SRAP primers have been used with other plants, Ferriol el. al. 2003 have reported the polymorphism of using SRAP primers in the *Cucurbita pepo* and *Cucurbita ovifera* ranged from 25% to 100% with an average 72.7%. Zeafizadeh and Goliev 2009 also found the polymorphism ranged from 25% to 100% with an average 56.4%. Comparing with other techniques, SRAP markers also provided high levels of polymorphism. Using RAPD-PCR and ISSR-PCR with 147 *H. tuberosus* genotype from Plant Gene Resources of Canada, % polymorphism was 88.71% and 93.22 %, respectively (Sudarat Khampa, 2008).

The dendrograms were developed through UPGMA, WPGMA, Single Linkage and Complete Linkage clustering methods over the original matrices calculated from Simple Matching, Dice and Jaccard similarity coefficients. In the calculation of dendrograms, NTSYS Pc 2.1 program was used. The cophenetic correlations from the genetic similarity matrices are shown in Table 4.2. The UPGMA clustering for Simple Matching was given a highest cophenetic correlation as 0.88262 and complete linkage clustering for Simple Matching was yielded a lowest correlation value as 0.78276. UPGMA clustering method with Simple Matching is evaluated as a convenient combination for detecting the genetic relationships between 47 *H. tuberosus* genotypes.

Table 4.2 Genetic similarity of 47 *H. tuberosus* genotypes from SRAP technique and the cophenetic correlation coefficients of matices obtained from cophenetic values

Clustering/simelarity	Simple matching	Jaccard	Dice
UPGMA	0.88262	0.88056	0.86930
WPGMA	0.86841	0.86788	0.85533
Single Linkage	0.81504	0.82972	0.82405
Complete Linkage	0.78276	0.81726	0.80355

Table 4.3 Consensus fork index among the dendrograms (UPGMA) produced by similarity coefficients among 47 *H. tuberosus* genotypes by SRAP technique.

Clustering/simelarity	Simple matching	Jaccard	Dice
Simple matching	*******		
Jaccard	0.88889	*******	
Dice	0.88889	1.00000	*******

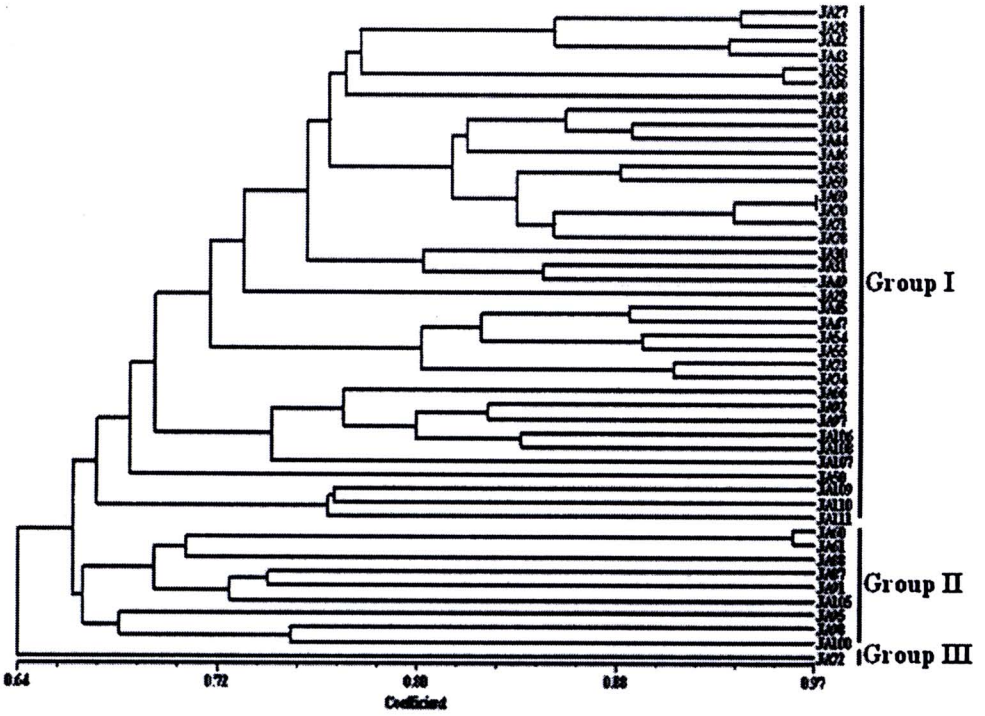


Figure 4.10 The dendrogram based on the SRAP primer data showing the genetic relationships among 47 *H.tuberosus* genotypes based on the UPGMA clustering for Simple Matching coefficients

The dendrograms derived from this combination were given in Fig. 4.10 (the rest dendrograms derived from the other combinations of genetic similarity coefficients and different clustering methods were place in Appendix B, a total of 47 *H. tuberosus* genotypes were grouped into three clusters (coefficient = 0.67). Group I was a major group contained 37 genotypes; JA27, JA28, JA29, JA30, JA31, JA32, JA34, JA35, JA36, JA42, JA43, JA44, JA45, JA46, JA47, JA48, JA49, JA50, JA54, JA55, JA58, JA59, JA66, JA69, JA70, JA71, JA73, JA74, JA78, JA92, JA97, JA106, JA107, JA108, JA109, JA110 and JA111, Group II contained 9 genotypes; JA60, JA61, JA87, JA88, JA91, JA95, JA98, JA105and JA100 and Group III had only one genotype; JA72. All originated cultivars in Canada were group together in Group I. Almost originated cultivars in USA were group together in Group II, excepted JA72 that was separated to Group III. For the

originated cultivars in USSR or France were distributed within Group I and II. The genotype groupings revealed agreement of the origin of *H. tuberosus* in Canada and USA. By RAPD technique, *H. tuberosus* origins of Canada were separated clearly from others (Sudarat Khampa, 2008). In a study of using only three SRAP primer pairs 58 polymorphic bands were generated from the 134 isolations of *A. morbosa*. Population of *A. morbora* collection from different region in Canada and United State show various degrees of genetic variabilities (Zhang et al., 2005).

The Simple matching similarity values ranged from 0.58-0.97. The closest samples have been found between JA69 and JA70 (S.I. = 0.97) and the most distant ones (S.I. = 0.58) have been found between JA32 and JA61.

4.2 Genetic diversity of *H. tuberosus* based on EST-SSR markers

4.2.1 Searching Jerusalem artichoke EST databases for resistance gene homologs

A total of 69 EST sequences homologous to disease resistance genes were retrieved from 21,994 Jerusalem artichoke transcript assemblies in the Compositae Genome Program Database (CGPdb;http://cgpdb.ucdavis.edu/cgpdb2/). The candidate EST sequences were substantiated annotation by BLASTX analysis against the NCBI Protein Database to genes involved in resistance identified in *Helianthus* or *Arabidopsis*. All of 69 ESTs presented homologous to *R*-genes with BLASTX similarities ranged from 28 to 88%. Of these 20, 25, 20 and 4 ESTs were NBS-LRR, CC-NBS (-LRR), TIR-NBS-LRR and *R* genes-mediated resistance signaling protein (EDS1 and RAR1) classed genes (Appendix A). Forty-one and 28 ESTs were homologous to *H. annuus* and *A. thaliana* or *A. arenosa* R-gene respectively. *H. tuberosus* has been a historically important source of *R*-genes transferred into *H. annuus* through backcross breeding (Vranceanu et al.,1980; Parker and Riches 1993). Several clusters of NBS-LRR loci identified from ESTs and RSAs are not linked to *R*-genes identified from resistance

phenotypes, but could be the source of as yet undiscovered *R*-genes (Radwan et al., 2008).

4.2.2 Sequence and genetic diversity analysis

To determine the conserved motifs of all 69 ESTs, the amino acid sequences translated from these ESTs was scanned against the nonredundant (nr) protein database by BLASTP searches for the presence of known domains. 49 of 69 translated amino acid sequences were able to identify their conserved motifs which spanned TIR, NBS, LRR or combinations. Moreover, Lipase3 and CHORD which are conserved motifs in R genes-mediated resistance signaling protein were also found (Appendix A).

A multiple alignment analysis was performed using ClustalW for nucleotide sequences and amino acid translations of the EST sequences. Phylogenetic analyses were performed on amino acid translation of the EST sequence data using the neighbor-joining method as implemented in ClustalW with default analysis parameters (Fig. 4.11). The translated amino acid sequences were clustered into 8 groups (GI-VIII). GI were the combination of NBS-LRR, CC-NBS-LRR, and TIR-NBS-LRR classes. GII were comprised of TIR-NBS-LRR class and R genes-mediated resistance signaling protein. GIII were solely CC-NBS-LRR class. GIV could be divided into three subgroups, GIVa mostly contained CC-NBS-LRR class with a few of NBS-LRR and TIR-NBS-LRR genes. GIVb, GV and GVII comprised of NBS-LRR and TIR-NBS-LRR classed genes. GIVc and GVIII were only one CC-NBS-LRR sequence. GVI contained mostly CC-NBS-LRR class and only one NBS-LRR class. Those ESTs were further searched for microsatellites.

Table 4.4 Summary of 69 annotated ESTs and conserved domains

R-gene class	No. of ESTs	No. of ESTs contained conserved domains	No. of ESTs derived SSR	No. of SSRs	SSR motif Unit			
					Tri-	Tetra-	Penta-	Hexa-
NBS-LRR	20	17	12	20	-	1	15	4
CC-NBS (-LRR)	25	16	15	31	5	2	15	9
TIR-NBS-LRR	20	13	12	16	1	1	13	1
EDS1	3	2	2	2	-	-	2	-
RAR1	1	1	-	-	-	-	-	-
Total	69	49	41	69	6	4	45	14

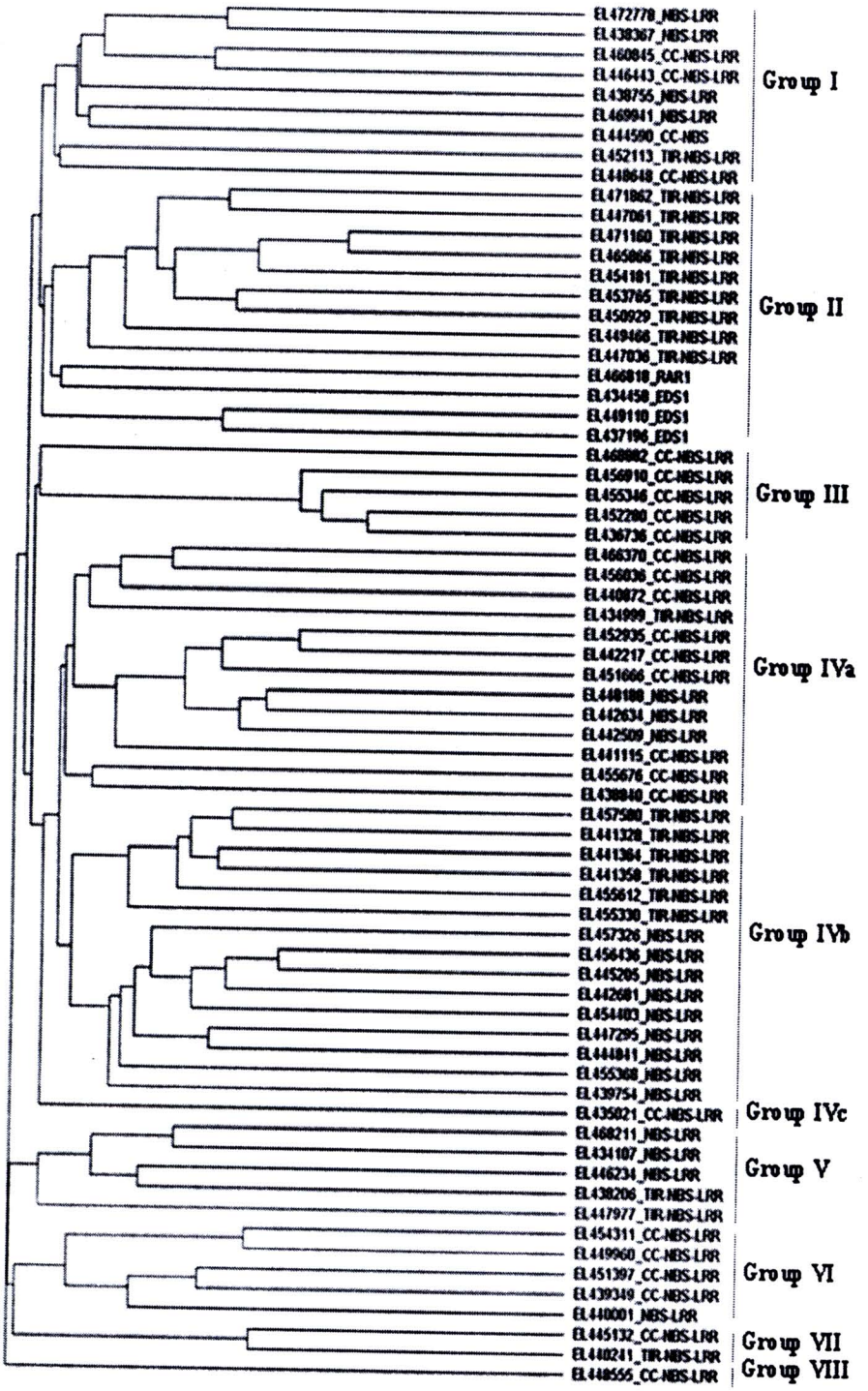


Figure 4.11 Neighbor-joining tree produced from an analysis of the similarity matrix of the amino acid sequences translated from 69 *H. tuberosus* ESTs.

4.2.3 Searching for microsatellites in ESTs

Searching for SSRs presented in 69 ESTs of *H. tuberosus* from Compositae Genome Project Database, collected in FASTA format, and then searched for SSRs using WebSat was carried out. A total of 41 ESTs contained SSRs which were 59.4% (Table. 4.5) suggesting an average frequency of SSR as ~1/1.19 kb. The overall frequency average of (redundant) SSRs in plants is 1 per 6.0 kb (Varshney et al., 2005).Various SSR frequency were reported by Morgante et al. (2002) with 2.1, 1.1 and 1.3 per kb for rice, maize and wheat, respectively. However, it may be noted that the frequency, distribution and abundance of SSRs can be highly variable depending on the SSR search criteria, the size of the dataset, and the database-mining tools (Varshney et al., 2005). The pentanucleotide repeat motifs were the most abundant SSRs, accounting for 65.2% followed by 20.3%, 8.7%, and 5.8% for hexa-, tri-and tetra-nucleotide repeats, respectively. Preeya et al. (2010) also found the pentanucleotide repeat motifs 70.01% from searching 9,164 EST sequences of *H. tuberosus* from Compositae Genome Project Database with the length of microsatellites varied from 10-30 bp. Generally, trinucleotide repeats were the most abundant in EST databases derived from a number of plant taxa, including wheat (Gupta and others 2003), barrel medic (Eujayl and others 2004), pine (Chagne and others 2004), and barley (Toth and others 2000), and the higher level of variability of these repeat motifs likely results from the fact that the gain/loss of trinucleotide repeat units has no effect on the protein-coding frame. Like pentanucleotide repeats that were the most one in this study, they also have no effect on their conserved protein-coding frame once occurring of the variation of nucleotide sequence.

Table 4.5 Summary of EST-SSRs indentified

No.	SSR motif	Type of motifs	Motif abundance
1	ACA	3	1
2	AGC	3	1
3	GAA	3	4
4	AAAT	4	1
5	AATC	4	1
6	TGCT	4	1
7	TTAA	4	1
8	AAAAT	5	1
9	AAAGA	5	1
10	AAAGC	5	1
11	AAGTT	5	2
12	ACAAA	5	1
13	ACCGT	5	1
14	ACTTG	5	1
15	AGAGA	5	1
16	AGCTC	5	1
17	AGCTT	5	1
18	AGGGT	5	1
19	ATATC	5	1
20	ATCTA	5	1
21	ATGAT	5	3
22	ATTGG	5	1
23	CAAAA	5	1
24	CAAAT	5	1
25	CACCA	5	1
26	CGGTT	5	1
27	CTCAA	5	1
28	CTCGT	5	1
29	CTTGA	5	1
30	GAAAC	5	1
31	GACAT	5	1
32	GCTTA	5	1
33	GTTAT	5	1
34	GTTTG	5	2
35	TACCA	5	1
36	TACTC	5	1
37	TATTT	5	2
38	TCAAT	5	1
39	TCCAA	5	1
40	TGATG	5	3
41	TTCAT	5	1
42	TTGTG	5	1
43	TTTAT	5	1
44	TTTCA	5	1
45	TTTCC	5	1
46	AATTTG	6	1
47	AGATGA	6	4
48	GATACT	6	1
49	GTATTT	6	1
50	GTTGCT	6	1
51	TCAGAT	6	1
52	TCCATG	6	1
53	TCCCAG	6	1
54	TTGTGA	6	1
55	TTTCTC	6	1
56	TTTGCA	6	1

4.2.4 Development of potentially functional EST–SSR Markers

The 27 non-redundant SSR-ESTs (comprising consensus sequences for 64 clusters and 267 singleton SSR-ESTs) were used for primer designing. Of the 27 potential EST-SSRs, 18 contained pentanucleotide repeat motifs, 7 contained hexanucleotide repeat motifs, 1 contained trinucleotide repeat motifs, and the last one contained both hexa- and trinucleotide repeat motifs. Moreover 23 primers were designed from ESTs contained R-gene conserved motif. The details of primer sequences and expected product size with SSR motifs are described in Table 4.6.

Table 4.6 Oligonucleotide primers used to amplify resistance gene analogues

Accession numbers	Class of R gene	Conserved domains	SSR motif	Primer name	Sequence (5'-3')	Expected size (bp)	Amplified fragments
Heli_tube.EL436736	CC-NBS-LRR	NBS	(TGATG)2	HEL436736	F: GAA CCA TTG TGT GGA AGA AT R: TTG ACA GTT TGT GGC AAT AG	150	yes
Heli_tube.EL438367	NBS-LRR	NBS	(CTCAA)2	WNKK	F: GCT TAG AGA TCT CAG TCA GC R: GGA GAA GTA GGC TCT TAC C	120	yes
Heli_tube.EL438755	NBS-LRR	NBS	(TCCATG)2	WTNP	F: CTC TAT TGT CGG TAT GGG TG R: GCA CCA TGT TTC TAT GTG CC	196	yes
Heli_tube.EL438840	CC-NBS-LRR	none	(ATATC)2	JNKK	F: CGC CAG CAA GTC AAC TTC AC R: CGG ATG GTC AAT GCT TGG AG	186	yes
Heli_tube.EL440001	NBS-LRR	NBS	(GTTAT)2	HEL440001	F: CTA TTT TGG TGC TTT TCC G R: CCT TCT TTT CGC AAT CAG TA	172	yes
Heli_tube.EL440872	CC-NBS-LRR	LRR	(TCAAT)2	HEL440872	F: CGG GGC ACT TAA AAG AAT A R: GTG ACA TAG AAG GGG AGT CA	155	yes
Heli_tube.EL441115	NBS-LRR	LRR	(ATTGG)2	YNKK	F: GGA GGT CCT AAC TGA GCT C R: CTT AAG TAA CGG TAG CTG CC	184	yes
Heli_tube.EL444590	CC-NBS	NBS	(AGATGA)2, (GAA)4, (AGATGA)3, (GAA)10	CAN	F: CCT TGG TCA TCG CTT AGA G R: GAG GGC AAT TGT CTC AAA GAG	256	no
Heli_tube.EL444841	NBS-LRR	LRR	(TTGTGA)2	HEL444841	F: TGA ACT CCG TCA TGT GAA TA R: CAA GTG ATG GGT GAA TCT CT	116	yes
Heli_tube.EL446443	NBS-LRR	NBS	(TCCCAG)2	HEL446443	F: ATC ACC AGC AAG GTA AGA AA R: TCC TCT TTC ACA CTC AAA CC	146	yes

Table 4.6 Oligonucleotide primers used to amplify resistance gene analogues (cont.)

Accession numbers	Class of R gene	Conserved domains	SSR motif	Primer name	Sequence (5'-3')	Expected size (bp)	Amplified fragments
Heli_tube.EL447036	TIR-NBS-LRR	TIR	(AGCTT)2	HEL447036	F: TCA AGG TTT CTG GTG GTT A R: TGT CAC GAT CTT TCT CCT T	103	yes
Heli_tube.EL448188	NBS-LRR	LRR	(TGATG)2	PPW	F: GCC TTG TTT GGA GGT CAT TC R: GGC ATG ACG AGC TCA AAT G	189	yes
Heli_tube.EL449110	EDS1	Lipase3	(AAAGC)2	HEL449110	F: CTG AGG TGG AGA AAG CAA T R: GAA TAT ACG GTC GCC TAC AA	176	yes
Heli_tube.EL449466	TIR-NBS-LRR	TIR	(GATACT)2	HEL449466	F: ATC GAA CAT CTT TAT GTG GC R: CAA TGA TAG AAG CCC TTG TT	132	yes
Heli_tube.EL449960	CC-NBS-LRR	none	(TTGTG)2, (CGGTT)2	NKK	F: CGT CTT CCT TTT GGG AGA CAG R: CTT GCC TGC TTT GGG AGA AAC	188	yes
Heli_tube.EL450929	TIR-NBS-LRR	TIR	(ACA)4	HEL450929	F: GAA CCC ATC AGA AGT GAG AA R: AAG TGC TTT CCT CCA AGA TT	103	yes
Heli_tube.EL451666	CC-NBS-LRR	LRR	(TCAGAT)2	HEL451666	F: CAA GGA AGT TAC CAG AAT CA R: CAA TTT ACG CAA TCT CCT AC	152	yes
Heli_tube.EL452935	CC-NBS-LRR	LRR	(CAAAT)2	HEL452935	F: ACT AAA ATA GGG GAA TTG GG R: CTC TTC ATC GTG TTC ACT CA	177	yes
Heli_tube.EL453765	TIR-NBS-LRR	TIR	(AAAAT)2	HEL453765	F: GGAAATATGGCTTTGCTTAC R: CCACCTCGACTGAGAAGTAG	180	yes
Heli_tube.EL454403	NBS-LRR	LRR	(CACCA)2	HEL454403	F: TCT TGA TCT CCG TCA TTC A R: GTC AGG TTC ACG AAA ACA A	153	yes
Heli_tube.EL455346	CC-NBS-LRR	NBS	(ATGAT)2	HEL455346	F: CCA TTG TGA GGA AGA ATT TG R: CAA TAC ATG GGC GTT GAC	160	yes

Table 4.6 Oligonucleotide primers used to amplify resistance gene analogues (cont.)

Accession numbers	Class of R gene	Conserved domains	SSR motif	Primer name	Sequence (5'-3')	Expected size (bp)	Amplified fragments
Heli_tube.EL457326	NBS-LRR	LRR	(GACAT)2	HEL457326	F: AGG ACA CCG GAT TTT AGT G R: TCT CCA ACT TCT TCA TCC TG	178	no
Heli_tube.EL457580	TIR-NBS-LRR	none	(TTTCC)2, (ACTTG)2	TNRA	F: GGG TTC TGG TGG ACA GAG TC R: GGA GAC ACT CAA TGC GTT TAA G	251	no
Heli_tube.EL465866	TIR-NBS-LRR	TIR	(TATTT)2	HEL465866	F: CGC CGT TAT TGT ATT CTC TC R: CTT CTC CAT ATT TCC CCT TT	164	no
Heli_tube.EL466370	CC-NBS-LRR	none	(TTTCTC)2, (GTTGCT)2	CNLA	F: TTT CTT CAT TCT GCA TCT TGG A R: CTG GTC AGG CCA ATG AGT TT	152	no
Heli_tube.EL471160	TIR-NBS-LRR	TIR	(TATTT)2	HEL471160	F: AAC TAT GCG GAT TCA TCT TG R: TGG CAG TTT ACT CTT TAC CC	180	yes
Heli_tube.EL472778	NBS-LRR	NBS	(TTTGCA)2	NLA	F: TGA ACT CGG ATG CTT ATG TGA A R: CCA AAG GCA GTC CTT GAC AT	143	yes

4.2.5 EST-SSR marker validation

A total of 27 designed EST-SSR primer pairs were used to amplify the DNA of *H. tuberous* samples. Of these, 22 (81%) primer pairs amplified the expected size of amplicons with considerable polymorphism (Fig. 4.12 and Table 4.6). The details of primer sequences, R-gene class annotation of the corresponding ESTs, SSR sequences and amplified capabilities among the tested *H. tuberous* were listed in Table 4.6. Five EST-SSR primers (CAN, CNLA, TNRA, HEL457326, HEL465866) could not successful amplify.

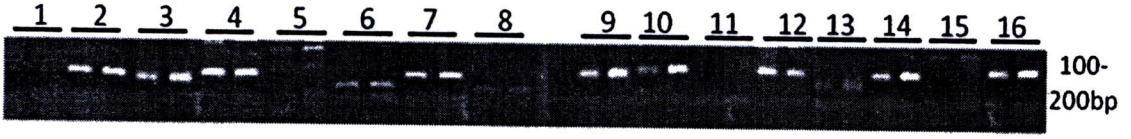


Figure 4.12 PCR amplification of F_1 HEL65xJA6 genomic DNA using EST-SSR primers as follows: (1) HEL457326, (2) HEL455346, (3) HEL446443, (4) HEL452935, (5) HEL471160, (6) HEL444841, (7) HEL454403, (8) HEL447036, (9) HEL451666, (10) HEL449110, (11) HEL465866, (12) HEL436736, (13) HEL449466, (14) HEL440001, (15) HEL440872, (16) HEL453765.

4.2.6 Gene isolation and sequence analysis

Genomic DNA was extracted from plant materials using modified methods of Tai and Tanksley (1990). NKK WNKK and PPW homologues were cloned using primers as indicated in section 3.4.6. The primers were designed based on alignments of the R-gene nucleotide sequences from NCBI (http://www.ncbi. nlm.nih.gov/) and the Compositae Genome Project Database (CGPDB; http://compgenomics.ucdavis.edu/compositae_index.php). The amplified fragments were excised from gels and subsequently cloned into pGem-T easy vector (Promega). The positive cloned which carried amplified fragments were screened using an antibiotic-resistance gene and an advantage of intracistronic á-complementation to regenerate â-galactosidase activity. And then, the colonies PCR were employed. The results of colonies PCR was shown on Fig. 4.13

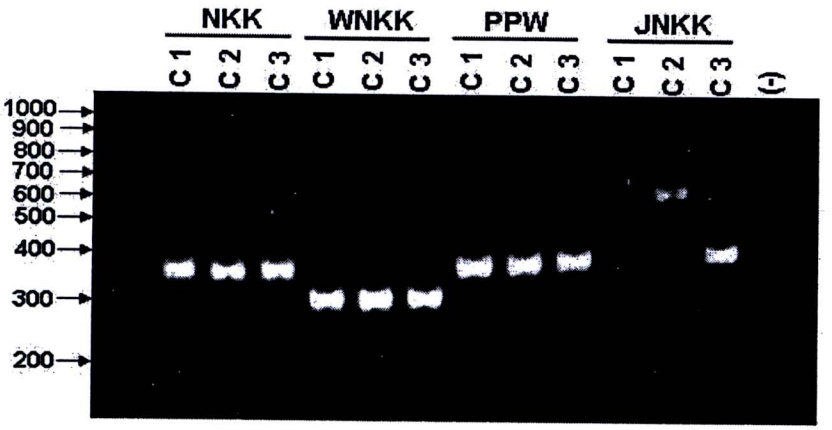


Figure 4.13 Screening the positive colonies of NKK, WNKK PPW and JNKK, revealed by 1% agarose gel, the number indicates clone accession number. (-): Negative control

The amplified PCR fragments were excised from 1.0% agarose gel and then sequenced at the Biomolecular Analysis Service Unit, Department of Biochemistry, Faculty of Medicine, Khon Kaen University. The analysis for similarity to the sequence and the non-redundant nucleotides databases at NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov) using the BLASTN program was performed. The 128, 95 and 91 bp partial fragments of *NKK* (EL449960), *WNKK* (EL438367) and *PPW* (EL448188), respectively were amplified from *H. tuberosus* cv JA 102. The *NKK* *WNKK* and *PPW* shared 92 %, 98 % and 89 % similarity at the nucleotide level with the gene from CGPDB database accession EL449960, EL448188 and EL438367, respectively (Fig. 4.14 -4.16). From the alignment result, the isolation of R-gene on *H. tuberosus* using the specific primers designed from EST-database was confirmed.

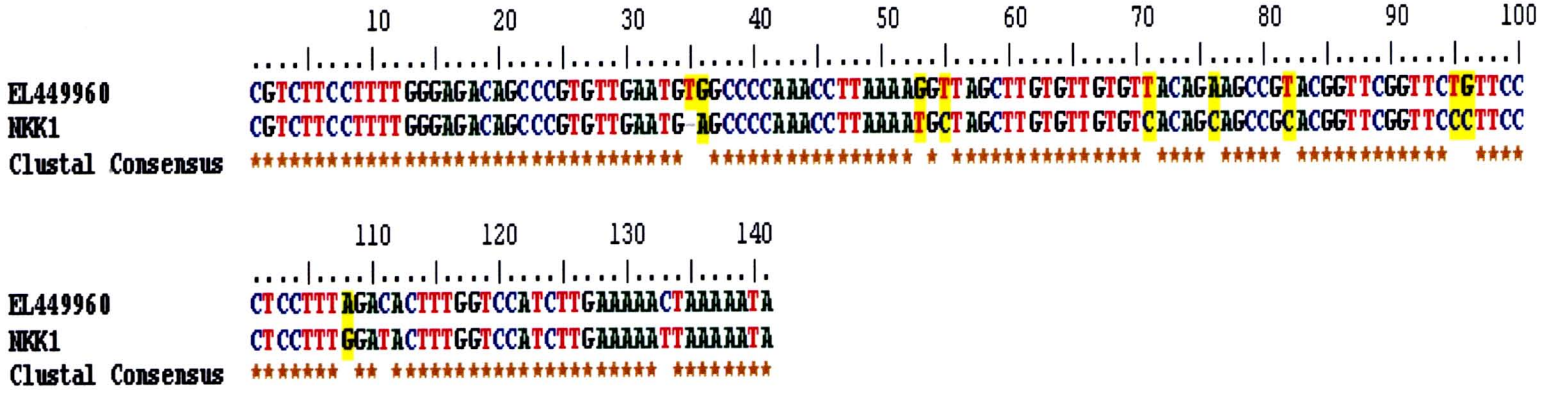


Figure 4.14 Comparison of the nucleotide sequence of *NKK* and EL 449960 available in the CGPDB database. Key: Single letters: nucleotide. "*": identical.

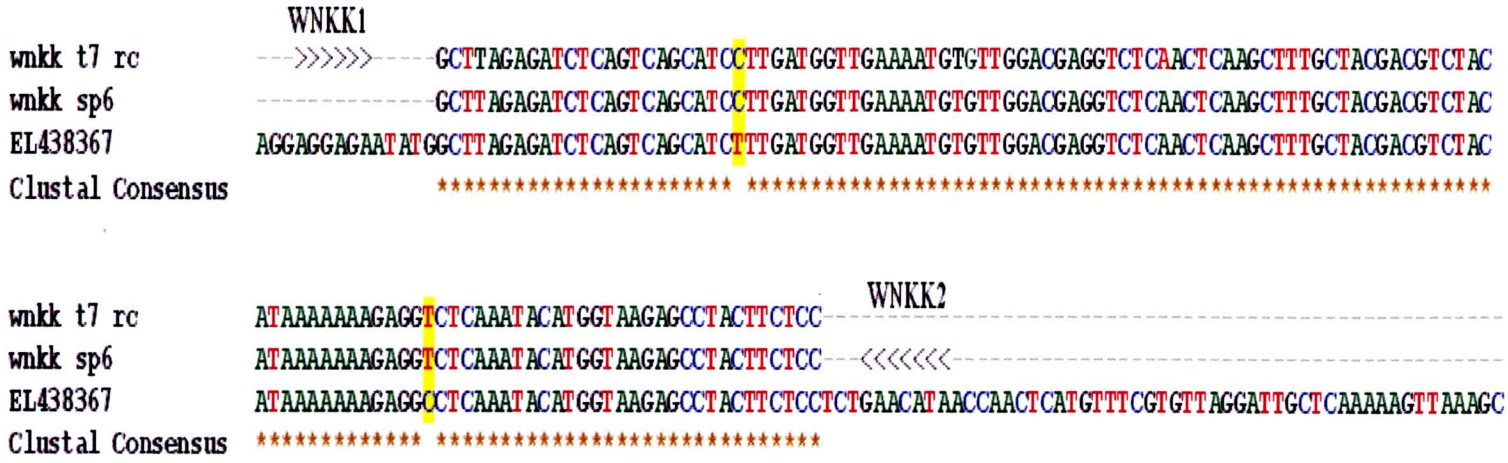


Figure 4.15 Comparison of the nucleotide sequence of *WNKK* and EL 438367 available in the CGPDB database. Key: Single letters: nucleotide. "*": identical.

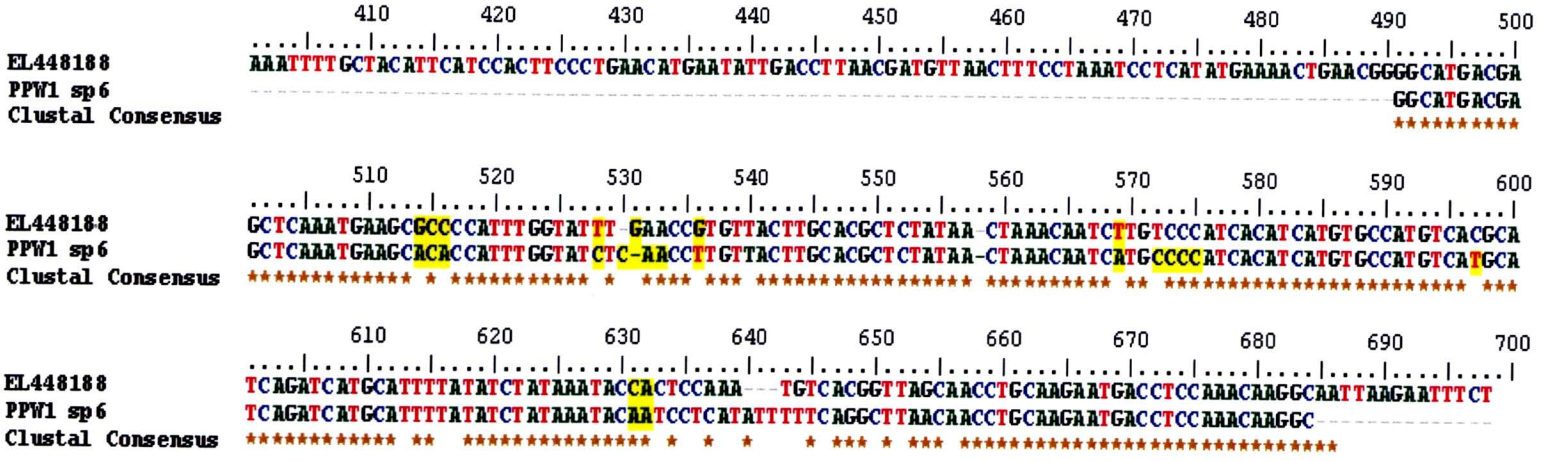


Figure 4.16 Comparison of the nucleotide sequence of *PPW* and EL 448188 available in the database. Key: Single letters: nucleotide. "*": identical.

4.2.7 Genetic divesity analysis

To ascertain the potential of EST-SSR markers in genetic diversity analysis of *H. tuberosus* germplasm, the six EST-SSR primers were used to preliminarily amplify the DNA of 3 *H. tuberosus* genotypes. The appropriated anealing temperatures for PCR were tried. The annealing temperature at 55 °C gave more specific bands than 53 °C (data not show). Most of the primers studied could amplify their corresponding genes to the expected sizes (Fig. 4.17, Table 4.7). Only WTNP amplified larger fragments than expected, reflecting the possible presence of introns within the genomic DNA sequence, and suggests that the presence of long introns between the sequences homologous to the primers in the genomic DNA may explain the lack of amplification in some of the five pairs that did not amplify in part 4.2.5.

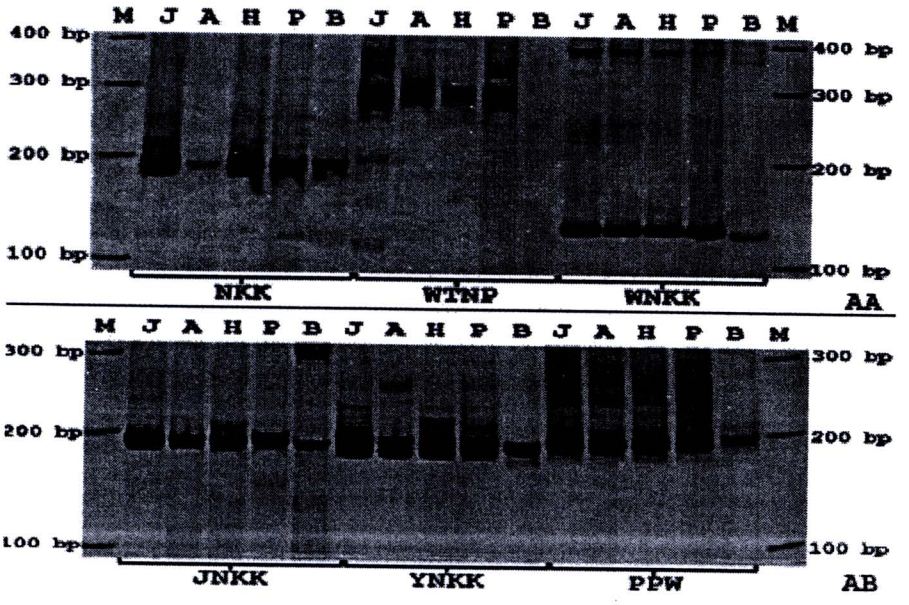
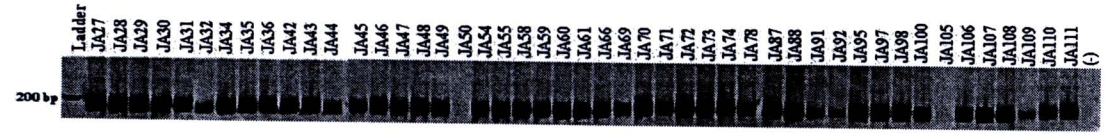


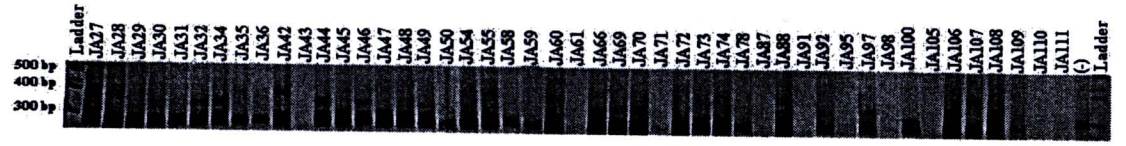
Figure 4.17 PCR products amplified with EST-SSR primers (NKK, WTNP, WNKK, JNKK, YNKK and PPW) on *H. tuberosus* genotypes; 'JA6' (J), 'AMES2747' (A),'HEL 308' (H), 'PI547237' (P), and *Tithonia diversifolia* (B), M : 100 bp Ladder plus, (Vivantis).

Consequently, the six EST-SSR primer pairs were used to preliminarily investigate genetic diversity of 47 *H. tuberosus* genotypes. These primer sets were given scorable bands with of a total 13 alleles (Fig. 4.18 and Table 4.7). Maximum 5 alleles were observed from WNTP primer. Minimum 1 allele was observed from NKK, WNKK and JNKK, which were monomorphic alleles. Of these, 9 alleles were polymorphic bands (69.23%). The PIC values ranged from 0 to 0.77. Mean band per primer and mean band frequency per primer were ranged from 1.0 to 2.38 and 0.48 to 1.0, respectively (Table 4.7). Since, the EST-SSR markers originated from highly conserved genomic regions, which may present lower degree of polymorphism compared to microsatellites originating from genomic libraries (Varshney et al., 2005). The 18 amplifiable EST-SSR markers revealed low to medium allelic diversity with PIC values ranging from 0–0.77 for *arabica* and *robusta* genotypes, respectively (Aggarwal et al., 2007). Moreover, the less polymorphism of EST-SSR marker also found in several germplasm characterization and genetic diversity studies (Scott K.D. et al., 2000; Cho, Y.G. et al., 2000; Eujayl I. et al., 2001; Chabane K. et al., (2005)). However, it is noteworthy that for detection of polymorphism, EST-SSRs derived from 30 ESTs were found to be superior to those derived from 50-ESTs (Scott et al., 2000; Holton et al., 2002; Gao et al., 2003; Gupta et al., 2003).

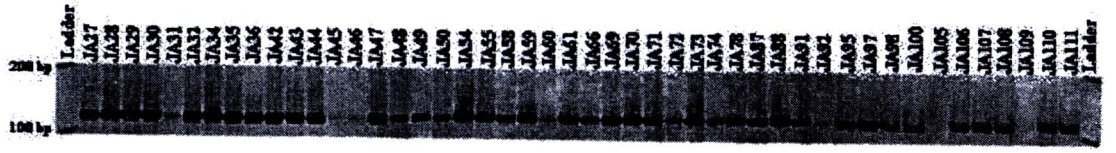
a) NKK



b) WTNP

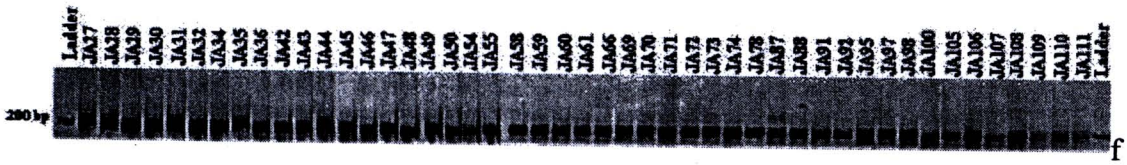


c) WNKK



d) JNKK

e) YNKK



f

) PPW

Figure 4.18 PCR amplification of 47 *H. tuberosus* samples with EST-SSR primers, a) NKK, b) WTNP, c) WNKK, d) JNKK, e) YNKK, f) PPW, Ladder: 100 bp DNA ladder plus, (Vivantis).

Table 4.7 Parameters of 47 genotypes of *H. tuberosus* amplified with 6 EST-SSR primers.

Primer	Total number of bands	Number of polymorphic bands (PPB%)	PIC	Mean band per primer	Mean band frequency per primer	Product size (bp)	Expected size (bp)
NKK	1	0 (0%)	0.00	1.00	1.00	190	188
WTNP	5	5 (100%)	0.77	2.38	0.48	290-320	196
WNKK	1	0 (0%)	0.00	1.00	1.00	120	120
JNKK	1	0 (0%)	0.00	1.00	1.00	200	186
YNKK	2	1 (50%)	0.47	1.62	0.81	195-205	184
PPW	3	3 (100)	0.64	2.09	0.70	195-205	189

Note: PPB = Percent of polymorphic bands; PIC Polymorphic information content

The data obtained from EST-SSR markers were compared with different similarity coefficients and clustering methods (Da Silva Meyer et al., 2004). Using similarity coefficients, Dice, Jaccard and Simple Matching, affects the results of UPGMA, WPGMA, Complete Linkage and Single Linkage clusters (Jackson et al., 1989; Duarte et al., 1999). The dendrogram obtained through these combinations were also compared. The cophenetic correlations from the genetic similarity matrices are shown in Table 4.8. The UPGMA clustering for Simple Matching gave a highest cophenetic correlation as 0.75884 and Single Linkage clustering for Dice was yielded a lowest correlation value as 0.59425 same as the results described from markers in previous section of this study. Comparing the results of clustering method/similarity coefficients, Simple Matching and Jaccard coefficients produced high cophenetic correlations; while, Dice coefficients had lower cophenetic correlations. Because of their highest correlation rate, Simple Matching similarity with UPGMA clustering method is evaluated as a convenient combination for detecting the genetic relationships among 47 genotypes of *H. tuberosus*. Since Dice coefficient and complete linkage clustering method had a lowest correlation value, bear in mind that the combination is an inconvenient composition for detecting the genetic

relationships of this study. Koopman et al. (2001) found that Jaccard similarity with UPGMA yielded highest correlation rate for AFLP data set from Lactuca, s.l. species. In 2010, Seli and Yegenoglu (2010) reported that Dice similarity with UPGMA yielded highest correlation value for RAPD data set from cultivated olives. In this study, simple matching similarity with UPGMA yielded highest correlation value but the differences between simple matching similarity coefficients with UPGMA and WPGMA clustering was not so far (0.75884 and 0.75573, respectively) (Table 4.8).

Table 4.8 Genetic similarity of 47 genotypes of *H. tuberosus* from EST-SSR technique and the cophenetic correlation coefficients of matices obtained from cophenetic values.

Clustering/similarity	Simple matching	Jaccard	Dice
UPGMA	0.75884	0.74796	0.71809
WPGMA	0.75573	0.74023	0.68157
Single Linkage	0.63986	0.62699	0.59425
Complete Linkage	0.67355	0.67394	0.63755

For the evaluation of trees from UPGMA clustering with genetic similarity coefficients, consensus indices were also calculated for each combination of coefficient and UPGMA clustering. The results from Consensus Indices were shown in Table 4.9. The results obtained from consensus indices shown that Consensus fork index was found (CI_C = 1.0000) in Jaccard and Dice coefficients. Simple Matching coefficient had near values with the Dice and Jaccard coefficients (CI_C = 0.88889). This is in line with the findings from dendrograms and shows why obtained different results from the dendrogram formed with Simple Matching coefficient as compared to the others.

Table 4.9 Consensus fork index among the dendrograms (UPGMA) produced by similarity coefficients among 47 genotypes of *H. tuberosus* by EST-SSR markers.

Clustering/similarity	Simple matching	Jaccard	Dice
Simple matching	*******		
Jaccard	0.88889	*******	
Dice	0.88889	1.00000	*******

Based on the dendrogram, a total of 47 *H. tuberosus* were grouped into three major clusters (coefficient = 0.75). Group I and group II contained 21 and 25 genotypes, respectively. Group III comprised only one genotype (JA50). The closest samples based on Simple matching similarity values (S.I. = 1.00) have been found many groups; 1) JA27 and JA 28, 2) JA30, JA32 and JA34, 3) JA69 and JA70, 4) JA42 and JA72, 5) JA88 and JA92, 6) JA35, JA59, JA71 and JA95, 7) JA 73 and JA74, 8) JA61 and JA110, 9) JA91, JA98 and JA111, and 10) JA45, JA47 and JA100 and the most distant ones (S.I. = 0.38) have been found as; 1) JA43 and JA66, and 2) JA69, JA70 and JA87 (Fig. 4.18)

The dendrogram derived from the simple matching similarity coefficient and UPGMA clustering method was given in Fig. 4.18. While, the dendrograms derived from other combinations of genetic similarity coefficients and different clustering methods were place in Appendix B).

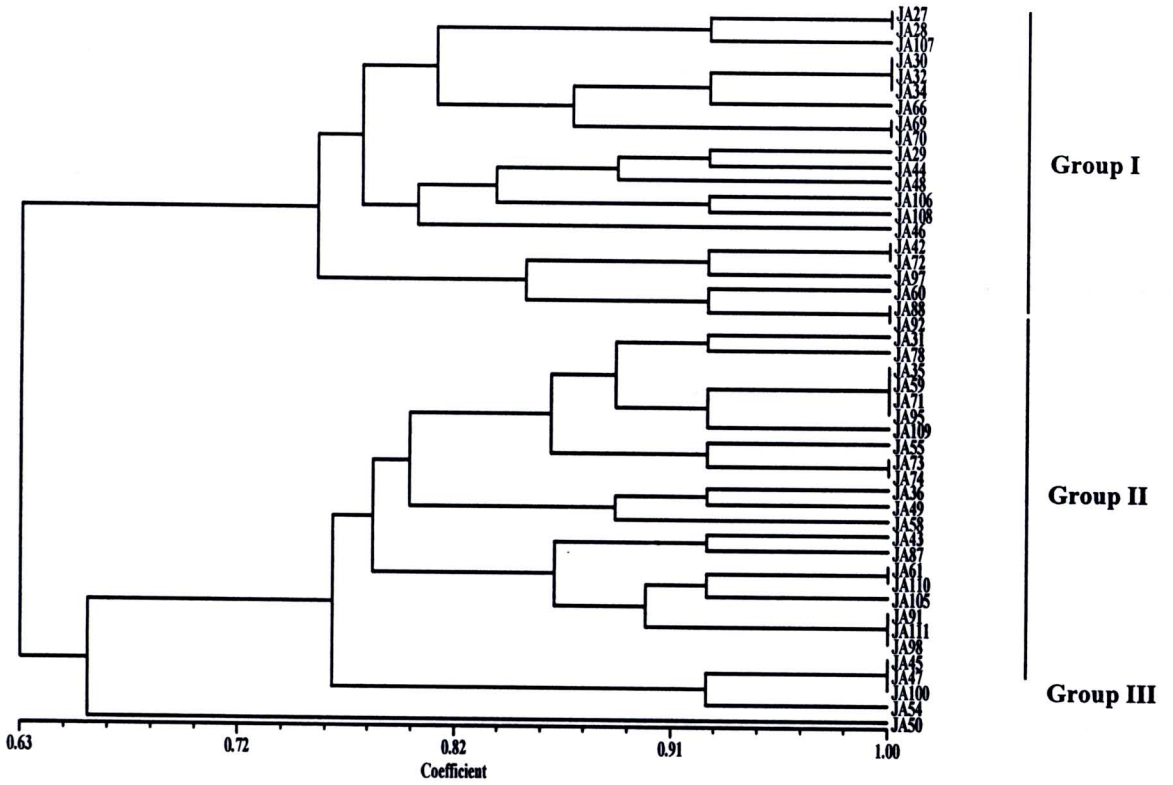


Figure 4.19 UPGMA dendrogram with Simple matching similarity coefficients from the EST-SSR data of 47 *H.tuberosus* genotypes.