MOLECULAR CHARACTERISATION OF NERICA LINES

Contributors: M-N Ndjiondjop, K Semagn, M Cissoko, MP Jones and S McCouch

Unit 1 – Molecular profiling of NERICA lines

As mentioned throughout this compendium, NERICA rices are interspecific inbred progeny derived from crosses between *Oryza sativa* × *O. glaberrima*. This chapter is a report of a study that evaluated 70 BC_2 interspecific lines developed by crossing a tropical japonica variety (WAB 56-104) as the recurrent parent to an *O. glaberrima* variety (CG 14) as the donor parent, followed by the use of anther culture to derive doubled haploids (DH) (26 lines) or eight generations of inbreeding to fix the lines (44 lines). Seven of these BC_2-derived inbred lines have been released as NERICA1–NERICA7. This study examined the relative contribution of each parent and the extent of genetic differences among these 70 sister lines using 130 well-distributed microsatellite markers, which cover 1725 cM of the rice genome. The average proportion of *O. sativa*-recurrent parent genome was 87.4%, while the observed average proportion of *O. glaberrima* donor genome was 6.3%. Non-parental alleles were detected in 83% of the lines but the overall average was 2.2%. Lines that had undergone eight generations of inbreeding in the field contained significantly more non-parental alleles (av. 2.7%) compared to the DH lines (av. 1.3%) that were developed from BC_2 anthers. Using both cluster and principal component analyses, two major groups were detected in these materials. The NERICA varieties (NERICA1 to NERICA7) clustered in one group while the remaining 63 lines clustered in another group, suggesting that the second group may offer significant opportunities for further selection and variety development.

Table 13. Summary of the proportion of genome for 70 NERICA varieties using 130 SSRs

		CG 14	WAB56-104	Heterogenous	Missing	Non parental
All 70 lines	Minimum	0.9	79.0	0.0	0.0	0.0
	Maximum	12.1	94.4	3.4	13.6	5.5
	Mean	6.3	87.4	0.4	3.7	2.2
7 NERICA lines	Mean	8.2	88.2	0.2	0.3	3.0
Other 63 lines	Mean	6.0	87.4	0.4	4.1	2.1
DHs n = 26)	Mean	5.5	87.5	0.5	5.2	1.3
Pedigree lines (n = 44)	Mean	6.7	87.4	0.4	2.8	2.7

Unit 2 – Microsatellites and agronomic traits for assessing genetic relationships among 18 NERICA varieties

Genetic differences and patterns of relationship among the first 18 NERICAs are largely unknown. A total of 102 polymorphic microsatellite markers were used to genotype 18 NERICA varieties. Subsets of seven NERICA varieties (NERICA1 to 7) were further characterized for 10 agronomic traits. The microsatellite data revealed no genetic difference between NERICA8 and 9. The absence of genetic distance and identical SSR haplotype distribution (banding pattern) observed between NERICA8 and 9 is highly likely to be due to lack of molecular difference at the DNA level but the possibility for seed admixture remains to be explored. This study, however, revealed the presence of a wide range of genetic differences among all other NERICA varieties, with the greatest being between NERICA6 and 17. Cluster and principal component analyses of the SSR data revealed distinct separation of NERICA1 to 7 from NERICA8 to 18.

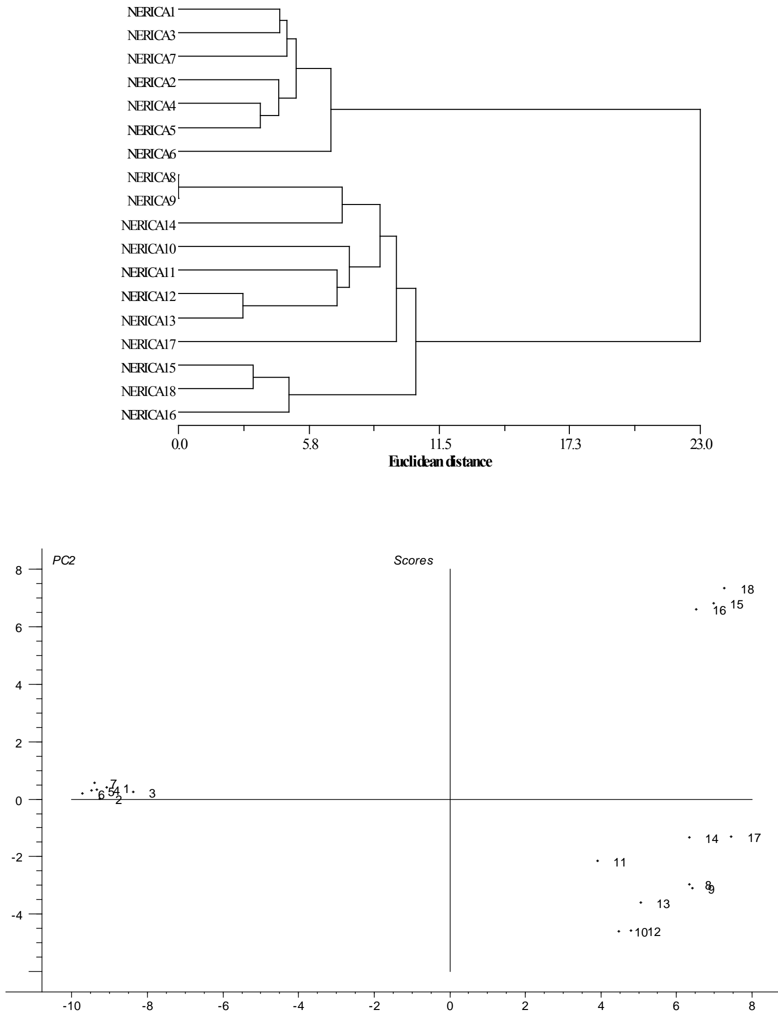

Figure 12. Cluster and principal component (PC) analyses performed using 104 SSR markers: UPGMA phenogram of 18 NERICA varieties based on Euclidean distance matrix, and score plot of NERICAs 1 to 18 from PC analysis. PC1 and PC2 explained 57% and 13%, respectively; numbers in the plot correspond to NERICA1 to 18.

Unit 3 – Molecular profiling of upland NERICA

Background information

The study reported above (Module 4/Unit 2) on the patterns of variation and genetic relationships among the named NERICA1 to NERICA18 concluded there was (i) a lack of genetic difference between NERICA8 and 9, (ii) a relatively wide range of genetic variability among all NERICAs, except NERICA8 and NERICA9, and (iii) a distinct pattern of variation that clustered NERICA1–7 in one group and NERICA8–18 in another (Semagn *et al.*, 2006).

However, the proportion and distribution of *O. glaberrima* introgressions in the larger collection of 70 sibling lines which are of great interest for future breeding efforts were not investigated.

It is noteworthy that NERICA is an extended family of several hundred interspecific progenies, of which 18 have been named by WARDA and released in various African countries.

The following section highlights the outcome of a study that analyzed 70 introgression lines to determine (i) the extent and map position of introgressions from *O. glaberrima*, (ii) the extent of heterozygous loci and non-parental alleles contained in the interspecific inbred lines, and (iii) the genetic relationships among the lines, with particular interest in evaluating the potential breeding value of the lines that have not been released as varieties.

The 70 interspecific lines (Table 14) were developed by crossing WAB 56-104 (*O. sativa* japonica), as the recurrent parent and CG 14 (*O. glaberrima*) as the donor parent. Twenty six of the lines were developed as double haploids (DH) derived from BC_2F_1 plants using anther culture (Guiderdoni *et al.*, 1992). The rapid genetic fixation achieved in the DH lines is expected to retain blocks of genes that would have been lost through conventional inbreeding (due to

sterility factors) and artificial selection (Jones *et al.*, 1997b). The remaining 44 samples were BC_2F_8 lines developed using pedigree selection (Jones *et al.,* 1997b). All 70 lines in this study had gone through two generations of backcrossing followed by either eight generations of selfing or doubling the haploid chromosome numbers and all behaved as inbred introgression lines.

Table 14. Pedigree and donor genome coverage (introgression) of 70 lines developed from WAB 56-104 as recurrent parent and CG 14 as donor parent.

Pedigree	Name or number	Introgression (%)	Pedigree	Name or number	Introgression (%)
WAB450-I-B-P-20-HB	NERICA7	7.3	**WAB450-B-16A2.5**	36	5.4
WAB450-I-B-P-38-HB	NERICA1	6.7	**WAB450-B-16A2.10**	37	6.2
WAB450-I-B-P-28-HB	NERICA3	3.4	WAB450-24-3-P3-1-HB	38	6.6
WAB450-I-B-P-138-HB	4	11.4	**WAB450-B-18A2.4**	39	5.7
WAB450-I-B-P-147-HB	5	5.2	**WAB450-B-19A2.5**	40	5.2
WAB450-12-2-BL1-DV5	6	5.2	**WAB450-B-18A2.6**	41	2.5
WAB450-I-B-P-135-HB	7	5.3	**WAB450-B-3A1.2**	42	8.3
WAB450-I-B-P105-HB	8	6.6	**WAB450-4-1-A22**	43	7.0
WAB450-I-B-P-91-HB	NERICA4	7.5	**WAB450-B-19A3.1**	44	7.4
WAB450-I-B-P-160-HB	NERICA6	12.1	WAB450-I-B-P-129-HB	45	6.4
WAB450-11-1-P31-1-HB	NERICA2	9.5	WAB450-I-B-P-82-2-1	46	7.3
WAB450-I-B-P-33-HB	12	6.3	WAB450-I-B-P-72-3-1	47	6.5
WAB450-11-1-1-P31-HB	NERICA5	11.0	WAB450-I-B-P-6-1-1	48	10.9
WAB450-24-3-2-P18-HB	14	10.2	WAB450-I-B-P-157-1-1	49	2.9
WAB450-11-1-3-P40-HB	15	7.4	WAB450-I-B-P-22-HB	50	8.0
WAB450-I-B-P-32-HB	16	8.1	WAB450-I-B-P-65-1-1	51	8.4
WAB450-12-2-BL1-DV1	17	8.0	WAB450-I-B-P-106-HB	52	8.7
WAB450-I-B-P-153-HB	18	9.8	WAB450-I-B-P-62-HB	53	2.8
WAB450-11-1-P40-1-HB	19	6.2	WAB450-I-B-P-51-1-1	54	4.5
WAB450-24-3-4-P18-3-1	20	7.0	**WAB450-4-1-A16**	55	9.9
WAB450-I-B-P-133-HB	21	8.1	**WAB450-4-1-A26**	56	4.9
WAB450-I-B-P-163-2-1	22	5.3	**WAB450-B-16A2.7**	57	5.5
WAB450-5-1-BL1-DV6	23	7.2	**WAB450-B-16A1.2**	58	11.0
WAB450-I-B-P-157-2-1	24	6.6	**WAB450-4-1-A6**	59	3.8
WAB450-I-B-P-23-HB	25	5.1	**WAB450-B-19A1.2**	60	4.6
WAB450-I-B-P-24-HB	26	5.5	**WAB450-B-16A1.8**	61	5.6
WAB450-24-3-P3-1-HB	27	4.5	**WAB450-B-19A2.8**	62	4.9
WAB450-6-2-9-MB-HB	28	7.7	**WAB450-B-19A1.9**	63	5.4
WAB450-4-A9	29	1.7	**WAB450-B-18A2.2**	64	3.0
WAB450-B-19A1.8	30	1.1	**WAB450-B-16A2.4**	65	5.7
WAB450-B-18A2.8	31	9.2	WAB450-I-B-P-139-HB	66	2.0
WAB450-16A1.6	32	3.3	WAB450-I-B-P-550-HB	67	3.5
WAB450-B-16A1.4	33	5.4	WAB450-I-B-P-26-1-1 68	68	10.3
WAB450-B-1A1.1	34	2.2	WAB450-5-1-4-B-1-H2	69	0.9
WAB450-4-1A14	35	8.4	WAB450-I-B-P-2-2-1	70	0.9

The 26 lines in boldface were double haploids derived from BC_2F_1, while the remainder were BC_2F_8 lines developed using repeated selfing and pedigree selection.

Highlights

The relative contribution of each parent and the extent of genetic differences among the 70 sister lines was established using 130 well-distributed microsatellite markers which cover 1725 centiMorgans (cM) of the rice genome. The average proportion of *O. sativa*-recurrent parent genome was 87.4% (1,508 cM), while the observed average proportion of *O. glaberrima* donor genome was 6.3% (108 cM).

Non-parental alleles were detected in 83% of the lines and contributed an average of 38 cM per line (~2.2% of genomic DNA). Lines that had undergone eight generations of inbreeding in the field contained significantly more non-parental alleles (av. 2.7%) compared to the DH lines (av. 1.3%) that were developed from BC_2 anthers. Using both cluster and principal component analyses, two major groups were detected in these materials. The NERICA varieties (NERICA1 to 7) clustered in one group while the remaining 63 lines clustered in another group, suggesting that the second group may offer significant opportunities for further selection and variety development.

Polymorphism and parental genome coverage

An initial polymorphism survey was conducted using DNA from the two parents (WAB 56-104 and CG 14). One hundred and thirty of the 164 SSR primers (79.3 %) screened were polymorphic between the two parents. The number of polymorphic markers per chromosome varied from eight on chromosomes 7 and 12 to 15 on chromosome 1 (Figure 13), and the overall average was 10.8 polymorphic markers per chromosome. The 70 lines were then genotyped with the 130 polymorphic SSR markers. Introgressions were detected in all individuals and on all chromosomes. *O. sativa* alleles were detected at all 130 marker loci in one or more individuals but only 57 markers (43.8%) showed introgressions from *O. glaberrima*. When the data from the 130 markers were used to estimate the proportion

of each parental genome in the 70 individuals, *O. glaberrima* DNA represented from 0.9 to 12.1% of the genome (Table 13; Figure 14) while *O. sativa* represented between 79.0 to 94.4%. The average proportion of the genome containing *O. glaberrima* alleles in the 70 lines was 6.3% (108 of 1,725 centiMorgans while it was 87.4% (1507.9 of 1,725 cM) for the *O. sativa* parent. Heterozygosity was observed at 19 marker loci in a total of 28 lines (40%). The frequency of heterozygosity ranged from 0.3 to 3.4 loci/line, and the average was only 0.4 per line.

Non-parental alleles were detected at 20 SSR loci (15.4%) in one or more of the 58 lines (82.9%). The frequency of non-parental alleles among the 58 lines varied from 0.4 to 5.5% non-parental alleles/line, with an average of 2.7%. When the genomic composition of the 44 lines developed through pedigree selection was compared with those of the 26 double haploids, there was significantly higher ($p < 0.05$) representation of *O. glaberrima* introgressions in the pedigree lines (6.7%) than in the double haploids (5.5%), but the proportion of recurrent parent (*O. sativa*) genome remained identical (87.4%). More significantly ($p < 0.001$), the number of loci containing non-parental alleles was twice as high in the pedigree lines (2.7%) compared to the double haploids (1.3%). The mean proportion of loci showing introgression of *O. glaberrima* in the seven NERICA varieties (NERICA1 to NERICA7) was 8.2%, which is significantly ($p < 0.05$) higher than in the 63 sister lines (6.0%). However, the proportion of non-parental alleles and of recurrent parent genome in the seven NERICA varieties were not different from their sister lines.

The distribution of introgressions varied among the 12 rice chromosomes. The chromosomes with the fewest *O. glaberrima* alleles (25% of SSR loci) and those with the highest proportion of *O. glaberrima* alleles (87.5% of markers) were chromosomes 3 and 12, respectively. When the map distances (cM) between markers were considered as the basis for estimating the extent of *O. glaberrima*

introgressions on each chromosome, chromosome 3 had the smallest amount of introgressed DNA (2.5%) while chromosome 6 had the highest (21.5%), and the overall average size of introgressed DNA per chromosome was 7.5% (Figure 12). The proportion of recurrent parent (*O. sativa*) genome varied from 64.3% on chromosome 6 to 94.7% on chromosome 11. Eight of the 12 chromosomes (all except chromosomes 3, 5, 9 and 12) contained heterozygous loci and the average heterozygosity for the eight chromosomes varied from 0.2 to 1.5%. Non-parental alleles were also observed in eight of the 12 chromosomes (all except chromosomes 5, 7, 10 and 11). The highest proportion of non-parental alleles (5.2%) was observed on chromosome 6 where the average across all 70 lines was 1.4% of loci. The proportion of non-parental alleles across chromosomes showed a high negative correlation ($r = -0.72$) with the proportion of recurrent parent (WAB 56-104) genome.

Genetic relationships among lines

Cluster analysis using the simple matching coefficients derived from SSR markers produced two major groups, with six sub-groups observed in the dendrogram in Figure 15. The seven released NERICA varieties belong to one major group (group-1), with NERICA6 being the most genetically divergent. Principal component analysis (PCA) also revealed two major groups. As shown in Figure 12, a plot of PC1 (12%) and PC2 (7%) clearly separated the two groups in the same way as the dendogram. There were two differences between the dendogram and principal component analysis: (i) NERICA3 was intermediate between the two groups in the PCA while NERICA6 was the outlier in the dendrogram, and (ii) the six subgroups from the dendrogram were not evident in the PCA.

Since NERICA1–NERICA7 were selected for variety release from among the 70 lines developed from the same parents, this study compared the average introgression in these seven varieties (Figure 13) with those in other sister lines and found that the released

varieties contained significantly more *O. glaberrima* DNA than the other lines. Based on results from field evaluation for phenotypic traits and participatory varietal selection, the present study suggests that the presence of specific *glaberrima* introgressions is associated with superior performance in the field. This hypothesis, however, remains to be tested using QTL and association analysis to identify which specific segments of the *O. glaberrima* genome are associated with superior agronomic performance in the upland rice production system.

The seven NERICA varieties represented only one of the two major groups revealed both by the cluster and principal component analyses (Figures 15 and 16), suggesting that the other lines may provide an opportunity for selection and additional varietal development. The possibility for further selection and varietal development within group-2 will be highly dependent on the availability of reliable morpho-agronomic data from multi-location trials.

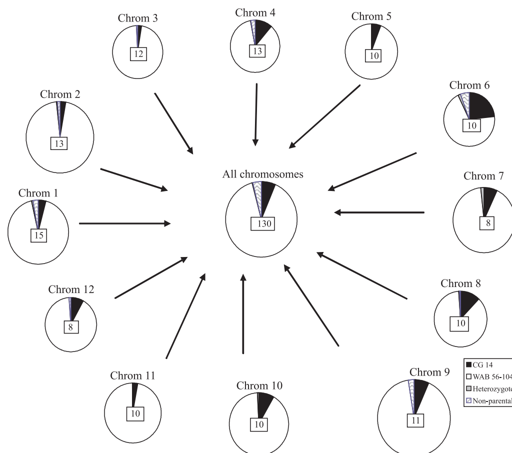

Figure 13. Pie charts for 12 rice chromosomes depicting the proportion of genome introgression in an interspecific rice population derived from CG 14 (donor) and WAB 56-104 (recurrent) parents. The pie charts were plotted from the graphical genotyping analyses outputs; the numbers in the center of the pies correspond to the number of SSR markers used in the study.

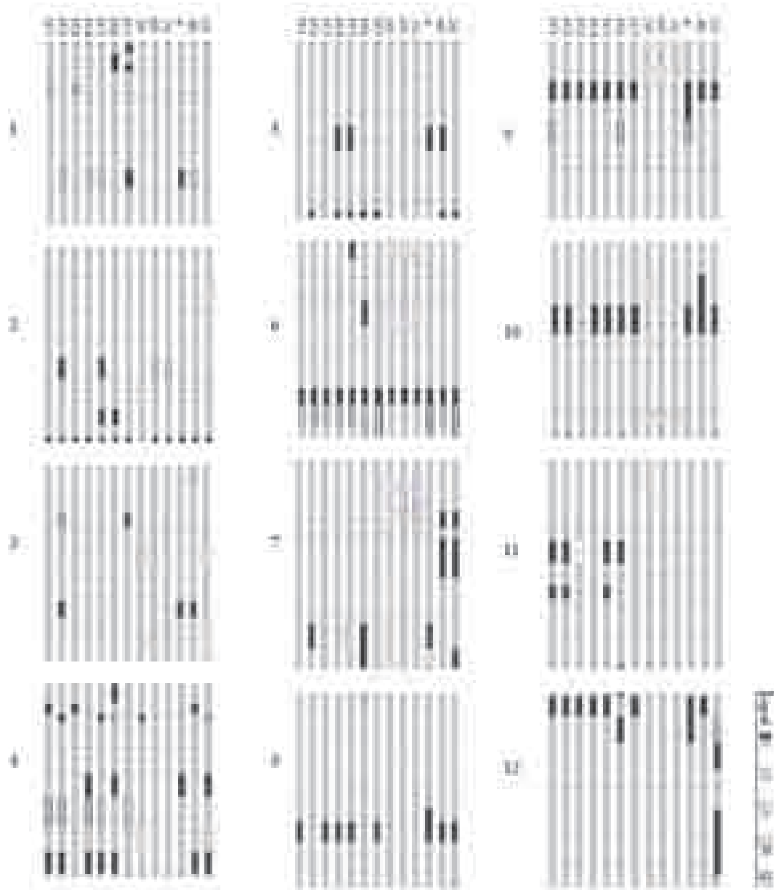

Figure 14a.
Graphical genotyping of the 13 interspecific lines using 130 microsatellite markers. Vertical bars represent the 12 chromosomes of rice, with chromosome number given on the left side of each bar.

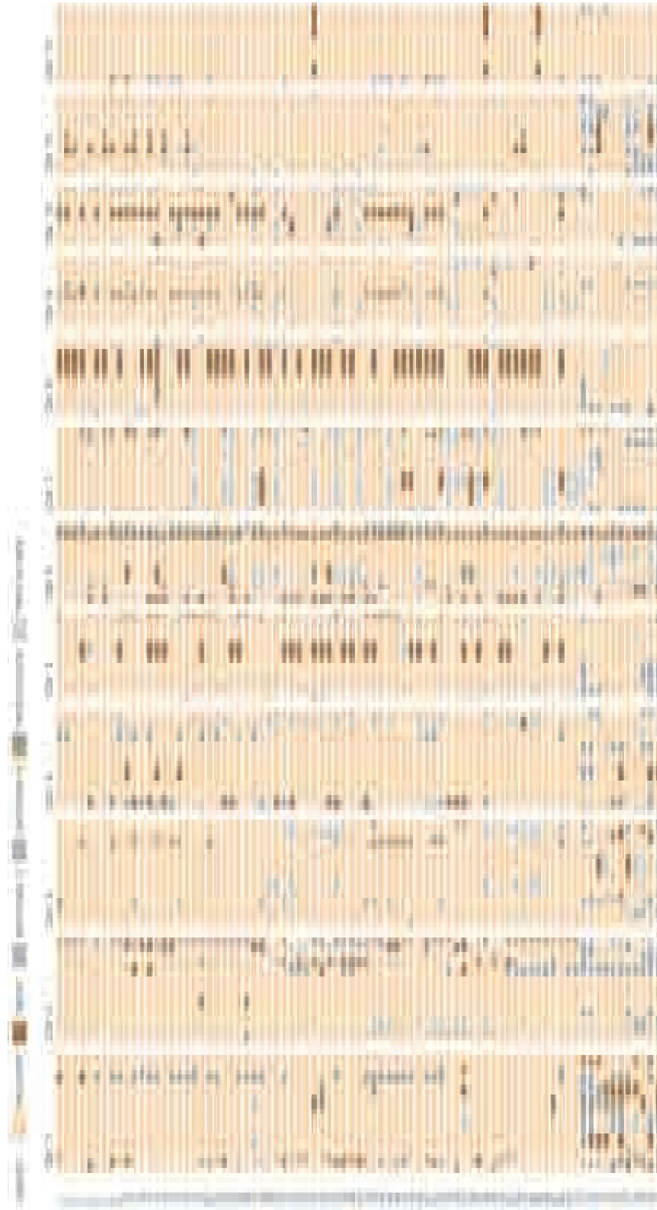

Figure 14b.
Graphical genotyping of the 81 interspecific lines using 130 microsatellite markers. Vertical bars represent the 12 chromosomes of rice, with chromosome number given on the left side of each bar.

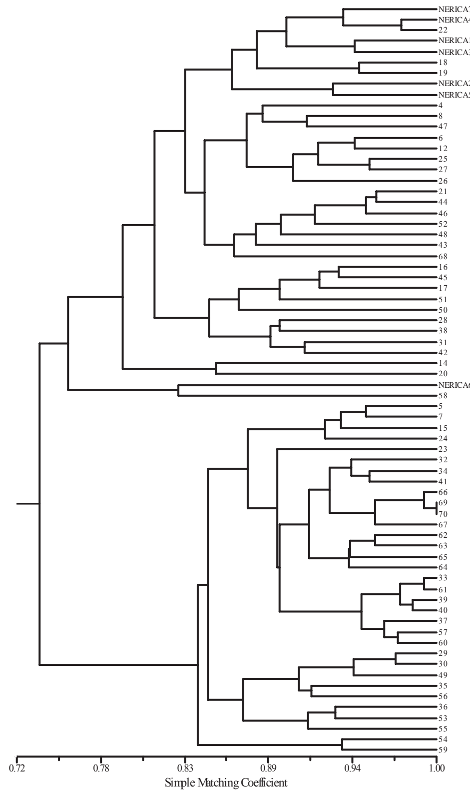

Figure 15. Dendrogram of the 70 interspecific lines using simple matching coefficients derived from 130 microsatellite markers. The lines were separated into two major groups and six subgroups although six other lines did not fit into the latter. Numbers correspond to the names as shown in Table 13. All lines within group-2, except the five lines indicated by arrows, contained introgression lower than the 6.3% average for the population.

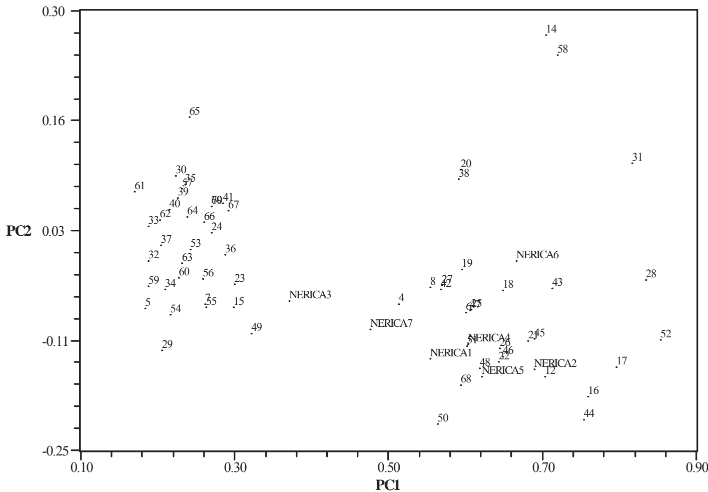

Figure 16. Score plot of the first two principal components from principal component analysis of the 70 interspecific lines genotyped with 130 microsatellite markers. Numbers in the plot correspond to the names as shown in Table 13.

Each chromosome is segmented by horizontal lines at the marker positions. Numbers on the top of the vertical bars refer to the 13 lines: N1 to N7 refers to NERICA1 to 7 (e.g. N1: NERICA1; N7: NERICA7); line number 30, 69 and 70 had the lowest introgression while line 4, 48 and 58 contained the highest introgression. Refer to Table 12 for pedigree for each line. Legend: A: CG 14 genome; B: WAB 56-104 genome; H: heterozygote; U: non-parental alleles; M: missing data.