CONSTRUCTION OF A HIGH DENSITY, COMPOSITE RFLP LINKAGE MAP FOR CULTIVATED SUNFLOWER (*Helianthus annuus* L.).

S.T. Berry, A.J. Leon*, P. Challis, C. Livini, R. Jones, C.C. Hanfrey, S. Griffiths, A. Roberts*.

ZENECA Seeds, Jealott's Hill Research Station, Bracknell, Berkshire. RG42 6EY. U.K.
* ZENECA Agricola, Balcarce Research Station, Ruta 226 Km 60.3, Casilla de Correo 30, 7620 Balcarce, Argentina.

Abstract
A high density, composite RFLP linkage map for the sunflower genome was constructed based on the segregation of 635 marker loci, detected by 500 probes, in 9 different F_2 populations. The genetic markers covered 1472 centiMorgans (cM) of the sunflower genome and were arranged in 17 linkage groups (A to Q), corresponding to the haploid number of chromosomes in this species. Three loci remained unlinked, but two of these markers were tightly linked to each other. Although the average interval size was 2.3 cM, there were still a number of persistent gaps (i.e. > 20 cM) as well as areas of acute marker clustering, some of which probably correspond to the location of centromeres. The high level of gene duplication within the sunflower genome is consistent with the hypothesised evolution of *H. annuus* through allopolyploidy; however homoeologous linkage groups could not be identified.

Key words
Restriction fragment length polymorphisms (RFLPs) - Linkage map - Composite RFLP map - *Helianthus annuus* - Sunflower

Introduction
Cultivated sunflower (*Helianthus annuus* L.) is a diploid, allogamous annual (2n=2x=34) which is grown primarily for its seed oil throughout the warm temperate regions of the five continents. The annual world production of sunflower oil is over 200 million tonnes per annum, which is currently third behind soybean and cottonseed oils. Despite the economic importance of sunflower, the application of biotechnology to this crop has been extremely limited. Recently, however, the use of restriction fragment length polymorphisms (RFLPs) as genetic markers has been reported in sunflower (BERRY *et al.* 1994; GENTZBITTEL *et al.* 1994) and these studies revealed that there is a high level of DNA polymorphism in the cultivated germplasm pool. This finding has allowed the rapid construction of the first RFLP linkage maps for sunflower (BERRY *et al.* 1995; GENTZBITTEL *et al.* 1995) which have already been used to identify both qualitative (MOUZEYAR *et al.* 1995, LEON *et al.* in press) and quantitative trait loci (LEON *et al.* 1995). The two published sunflower maps differ in that the one described by BERRY *et al.* (1995) is based on a single F_2 mapping population, whereas the one reported by GENTZBITTEL *et al* (1995) is a composite map produced from three F_2 and two BC_1 populations. The use of a single intraspecific population (ZENB8xHA89) only allowed BERRY *et al.* (1995) to map 55% of the available sunflower RFLP probes. In this paper we describe the addition of 400 new loci to the ZENB8xHA89 map using the segregation data from another 8 F_2 populations and the computer program JoinMap (STAM 1993) in order to produce a single composite map.

Materials and methods
The methods for RFLP analysis and linkage map construction are described by BERRY *et al.* (1994 and 1995). Six hundred and twelve low-copy probes selected from the cDNA (582) and *PstI* (30) genomic libraries described by BERRY *et al.* (1994) were radioactively-labelled and hybridised to Southern blots of the 13 inbred lines that were used to make the 9 F_2 mapping populations (Table 1). Probes that were identified as being polymorphic in a given population were then hybridised to Southern blots of F_2 DNA and the data scored and analysed according to BERRY *et al.* (1995). In addition to the RFLP loci, one isozyme marker (*Pgd-3*) was mapped on the ZENB8xHA89 F_2 according to CARRERA *et al.* (1995), as well as the dominant gene (*Hyp*) controlling white pigmentation of the hypodermis seed layer (LEON *et al.* in press).

Table 1. The number of individuals and RFLP loci mapped in each of the nine F_2 populations used to construct the sunflower composite map.

Population	No. of plants	No. of loci mapped
ZENB8xHA89	289	250
ZENB8xPAC2	70	334
ZENB8xZENR7	142	91
BSA52xRHA297	58	114
HA89xRHA271	94	118
ZENR7xRHA801	94	93
HA89xZENR9	115	38
ZENR1xZENR8	207	103
ZENB4xHA300	218	152

Linkage maps for each of the nine individual F_2 populations were constructed using MAPMAKER version 3.0 (LANDER *et al.* 1987). The composite map was produced using JoinMap (STAM 1993) by analysing the RFLP data on a group by group basis, using the default LOD score of 0.05 and the Kosambi mapping function. The integration of distinct genetic linkage maps requires the use of at least two common markers per linkage group between any two populations. These markers then act as anchor points allowing the merging of linkage groups in the correct orientation from the different crosses.

Results and discussion

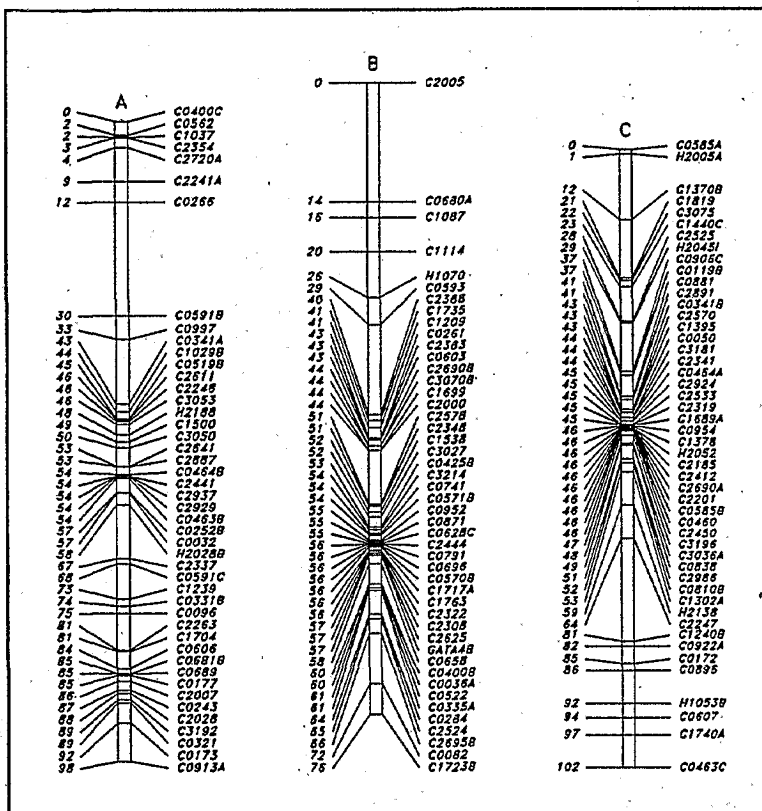
To-date 500 of the available 612 low-copy RFLP probes have been mapped on at least one of the nine F_2 populations listed in Table 1. These 500 probes detected 633 different loci within the sunflower genome, the majority of which were scored as codominant markers and followed the expected 1:2:1 segregation ratio. However, in each F_2 population there were regions of distorted segregation which were caused by a reduction in one or other of the parental homozygous classes (data not shown). Distorted segregation is a common phenomenon in the published plant RFLP maps and it is thought to be caused by selection against individual genotypes during either sporogenesis, gametogenesis, seed development, germination or plant growth. The 633 RFLP loci, plus the *Pgd-3* and *Hyp* loci, were arranged into 17 linkage groups (Figure 1), which probably correspond to the 17 haploid chromosomes of *H. annuus*. The letter used here to identify each group (A to Q) is completely arbitrary as no classical genetic map exists for sunflower. The composite map shown in Figure 1 covers 1472 cM of the sunflower genome and represents the most complete and saturated genetic map available for *H. annuus*. The process of consensus mapping has effectively added a further 400 loci to the ZENB8xHA89 map, which covered 1380 cM and was thought to represent 60-80% of the sunflower genome (BERRY *et al.* 1995). It is now clear that this earlier estimate of genome coverage was too conservative and, in fact, the composite map covers approximately 90% of the genome based on an estimated total genome length of 1650 cM (GENTZBITTEL *et al.* 1995). The independent segregation of 3 loci (C0592A and C1370A which are 2 cM apart and H2028A) also confirms that at least 100 cM of the sunflower genome are still missing.

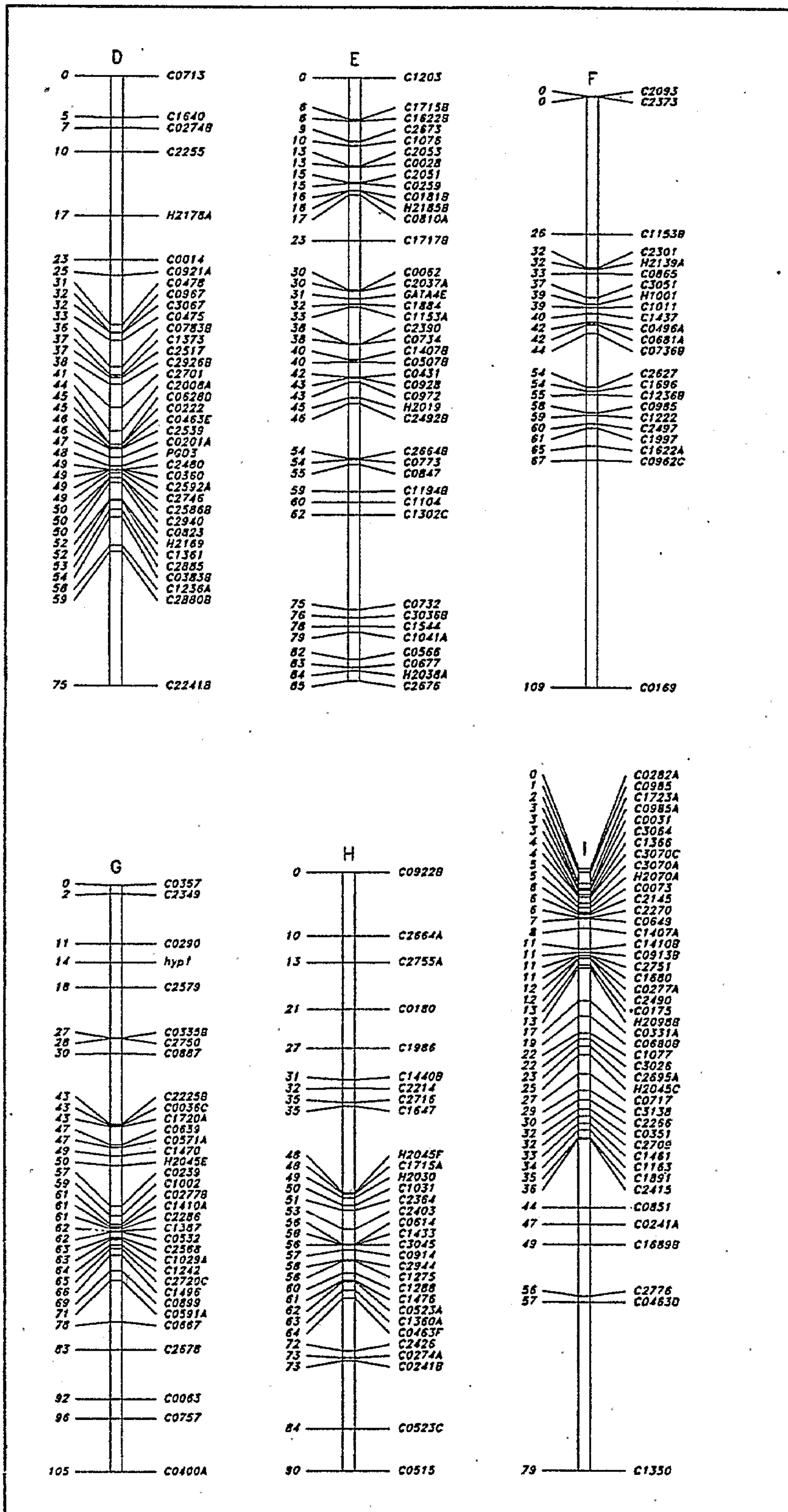
Due to the variation in recombination frequency across the genome, the distribution of markers within linkage groups is far from uniform. For example, large gaps of greater than 20 cM are present on linkage groups D, F, I, M and O (Figure 1). Gaps such as these are common in the published plant RFLP linkage maps and either represent regions of high recombination or reflect an under-representation of clones from these areas in the libraries used as probe sources. As well as regions devoid of marker loci, there are also areas of acute marker clustering such as those seen on linkage groups B, C and J (Figure 1). In some plant species these areas of reduced recombination have been found to correspond to the position of the centromere. In general, the ordering of marker loci in the composite map is fairly consistent with the individual F_2 maps (data not shown); however the exact order within these tight clusters of marker loci is uncertain.

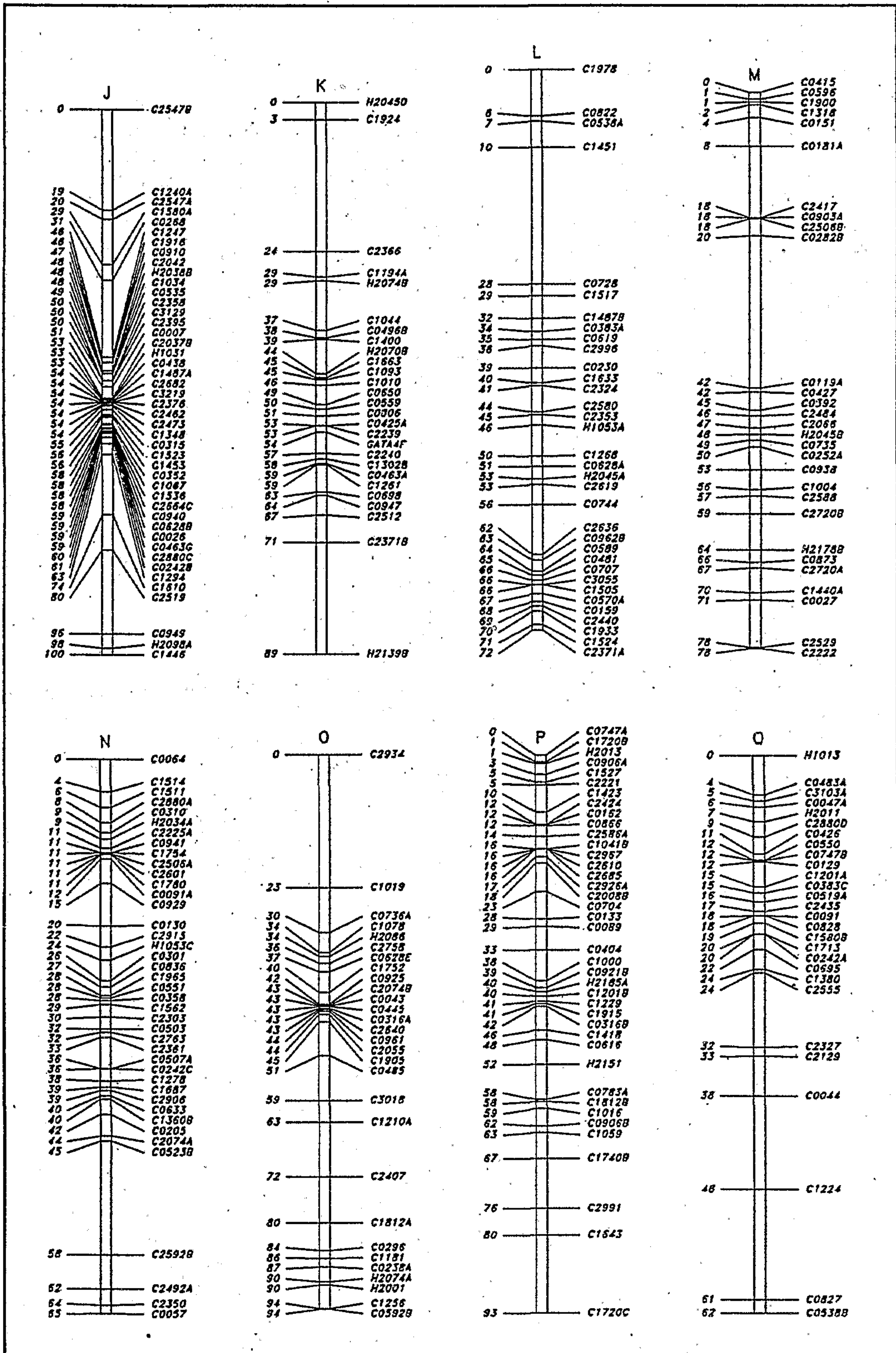
The presence of 133 duplicated loci on the linkage map supports the hypothesis that *H. annuus* has evolved through allopolypoidy via the hybridisation of two *Viguiera* progenitor species of n=8 and n=9 (HEISER and

SMITH 1955); however there is no apparent conservation of linkage blocks between homoeologous chromosomes as seen in the RFLP linkage maps of other allopolyploids, such as maize (HELENTJARIS *et al.* 1988). The sunflower genome may be similar to that of soybean where the collinearity between the progenitor genomes is being lost through the process of diploidisation (ZHU et al. 1994). An alternative hypothesis is that a large number of the independent duplications within the sunflower genome are due to pseudogene families (GENTZBITTEL *et al.* 1995). In fact, the *Pst*I clone H2045, which mapped to 9 unlinked loci, is part of the reverse transcriptase gene from a *copia*-like retrotransposon (unpublished data). The presence of such elements in the sunflower genome would allow the reverse transcription of endogenous mRNA to take place and hence the formation of pseudogenes.

Figure 1. A composite sunflower linkage map of 633 RFLP loci identified by 500 low-copy DNA probes. The 17 linkage groups are listed at the *top*, the loci are listed on the *right* and the cumulative map distances in centiMorgans (cM) are shown on the *left*. Loci nomenclature is described by BERRY *et al.* (1995). The *Pgd-3* and *Hyp* loci map to linkage groups D and G respectively.




D
0 C0713
5 C1640
7 C0274B
10 C2255
17 H2178A
23 C0014
25 C0921A
31 C0478
32 C0967
32 C3067
33 C0475
36 C0783B
37 C1373
37 C2517
38 C2926B
41 C2701
44 C2008A
45 C0628D
45 C0222
46 C0463E
46 C2539
47 C0201A
48 PG03
49 C2480
49 C0360
49 C2592A
49 C2746
50 C2586B
50 C2940
50 C0823
52 H2169
52 C1361
53 C2885
54 C0383B
58 C1236A
59 C2880B
75 C2241B
E
0 C1203
6 C1715B
6 C1622B
9 C2673
10 C1076
13 C2053
13 C0028
15 C2051
15 C0259
16 C0181B
16 H2185B
17 C0810A
23 C1717B
30 C0062
30 C2037A
31 GATA4E
32 C1884
33 C1153A
38 C2390
38 C0734
40 C1407B
40 C0507B
42 C0431
43 C0928
43 C0972
45 H2019
46 C2492B
54 C2664B
54 C0773
55 C0847
59 C1194B
60 C1104
62 C1302C
75 C0732
76 C3036B
78 C1544
79 C1041A
82 C0566
83 C0677
84 H2038A
85 C2676
F
0 C2093
0 C2373
26 C1153B
32 C2301
32 H2139A
33 C0865
37 C3051
39 H1001
39 C1011
40 C1437
42 C0496A
42 C0681A
44 C0736B
54 C2627
54 C1696
55 C1236B
58 C0985
59 C1222
60 C2497
61 C1997
65 C1622A
67 C0962C
109 C0169
G
0 C0357
2 C2349
11 C0290
14 hyp1
18 C2579
27 C0335B
28 C2750
30 C0887
43 C2225B
43 C0036C
43 C1720A
47 C0639
47 C0571A
49 C1470
50 H2045E
57 C0239
59 C1002
61 C0277B
61 C1410A
61 C2286
62 C1387
62 C0532
63 C2568
63 C1029A
64 C1242
65 C2720C
66 C1496
69 C0899
71 C0591A
78 C0667
83 C2678
92 C0063
96 C0757
105 C0400A
H
0 C0922B
10 C2664A
13 C2755A
21 C0180
27 C1986
31 C1440B
32 C2214
35 C2716
35 C1647
48 H2045F
48 C1715A
49 H2030
50 C1031
51 C2364
53 C2403
56 C0614
58 C1433
56 C3045
57 C0914
58 C2944
58 C1275
60 C1288
61 C1476
62 C0523A
63 C1360A
64 C0463F
72 C2426
73 C0274A
73 C0241B
84 C0523C
90 C0515
I
0 C0282A
1 C0985
2 C1723A
3 C0985A
3 C0031
3 C3064
4 C1366
4 C3070C
5 C3070A
5 H2070A
6 C0073
6 C2145
6 C2270
7 C0649
8 C1407A
11 C1410B
11 C0913B
11 C2751
11 C1680
12 C0277A
12 C2490
13 C0175
13 H2098B
17 C0331A
19 C0680B
22 C1077
22 C3026
23 C2695A
25 H2045C
27 C0717
29 C3138
30 C2256
32 C0351
32 C2709
33 C1461
34 C1163
35 C1891
36 C2415
44 C0851
47 C0241A
49 C1689B
56 C2776
57 C0463D
79 C1350


J
0 C2547B
19 C1240A
20 C2547A
29 C1580A
31 C0268
46 C1247
46 C1916
47 C0910
48 C2042
48 H2038B
48 C1034
49 C0535
50 C2358
50 C3125
50 C2395
51 C0007
53 C2037B
53 H1031
53 C0438
54 C1487A
54 C2682
54 C3219
54 C2376
54 C2462
54 C2473
54 C1348
55 C0315
56 C1523
56 C1453
58 C0352
58 C1087
58 C1336
58 C2564C
59 C0940
59 C0628B
59 C0026
59 C0463C
60 C2880C
61 C0242B
63 C1294
74 C1610
80 C2519
96 C0949
98 H2098A
100 C1446
K
0 H2045D
3 C1924
24 C2366
29 C1194A
29 H2074B
37 C1044
38 C0496B
39 C1400
44 H2070B
45 C1663
45 C1093
46 C1010
49 C0650
50 C0559
51 C0006
53 C0425A
53 C2239
54 GATA4F
57 C2240
58 C1302B
59 C0463A
59 C1261
63 C0698
64 C0947
67 C2512
71 C2371B
89 H2139B
L
0 C1978
6 C0822
7 C0538A
10 C1451
28 C0728
29 C1517
32 C1487B
34 C0383A
35 C0619
36 C2996
39 C0230
40 C1633
41 C2324
44 C2580
45 C2353
46 H1053A
50 C1268
51 C0628A
53 H2045A
53 C2619
56 C0744
62 C2636
63 C0962B
64 C0589
65 C0481
66 C0707
66 C3055
66 C1505
67 C0570A
68 C0159
69 C2440
70 C1933
71 C1524
72 C2371A
M
0 C0415
1 C0596
1 C1900
2 C1318
4 C0151
8 C0181A
18 C2417
18 C0903A
18 C2506B
20 C0282B
42 C0119A
42 C0427
45 C0392
46 C2484
47 C2066
48 H2045B
49 C0735
50 C0252A
53 C0938
56 C1004
57 C2588
59 C2720B
64 H2178B
66 C0873
67 C2720A
70 C1440A
71 C0027
78 C2529
78 C2222
N
0 C0064
4 C1514
6 C1511
8 C2880A
9 C0310
9 H2034A
11 C2225A
11 C0941
11 C1754
11 C2506A
11 C2601
11 C1780
12 C0091A
15 C0929
20 C0130
22 C2913
24 H1053C
26 C0301
27 C0836
28 C1965
28 C0557
28 C0358
29 C1562
30 C2303
32 C0503
32 C2763
33 C2367
36 C0507A
36 C0242C
38 C1278
39 C1687
39 C2906
40 C0633
40 C1360B
42 C0205
44 C2074A
45 C0523B
58 C2592B
62 C2492A
64 C2350
65 C0057
O
0 C2934
23 C1019
30 C0736A
34 C1078
34 H2086
36 C2758
37 C0628E
40 C1752
42 C0925
43 C2074B
43 C0043
43 C0445
43 C0316A
43 C2640
44 C0961
44 C2055
45 C1905
51 C0485
59 C3018
63 C1210A
72 C2407
80 C1812A
84 C0296
86 C1181
87 C0238A
90 H2074A
90 H2001
94 C1256
94 C0592B
P
0 C0747A
1 C1720B
1 H2013
3 C0906A
5 C1527
5 C2221
10 C1423
12 C2424
12 C0152
12 C0866
14 C2586A
16 C1041B
16 C2957
16 C2610
16 C2685
17 C2926A
18 C2008B
23 C0704
28 C0133
29 C0089
33 C0404
38 C1000
39 C0921B
40 H2185A
40 C1201B
41 C1229
41 C1915
42 C0316B
46 C1418
48 C0616
52 H2151
58 C0783A
58 C1812B
59 C1016
62 C0906B
63 C1059
67 C1740B
76 C2991
80 C1643
93 C1720C
Q
0 H1013
4 C0483A
5 C3103A
6 C0047A
7 H2011
9 C2880D
11 C0426
12 C0550
12 C0747B
12 C0129
15 C1201A
15 C0383C
16 C0519A
17 C2435
18 C0091
18 C0828
19 C1580B
20 C1713
20 C0242A
22 C0695
24 C1380
24 C2555
32 C2327
33 C2129
38 C0044
46 C1224
61 C0827
62 C0538B

References

Berry S.T., Allen R.J., Barnes S.R., Calgari P.D.S. (1994) Molecular marker analysis of *Helianthus annuus* L. 1. Restriction fragment length polymorphism between inbred lines of cultivated sunflower. Theor. Appl. Genet. 89:435-441.

Berry S.T., Leon A.J., Hanfrey C.C., Challis P., Burkholz A., Barnes S.R., Rufener G.K., Lee M., Caligari P.D.S. (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. Theor. Appl. Genet. 91:195-199.

Carrera A., Poverene M., Roberts A., Berry S.T., Leon A.J. (1995) Localizacion del locus isoenzimatico *Pgd-3* en el mapa de girasol. Acatas del XXVI Congreso Argentino de Genetica, Bariloche, Argentina, p 43.

Gentzbittel L., Zhang Y.X., Vear F., Griveau Y., Nicolas P. (1994) RFLP studies of genetic relationships among inbred lines of cultivated sunflower, *Helianthus annuus* L.: Evidence for distinct restorer and maintainer germplasm pools. Theor. Appl. Genet. 89:419-425

Gentzbittel L., Vear F., Zhang Y.X., Berville A., Nicolas P. (1995) Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). Theor. Appl. Genet. 90:1079-1086.

Heiser C.B., Smith D.M. (1955) New chromosome numbers in *Helianthus* and related genera. Proc. Ind. Acad. Sci. 64:250-253.

Helentjaris T., Weber D., Wright S. (1987) Identification of the genome location of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics 118:353-363

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newburg L. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181

Leon A.J., Lee M., Rufener G.K., Berry S.T., Mowers R.P. (1995) Use of RFLP markers for genetic linkage analysis of oil percentage in sunflower seed. Crop Sci. 35:558-564.

Leon A.J., Lee M., Rufener G.K., Berry S.T., Mowers R.P. (in press) Genetic mapping of seed hypodermis colour and its relation with seed oil content in sunflower (*Helianthus annuus* L.). Crop Sci.

Mouzeyar S., Roeckel-Drevet P., Gentzbittel L., Philippon J., Tourvieille De Labrouche D., Vear F., Nicolas P. (1995) RFLP and RAPD mapping of the sunflower *Pl1* locus for resistance to *Plasmopara halstedii* race 1. Theor. Appl. Genet. 91:733-737

Stam P. (1993) Construction of integrated genetic maps by means of a new computer package: JoinMap. Pl. Journal 3:739-744.

Zhu T., Schupp J.M., Oliphant A., Keim P. (1994) Hypomethylated sequences: Characterisation of the duplicate soybean genome. Mol. Gen. Genet. 244:638-645