

New York State Agricultural Experiment Station
Geneva, New York

QUIKMAP: A MICROSOFT® EXCEL SPREADSHEET FOR MAKING A QUICK COMPARISON
BETWEEN MARKERS IN A LINKAGE GROUP AND A NEW MARKER

by

N.F. Weeden & J. Barnard

Computer Centre

Technical Report 20-2

QUIKMAP: A MICROSOFT® EXCEL SPREADSHEET FOR MAKING A QUICK COMPARISON BETWEEN MARKERS IN A LINKAGE GROUP AND A NEW MARKER

N.F. Weeden & J. Barnard

New York State Agricultural Experiment Station
Cornell University

1 Linkage analysis

The primary use of QUIKMAP is to rapidly and easily identify linkages between a new segregating marker and any marker previously scored in the population. Depending on how the alleles are designated (i.e. if the parental genotypes are unknown or not specified during data assembly) linkage can be identified by a match-count close to maximum or close to zero (in the case of repulsion phase linkage). In general, a value greater than 80% of the maximum number of matches or less than 20% of the maximum is suggestive of linkage. QUIKMAP is also helpful in assembling a map from partial linkage groups because one can quickly determine if the end marker on one linkage group exhibits linkage with an end of any other linkage group. Note that loci giving distorted segregation patterns tend to group together as a result of the high number of matches or mismatches found when the segregation pattern of two such loci are compared. Caution should be used when interpreting the results from such loci. However, QUIKMAP is particularly useful for testing linkage relationships of a marker for which only partial data is available (e.g. one side of a gel became blurred or only one genotype is clear).

2 Methodology

Sample genotypes and markers in the linkage group are represented by matrix **G**

$$\begin{pmatrix} G_{11} & G_{12} & \dots & G_{1m} \\ G_{21} & G_{22} & \dots & G_{2m} \\ \dots & \dots & \dots & \dots \\ G_{n1} & G_{n2} & \dots & G_{nm} \end{pmatrix}$$

where n is the number of samples and m is the number of markers.

The test marker is represented by vector **T**

$$\begin{pmatrix} T_1 \\ T_2 \\ \dots \\ T_n \end{pmatrix}$$

Comparisons between the test marker and the linkage group are placed in matrix **C**

$$\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1m} \\ C_{21} & C_{22} & \dots & C_{2m} \\ \dots & \dots & \dots & \dots \\ C_{n1} & C_{n2} & \dots & C_{nm} \end{pmatrix}$$

where the elements of **C** have values of unity or zero depending on the whether there is a match between the column of **G** and **T**. If an element of **T**, T_i contains the symbol 'x', then no match will be assessed even if 'x' is present in the corresponding row of **G**. That is,

$$C_{ij} = \begin{cases} 1 & \text{if } G_{ij} = T_i \text{ and } T_i \neq x, \\ 0 & \text{otherwise.} \end{cases}$$

3 Microsoft® Excel Spreadsheet for Quikmap

The structure of the spreadsheet follows the above description. Rows correspond to samples, the first m columns correspond to the markers in the linkage group, column $m+1$ holds the test markers, and columns $m+2 \dots 2m+1$ hold the comparisons.

For a concrete example, suppose that the linkage group comprises 4 markers and that 10 samples are scored. Reserving row 1 for header information, data is entered into rows 2...11, columns A...D. Test maker data is entered into column E.

The formula

$$=IF(AND(A2=\$E2, \$E2<>"x"), 1, 0)$$

is entered into cell F2 and the rest of the column 'filled down' [Edit menu]. Columns G...I are 'filled right' from column F.

The final row (row 12 in the example) counts the matches. Enter =SUM(F2:F11) into cell F12 and 'fill right' to cell I12.

	A	B	C	D	E	F	G	H	I
1	<i>p26c</i>	<i>gpi</i>	<i>s44i</i>	<i>c34</i>	<i>test</i>				
2	f	s	f	f	s	0	1	0	0
3	s	s	f	f	s	1	1	0	0
4	f	s	f	s	s	0	1	0	1
5	f	f	f	s	f	1	1	1	0
6	s	s	d	f	x	0	0	0	0
7	f	s	f	f	s	0	1	0	0
8	d	s	f	f	s	0	1	0	0
9	s	f	f	s	s	1	0	0	1
10	f	s	f	s	f	1	0	1	0
11	s	s	f	f	f	0	0	1	1
12					<i>Matches=</i>	4	6	3	3