Excess Heterozygosity Contributes to Genetic Map Expansion in Pea Recombinant Inbred Populations

M. R. Knox¹ and T. H. N. Ellis

Department of Crop Genetics, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

Manuscript received April 4, 2002
Accepted for publication June 27, 2002

ABSTRACT

Several plant genetic maps presented in the literature are longer than expected from cytogenetic data. Here we compare F₂ and RI maps derived from a cross between the same two parental lines and show that excess heterozygosity contributes to map inflation. These maps have been constructed using a common set of dominant markers. Although not generally regarded as informative for F₂ mapping, these allowed rapid map construction, and the resulting data analysis has provided information not otherwise obvious when examining a population from only one generation. Segregation distortion, a common feature of most populations and marker systems, found in the F₂, but not the RI, has identified excess heterozygosity. A few markers with a deficiency of heterozygotes were found to map to linkage group V (chromosome 3), which is known to form rod bivalents in this cross. Although the final map length was longer for the F₂ population, the mapped order of markers was generally the same in the F₂ and RI maps. The data presented in this analysis reconcile much of the inconsistency between map length estimates from chiasma counts and genetic data.

Genetic maps are the theoretical placement, based on experimental data, of markers along a conceptual linkage group. Map length and marker order are influenced by many factors, including the frequency of double recombinants, errors in scoring and data input, environmental effects and sex differences on meiosis; in turn, marker type (codominant or dominant), the generation, size, and type of the population [F₂, recombinant inbred (RI), backcross (BC), doubled haploid (DH)] bear on these factors (Haldane 1919; Kosambi 1944; Lander et al. 1987; Lincoln and Lander 1992; Nilsson et al. 1993; Stam and Van Ooijen 1995; Staub et al. 1996; Sybenga 1996). Different mapping programs sometimes propose different marker orders from the same primary data, but having obtained a map, the order of markers is critical for map-based cloning or for marker-assisted selection in breeding programs.

With the advent of PCR-based marker systems, it is possible to obtain multiple markers from a few primer combinations and so generate maps rapidly, e.g., amplified fragment length polymorphism (AFLP; Vos et al. 1995) and sequence-specific amplification polymorphism (SSAP; Waugh et al. 1997; Ellis et al. 1998), even though both of these marker types are dominant.

For mapping, it is preferable for markers to segregate in a Mendelian fashion. Markers deviating from the expected ratios are described as having segregation distortion and have been reported for interspecies and intersubspecies crosses in rice (Harushima et al. 1996; Fukuta et al. 2000; Lu et al. 2000), pearl millet (Liu et al. 1994), tomato (Paran et al. 1995), alfalfa (Kaló et al. 2000), and peanut (Halward et al. 1993). Interspecies crosses tend to exhibit more segregation distortion but intraspecific crosses are also susceptible [common bean (Paredes and Gepts 1995), Medicago spp. (Jenczewski et al. 1997), and tomato (Helentjaris et al. 1986)]. This was borne out by Zamir and Tadmor (1986), who found significantly higher deviations occurring in interspecies compared to intraspecific crosses for lentil, pepper, and tomato using morphological and isozyme markers. Both codominant and dominant marker types are affected by segregation distortion [restriction fragment length polymorphism (RFLP) mapping in alfalfa (Brummer et al. 1993), Arabidopsis thaliana (Lister and Dean 1993), rice (Xu et al. 1997), rye (Korzun et al. 1998), and tomato (Paterson et al. 1991); isozyme loci of alfalfa (Kiss et al. 1993), barley (Konishi et al. 1992), pea (Weeden and Marx 1987), and pepper (Tanksley 1984); microsatellite in soybean (Jin et al. 1999); AFLP in conifer (Nikaido et al. 1999), rice (Virk et al. 1998), and tea (Hackett et al. 2000); randomly amplified polymorphic DNA (RAPD) in wheat (Manabe et al. 1999), Medicago spp. (Jenczewski et al. 1997; Kaló et al. 2000), and pea (Laucou et al. 1998); and morphological markers in pepper (Zamir and Tadmor 1986)].

Xu et al. (1997) found linked clusters of markers with skewed ratios mapping to corresponding regions in F₂, DH, BC, and RI linkage maps from interspecies crosses in rice. Mapping markers with skewed ratios may result in false association of markers to regions not otherwise

¹Corresponding author: Department of Crop Genetics, John Innes Centre, Norwich Research Park, Colney Ln., Norwich NR4 7UH, UK. E-mail: maggie.knox@bbsrc.ac.uk

Genetics 162: 861–873 (October 2002)
linked. Xu et al. (1997) attempted to overcome these difficulties by estimating the location of these markers with skewed ratios by an association method, while Kaló et al. (2000) estimated linkage in diploid alfalfa using the "special" maximum-likelihood equations of Lorieux et al. (1995).

The main pea RI mapping populations in our laboratory are from intraspecific Pisum sativum crosses. These populations were raised specifically for linkage analysis at the RI (F8 and later) stage and have been studied extensively at F11. In a study of the relationship between the genetic and cytogenetic maps in pea (Hall et al. 1997a,b) a large difference in map length, 1800 vs. ~750 cM, was found between these maps. A pea linkage map, consisting mainly of RAPD markers, was also found to have an excess map length at 1139 cM (Laugou et al. 1998). Map expansion in general and the disparity between the genetical and physical maps have been a major subject of discussion for a range of species (Taylor 1978; Burr et al. 1988; Burr and Burr 1991; Tanksley et al. 1992; Nilsson et al. 1993; Paran et al. 1995; Puchta and Hohn 1996; Sybenga 1996).

There is compelling cytogenetic evidence that in eukaryotes chiasmata are the manifestations of recombination sites, that the frequency of chiasmata per bivalent is proportional relative to the length of the chromosome, and this relationship is generally well conserved between species (Jones 1987; Rees and Narayan 1988; Tease and Jones 1995). Chiasma counts could be underestimated, but for the genetic map length to double the chiasma counts would also need to double and this is an unlikely oversight. Recombination nodules (RN), which are visible under the electron microscope, lodged within the synaptonemal complex at mid to late pachytene, are thought to be situated at sites of exchange events as the precursors for chiasmata and have been reported for Allium sps. (Albin and Jones 1988), locust and grasshopper (Bernelot-Moens and Moens 1986), tomato (Herickhoff et al. 1993; Sherman and Stack 1995), and lily (Stack et al. 1989). Recombination nodules are easier to resolve compared to chiasmata as they are visualized during the stages of meiosis when the chromosomes are less condensed. In all these cases there was a one-to-one relationship between chiasma and RN, and the correspondence between chiasma frequency, recombination, and map length has recently been corroborated by King et al. (2002), suggesting that chiasma counts were accurate. The RN map for tomato was still shorter by ~180 map units compared to the molecular map (Sherman and Stack 1995).

Scoring and typing errors have a considerable influence on map length estimates, resulting in extra, but nonexistent, crossovers altering either marker order or map length, or both (Buétow 1991; Lincoln and Lander 1992). Some scoring errors can be resolved by thorough checking of data but apparent "false errors" arising, for example, from DNA methylation events are not data-scoring errors (Knox and Ellis 2001).

The aim of this study was to investigate the problem of excess map length and to determine whether the map expansion function associated with inbreeding contributed to excess map length. To do this the behavior of the same set of markers in F2 and RI populations, for the same pea cross using karyotypically characterized and compatible parental lines, was analyzed. Here we present F2 and RI linkage data and describe F2 segregation distortion that is not present in the RI population. The excess of heterozygosity in the F2 along with data analysis from the RI core-mapping data goes some way in explaining the problem of relating the pea genetic and cytogenetic maps. The impact of an excess or deficiency of heterozygosity is generally relevant to genetic mapping in other species.

MATERIALS AND METHODS

Plant material: The JJ15 × JJ399 RI (F1) population has been described previously (Ellis et al. 1998). A cross using the same parental lines was carried out as part of a cytogenetic and genetic map integration study (Hall et al. 1997a,b). A total of 120 F2 seeds from this cross (a gift from J. S. Parker) were planted in the glasshouse at the John Innes Centre to produce the F2 generation.

Molecular markers: Markers derived from insertion sites for the pea Ty1-copia retroelement PDR1 (Lek et al. 1990) and an Spm-like transposable element have been used in this study. The SSAP method used to generate markers (Tps1) related to PDR1 has been described previously (Ellis et al. 1998); the specific primer (PPT, 5'-ATTCCACCGTTGAGGGGAG 3') was derived from the polyuridine tract of the element. Spm markers were also generated using the SSAP method but in this case the specific primer (SPM, 5'-TGTGGGCTAGTGGG 3') was based on the 12-bp terminal inverted repeat of an Spm-like element (Shirsat 1988), with GCC as an additional 3 bases of selection at the 3' end. The adapter primer (Taq, 5'-ATGAGTCTGAGAAAGC 3'), for a Taq1 digest, also required 2 or 3 bases of selection at the 3' end, depending on which element was being amplified. The PDR1 markers were generated with Taq + AA, AC, AG, AT, TA, TC, TG, TT, GT, and the Spm markers with Taq + ATA, ATC, ATG, ATT selection. Marker names are designated according to the type of element from which they are derived, Tps1 and Spm, shortened to T and S, respectively, in the tables, followed by the marker number and either "p" or "m" denoting the parental line from which they originate, JJ15 and JJ399, respectively.

Linkage mapping: For the F2 and RI map comparison analysis both Mapmaker (MM; Lander et al. 1987) and JOINMAP (JM; Stam and Van Ooijen 1995) were used to estimate the recombination frequencies and linkage data for map construction, using the transposable element markers (Table 1). Additionally, RI mapping data consisting of 347 markers, comprising RFLP, morphological, and molecular markers, are described in the text as the core mapping data (Ellis et al. 1998) and were used to construct two maps (Table 1, maps A and B) for map length and crossover number comparisons. Map A was constructed using only JM, with some adjustments to remove spurious linkages sometimes proposed by the mapping program; both MM and JM were used to construct map B with minor adjustments to accommodate map order differences between the two programs and marker order consensus.
from other pea mapping populations (Hall et al. 1997b; Ellis et al. 1998). In all cases the Haldane mapping function was used. Microsoft Excel was used to make graphs and charts of segregation ratios and $\chi^2$ values, and simple statistics were calculated using Minitab.

**RESULTS**

**Molecular markers:** The total number of markers mapped in the JI15 × JI399 RI core mapping data set was 347; a subset of these markers was selected from the SSAP data set for comparison of F2 and RI linkage analyses (Table 1). These dominant SSAP markers were chosen for this study as they have a rapid throughput for linkage analysis and a highly reproducible banding pattern. The SSAP banding pattern generated using the Spm primer gave a more complex pattern, providing an abundance of markers from four primer combinations, compared to nine with the Tps primer, finally to provide approximately the same number of markers (Table 2). The abundance of bands with the Spm primer reflects the high copy number of this element in the pea genome, assessed at $\sim 10^6$ from this study, compared to a copy number of $\sim 200$ of PDR1 (Ellis et al. 1998) per haploid genome.

Segregation ratios for the SSAP subset of markers, scoreable from both the F2 and RI populations, were analyzed and the data were split into markers acceptable for mapping and those with segregation distortion. Both parental lines contributed equally to SSAP alleles for both marker types (data not shown). A total of 217 Tps and Spm markers were scoreable for F2 and RI map comparison with the majority (137) having segregation ratios with $x^2 < 2$, of which 62 were Tps1 and 75 were Spm (Table 2). Eighty markers were not used; this was either due to extreme segregation distortion in the F2 population, discussed in detail later, or due to incomplete marker scoring in one population or the other.

**Recombination frequency:** Pairwise recombination fractions, $r$ values, were calculated using Mapmaker for the 137-marker subset and these were compared for the F2 and RI populations (Table 1). The pairwise comparison of $r$ values for all marker pairs showed a wide scatter of points (Figure 1), reflecting variation in the estimate of $r$ values for equivalent pairs in the two populations. Splitting these data into marker pairs linked in coupling and in repulsion (Figure 1) is informative. For the coupling data, the F2 and RI recombination fractions are well correlated (Pearson correlation coefficient = 0.85). The scatter is more pronounced for the pairwise analysis of markers linked in repulsion where the F2 $r$ values do not correlate well with the RI (Pearson correlation coefficient = 0.60). The extended line of points at $r = 0.049$ in the F2 corresponds to a range of $r = 0.375$ in the RI. This is due to the way the mapping program analyzes the data to accommodate the situation in which no double recessive homozygotes are seen in the F2; a minimum value is used when calculating recombination values from the maximum-likelihood tables and formulas of Allard (1956). Removing the 175 points corresponding to these low $r$ values in the F2 reduced the Pearson correlation coefficient for the repulsion data to 0.36.

Regression analysis suggests that $n_0 = (0.01 \pm 0.005) + (0.83 \pm 0.02) n_1$. Thus there is a systematic difference between the $r$ values for the two populations, such that there is $\sim 20\%$ less recombination in the RI than in the F2; presumably this reflects some environmental component affecting the overall chiasma/recombination frequency in the two crosses. The variation in $r$ values, for the coupling data, is consistent with the expected sampling error (Figure 1).

**Map construction for F2 and RI comparison:** Coupling and repulsion data were used to assemble linkage groups. For both the F2 and RI comparison data sets MM and JM propose 11 linkage groups at a LOD score $\geq 3$, with the same markers in each group. Adding together all
Marker numbers and the percentage with segregation distortion

<table>
<thead>
<tr>
<th>Marker type</th>
<th>Primer combination</th>
<th>Scored</th>
<th>Mapped in F₂ and RI</th>
<th>Distorted in F₂</th>
<th>( \chi^2 \geq 4 )</th>
<th>( \chi^2 &lt; 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tps</td>
<td>PPT/TaqAA</td>
<td>11</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>9</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>16</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AT</td>
<td>11</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>TA</td>
<td>11</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>TC</td>
<td>9</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>TG</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>13</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>GT</td>
<td>16</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td>102</td>
<td>62</td>
<td>5</td>
<td>8</td>
<td>(4.9%)</td>
</tr>
<tr>
<td>Spm</td>
<td>SPM/TaqATA</td>
<td>28</td>
<td>15</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>ATC</td>
<td>29</td>
<td>22</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>ATG</td>
<td>28</td>
<td>18</td>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>ATT</td>
<td>30</td>
<td>20</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td>115</td>
<td>75</td>
<td>13</td>
<td>14</td>
<td>(11.3%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>217</td>
<td>137</td>
<td>18</td>
<td>22</td>
<td>(8.3%)</td>
</tr>
<tr>
<td><strong>Tps and Spm</strong></td>
<td></td>
<td>18</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>(8.3%)</td>
</tr>
<tr>
<td><strong>Mapped in RI</strong></td>
<td></td>
<td>18</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>(8.3%)</td>
</tr>
</tbody>
</table>

Marker number obtained from each primer combination for the two types of element and the final number used for the F₂ and RI mapping comparison is indicated. The markers with segregation distortion in the F₂, classified as having a \( \chi^2 \geq 2 \), are expressed as a percentage.

11 groups the F₂ map is longer than the RI map, ∼710 and 565 cM, respectively (Table 1), consistent with the 20% reduction in the RI recombination fraction. In Figure 2 the linkage groups are identified according to the RI core map (Ellis et al. 1998). Placing the 11 groups into the expected 7 from the core map influences only linkage groups V–VII (Figure 2), interval differences and variation in length between the maps being immediately obvious. Linkage groups I–IV are intact in both maps and these groups have the most markers, whereas V–VII are each composed of more than one subgroup. The corresponding map length linking these subgroups, based on the core mapping data, is indicated in Figure 2. The order of markers is not identical for both maps with differences indicated in Figure 2 by lines and asterisks.

**Markers with segregation distortion:** Of the 217 Tps1 and Spm markers, 18% (40) are distorted in the F₂, and for 10% the distortion is extreme (Table 2). Of these 40 markers most segregate ∼1:1 in the RI population and are identified by primer combination in Table 2. These 40 markers were split into two groups, 18 with a \( \chi^2 \) from 2 to <4 and 22 with a \( \chi^2 \geq 4 \) \((P < \sim 0.05)\) in the F₂ (Table 2). This criterion was chosen because \( \chi^2 \sim 2 \) corresponds to \( P \sim 0.1 \), a stringent test for F₂ mapping of dominant markers. In the core RI map 36 of the 40 were mapped. The data relating to the 22 markers with more extreme distortion in the F₂ are summarized and compared to the RI data in Table 3. Table 3 also shows the linkage groups to which these markers have been assigned. Four markers are not mapped in the core RI map, i.e., S92m, S71p, S40p, and these show segregation distortion in both populations. Figure 2 shows the corresponding regions where the 18 mapped markers with extreme distortion were positioned on the core RI map.

**Analysis of the core JI15 × JI399 RI mapping data:** The core map data set consists of 347 markers; for 89 RI lines, there are 13,675 JI15 alleles and 13,678 JI399 alleles. The excess number of JI15 or JI399 alleles for each of the 89 RI lines is plotted in Figure 3. The extremes are RI line 88 having 158 JI399 alleles.

An identical subset of 80 RI lines was selected from maps A and B (Table 1) for analysis of the distribution of recombination events using an approach similar to the “graphical genotypes” of Young and Tanksley (1989) and reviewed in Rieseberg et al. (2000). This
subset was chosen to minimize missing data; >90% of
this subset was scored for >80% of the markers. For
the 560 chromosomes the crossover number per line is
illustrated in Figure 4a and the corresponding fre-
quency distribution is shown in Figure 4b. There are
33 nonrecombinant linkage groups; Figure 4c shows the
allocation and frequency distribution of these nonre-
combinant linkage groups, and they derive from 30 RI
lines. Three RI lines (18, 32, and 87) each have 2 paren-
tal linkage groups. JI15 and JI399 are represented in 18
and 15 nonrecombinant linkage groups, respectively
(Figure 4c).

Crossover analysis for this subset also highlights close
double recombination events (where a single marker
or a pair of markers, for a given RI line, is of different
parental origin compared to the consensus of the flan-
ing markers). Close double crossovers are present in maps
A and B, and these data are summarized in Table 4.

DISCUSSION

Molecular markers, recombination frequency, and
map comparison: Dominant markers are not ideal for
the generation of F2 maps because they fail to identify
heterozygotes. However, in this study they have facil-
tated the comparison of F2 and RI mapping data. If
selfing is taken into account according to Haldane
and Waddington (1931), then we would expect the
recombination fraction to be the same for both popula-
tions if the influence from environmental effects is ab-
sent. Figure 1 shows this is not the case as there is a
scatter of points, most pronounced in the repulsion
phase. Two major influences contribute to the scatter.
The first comes from the reduced information content
of dominant markers for F2 mapping, especially in the
repulsion phase, and the second is due to sampling
error inherent to the data sets.

Regression analysis suggests that there is a consistent
difference between the two data sets, with ~20% less
recombination in the RI; this presumably reflects envi-
ronmental effects on meiosis and recombination that
differed between the populations that were generated
independently. These influences on r are reflected in
the difference of 145 cM of map length between the F2
and RI. Differences in marker order (Figure 2) presum-
ably derive from the different data sets and reflect sam-
ping variation. The summation of errors from individ-
ual interval distances for each map will contribute to
map differences. The fractional error in r will be the

\[
\frac{1}{\sqrt{n}} \quad \text{or} \quad \frac{2}{\sqrt{n}}
\]

Figure 1.—Comparison of recombination frequencies in F2
and RI populations. For the cross JI15 × JI399, the observed
recombination frequency for the pairwise combination of
markers for the F2 population (ordinate) is plotted against
the corresponding observed recombination frequency in the
RI population (abscissa). These data are plotted separately
for the markers linked in coupling, repulsion, and for both
phases combined. The recombination frequencies were calcu-
lated for all pairwise combinations of 137 markers (Table 1),
and the data were plotted where a value of <0.4 was observed
in both populations. Lines are drawn showing predicted stan-
dard errors on the estimates of recombination frequencies.
The top lines are +1 (a) and +2 (c and d) standard errors
in the F2 recombination frequency, given the recombination
frequency in the RI. The bottom lines are −1 (b) and −2 (e)
standard errors in the RI recombination frequency, given the
F2 recombination frequency. In the combined plot, line c refers
to the repulsion phase and d to the coupling phase. For the F2,
standard errors were calculated for a population size of 100
individuals according to Allard (1956). For the RI population
standard errors were estimated as \( \sqrt{\frac{1 - r}{n}} \), where \( n = 100 \).
Figure 2.—Comparison of the F2 and RI (Table 1, 137-marker) maps showing relative lengths; regions where marker orders disagree between the two maps are indicated by lines, depicting a one-to-one difference, and an asterisk shows more than one difference. For linkage groups V–VII the numbers (centimorgans) indicate the approximate map length separating the subgroups, joined by dotted lines, on the basis of the core (Table 1, 347-marker) RI data set. Brackets to the right of the RI linkage groups show the regions where markers with extreme segregation distortion are located in relation to the core RI map; in some cases, e.g., for groups IV, VI, and VII, the brackets are external to the 137-marker subset RI map (Table 1).

same as the fractional error in the sum of all the interval distances.

Despite the disadvantage of dominant markers the mapping programs, MM and JM, have independently grouped the markers identically into 11 linkage groups for both the F2 and RI data sets. These 11 groups were ordered simply in relation to the core pea RI map. Linkage groups I–IV are intact for both maps, whereas linkage groups V–VII, for both maps, are composed of more than 1 group (Figure 2). This trend is reflected in the RI core linkage map (Table 1), where most of the markers belong to groups I–IV, generating fairly dense linkage groups compared to groups V–VII, which are generally more sparse. This was not the case for the 240-marker RAPD map for pea (Table 1), where linkage groups I and III were made up from 2 separate groups when compared to the core linkage map (Table 1).

**Segregation distortion and its effect on gametes and F2 ratios:** Markers exhibiting segregation distortion were split into two groups: those with \( \chi^2 \) in the range \( \geq 2 \) to \( <4 \) (18 markers) and those with \( \chi^2 \geq 4 \) (22 markers), of which 4 were not mapped in the RI population (Table 3). In both groups there was an approximately equal representation of JI15 and JI399 alleles (Table 3). From Table 2 it appears that almost twice as many Spm markers are affected by segregation distortion compared to Tps markers, but this difference is not statistically significant.

There is a range of types of segregation distortion in an F2 population; these include deviation from the 1:1 allele frequency and deviation from the expected heterozygote frequency. A range of distorted marker types is represented in F2 mapping studies with isozyme loci for pea (Weeden and Marx 1987), with codominant and dominant markers for alfalfa (Kiss et al. 1993), and with RFLP and isozyme loci for common bean (Paredes and Gepts 1995). These types are difficult to distinguish when using dominant markers because the heterozygotes and one homozygote class cannot be differentiated. However, in a recombinant inbred population domi-
TABLE 3

Excess heterozygosity and ratio distribution

<table>
<thead>
<tr>
<th>Marker</th>
<th>Linkage group</th>
<th>F2</th>
<th>RI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA + Aa</td>
<td>aa</td>
<td>Ratio</td>
</tr>
<tr>
<td>S33m</td>
<td>V</td>
<td>71</td>
<td>44</td>
</tr>
<tr>
<td>S108m</td>
<td>V</td>
<td>72</td>
<td>43</td>
</tr>
<tr>
<td>T140m</td>
<td>V</td>
<td>72</td>
<td>42</td>
</tr>
<tr>
<td>T28m</td>
<td>V</td>
<td>76</td>
<td>42</td>
</tr>
<tr>
<td>S43m</td>
<td>VII</td>
<td>78</td>
<td>40</td>
</tr>
<tr>
<td>S111m</td>
<td>IV</td>
<td>94</td>
<td>19</td>
</tr>
<tr>
<td>S15p</td>
<td>VI</td>
<td>95</td>
<td>18</td>
</tr>
<tr>
<td>T80m</td>
<td>II</td>
<td>96</td>
<td>18</td>
</tr>
<tr>
<td>S11m</td>
<td>II</td>
<td>98</td>
<td>18</td>
</tr>
<tr>
<td>T164p</td>
<td>VII</td>
<td>96</td>
<td>18</td>
</tr>
<tr>
<td>S16m</td>
<td>VII</td>
<td>95</td>
<td>18</td>
</tr>
<tr>
<td>S23m</td>
<td>I</td>
<td>101</td>
<td>17</td>
</tr>
<tr>
<td>T79p</td>
<td>II</td>
<td>98</td>
<td>16</td>
</tr>
<tr>
<td>S24p</td>
<td>VII</td>
<td>102</td>
<td>16</td>
</tr>
<tr>
<td>S6p</td>
<td>II</td>
<td>99</td>
<td>14</td>
</tr>
<tr>
<td>T161m</td>
<td>VII</td>
<td>97</td>
<td>14</td>
</tr>
<tr>
<td>S113p</td>
<td>VII</td>
<td>109</td>
<td>9</td>
</tr>
</tbody>
</table>

Unmapped markers

- T125p — | 50 | 52 | 1:1 | 48 | 31 | 2:1 |
- S92m — | 103| 11 | 9:1 | 61 | 26 | 2:1 |
- S71p — | 110| 8  | 13:1| 70 | 19 | 4:1 |
- S40p — | 113| 5  | 23:1| 73 | 16 | 5:1 |

Marker ratios are scored in the F2 and RI populations for the set of 22 with extreme distortion ($\chi^2 \geq 4$; Table 2) and the linkage groups to which the 18 mapped in the RI core map. The classifications of the marker types are $d$, the deficient heterozygote class, and $e$, the excess heterozygote class in the F2.

Map Expansion in Pea

Heterozygote advantage in the F2 seed could account for this, and seed abortion in the F2 was noted for this cross (Lee et al. 1988). The process leading to heterozygote excess is not known but the characteristic is not uncommon in plant species (Weeden and Marx 1987; Kiss et al. 1993; Liu et al. 1994; Paredes and Gepts 1995; Kearsey et al. 1996; Faris et al. 1998; Kaló et al.

![Figure 3](image-url)
Almost one-third of the isozyme markers in lentil (Zamir and Tadmor 1986). This was also the case for two RFLP markers in interspecific rice crosses mapping to chromosomes 10 and 11 in an F2 population (Xu et al. 1997).

The assessment of the ratio of homozygotes to heterozygotes within an F2 population cannot be made when using dominant markers. The expectation, based on Mendelian segregation for a monohybrid cross, is that the ratio will approximate to a 3:1 (AA / Aa:aa) corresponding to a 1:2:1 (AA:aa), ratios deviating from this having segregation distortion. On inbreeding, markers of the e and d classes will achieve a 1:1 ratio in an RI population. For example, they could segregate in a Mendelian manner after the F2; i.e., the heterozygotes are reduced by one-half after each successive meiosis until all loci are of one parental allele or the other. Alternatively the distorted heterozygote ratio may be maintained at each meiosis throughout the generations of inbreeding or in a genotype-specific pattern. These heterozygote-reducing mechanisms give different rates at which homozygosity would be reached but all generate homozygous RI lines. Lister and Dean (1993) reported 0.42% heterozygosity remaining at the F8 generation in A. thaliana. In an intraspecific cross of tomato, 2.1% heterozygosity remained at the F6:F7 generation (Saliba-Colombani et al. 2000), compared to 15% remaining in an interspecies cross at F7 (Paran et al. 1995). Residual heterozygosities of 1.6 and 2.7% were found in two maize F10 populations (Burr and Burr 1991). One potential cause of residual heterozygosity could be that many parental lines are not completely homozygous.

**Excess heterozygosity and map expansion:** The theory of the process by which excess heterozygosity leads to map expansion is based on the relationship between Mendelian segregation ratios, recombination frequency, and crossover number. This relationship is the fundamental principle on which linkage mapping is based (Mendel 1866; Morgan 1911; Sturtevant 1913; Haldane and Waddington 1931). The persistence of heterozygotes predicted by Haldane and Waddington (1931), for different crosses, has been taken into account in genetic mapping in animal and plant systems (Taylor 1978; Burr et al. 1988; Burr and Burr 1991).

In a given meiosis four possible gametes can be derived from one chiasma exchange between four chromatids. Two of these will be parental and two will be recombinant for markers either side of the chiasma. Thus one chiasma corresponds to 0.5 recombinants or 50 cM. For each marker in an F2 population it is expected that the segregation will be 1:2:1, but, if this ratio is distorted to 1:h1 (where h > 2), then the excess number of chiasma in the F2 is [2h/(h + 2)] − 1. So, for example, if the segregation ratio is 1:3:1, then the fraction of heterozygotes is 0.6 and thus the proportion of recombinants will be increased by a factor of 1.2, i.e., an excess of 0.2 chiasma. Therefore 10 such segments are equivalent to 2 extra chiasmata or 100 cM more added to the map.

**Figure 4.—** A comparison of crossover number based on 560 chromosomes (80 RI lines and 7 linkage groups) for the two maps A and B using the “core” data set (Table 1). (a) The number of crossovers for each of the 80 RI lines (sorted in ascending order according to the values for map A). (b) The crossover frequency distribution. (c) The number and distribution of the 33 nonrecombinant linkage groups (zero position in b) shown in relation to the linkage group (Roman numerals) and the equivalent chromosome (Arabic numbers) to which they belong.
length. Where the excess heterozygote ratio persists during the development of an RI population the effect on genetic map length will be even greater. If the recombination fraction in a small interval is \( r \) in the \( F_2 \) then in the RI the proportion of observed recombinants will be \( r(h + 2)/2 \). This suggests that map expansion, resulting from the persistence of heterozygotes, is misrepresented by the map expansion function of Haldane and Waddington (1931) when \( F_2 \) segregation ratios are distorted.

The chiasma counts are \( \sim 16 \) each in the Ji15 and Ji399 parental lines and \( \sim 14 \) in the \( F_1 \) hybrid of Ji15 \( \times 399 \) (Hall et al. 1997b). If we take 15 as average, the expected map length is \( 15 \times 50 = 750 \) cM (Table 1). In this study, 16% of markers used in generating the RI map showed distortion in the \( F_2 \) but not the RI population (Table 3). Figure 5 shows the effect this 16% has on the distribution of excess chiasma among the linkage groups and adds 447 cM (8.94 \( \times 50 \) cM) of extra map length. Maps A and B are 990 and 1800 cM, respectively (Table 1), so the additional map length due to segregation distortion is more than enough to account for the excess length of map A. In the case of linkage group V there is a reduced number of chiasma (discussed later). Even a slight deviation from a 1:2:1 ratio when mapping in the \( F_2 \) will alter map length. This analysis focuses on excess heterozygosity and its effect on map expansion but markers with deviations from the expected 1:1 allele ratio may also deviate in heterozygote ratio and so will have an impact on map length. Markers with skewed ratios found in \( F_2 \) mapping are generally reported as such (Kiss et al. 1993; Xu et al. 1997; Faris et al. 1998; Kaló et al. 2000).

**Nonrecombinant linkage groups and heterozygote deficiency:** If there is a deficiency of heterozygotes, then fewer recombination events than expected would be observed, and map length would be correspondingly reduced. One consequence of this would be the observation of more nonrecombinant linkage groups than expected.

In the RI data set one-sixteenth (33) of the total number of linkage groups were wholly parental. Selfing an \( F_1 \) gives a one-in-four chance that one of the two parental gametes is nonrecombinant in one meiosis; thus the chance of forming a zygote with a nonrecombinant linkage group is \( 1/16 \); this ignores effects in subsequent generations.

The 33 nonrecombinant linkage groups are not distributed evenly; linkage groups II and V have two-thirds of the share with 10 each (Figure 4c). Linkage group V has the shortest map in this cross. It is known from male meiotic analysis in pea that chromosome 3 (equivalent to linkage group V) persistently forms rod bivalents in this cross, *i.e.*, with a crossover in only one arm (Hall et al. 1997a). This correlates with the reduced chiasma numbers found in linkage group V from our segregation distortion analysis and the presence of four markers of

| Linkage group | Marker 1 | 2 | 11 | 13 | 17 | 37 | 38 | 43 | 45 | 50 | 57 | 61 | 63 | 69 | 72 | 81 | 82 | 90 | 91 | 93 | 98 | 99 |
| I             | NA      | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| II            | T77p    | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| III           | NA      | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| IV            | S11t    | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| V             | S108m   | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| VI            | S15p    | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| VII           | T161p   | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| VIII          | S119p   | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| IX            | S24p    | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| X             | T162p   | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|
| XI            | S21    | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|

**TABLE 4**

Close double-recombination events

| RI line | Marker 1 | 2 | 11 | 13 | 17 | 37 | 38 | 43 | 45 | 50 | 57 | 61 | 63 | 69 | 72 | 81 | 82 | 90 | 91 | 93 | 98 | 99 |
| DR      | DR      | DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR| DR|

DR: RI lines where there is a close double-recombination event for markers with extreme segregation distortion (Table 3). These double recombinants cannot be removed from the core map by local changes to marker order without generating additional problems and are found in both the A and B maps.
Figure 5.—The distribution of excess chiasmata among linkage groups and the extent to which the 16% of distorted markers contribute to map length. The magnitude of the effect is plotted as “excess chiasmata”; this is calculated from the F2 segregation ratio (1:4:1). The value of excess chiasmata is \[2k/(h + 2) - 1\] (where \(k\) is the heterozygote proportion).

The one other marker with a deficiency of heterozygotes in our study was found to map to the bottom end of linkage group VII. A deficiency of heterozygotes was also found in another pea study, using isozyme markers and a RI population, are also susceptible to map length variation from markers with segregation distortion (Xu et al. 1997; Virk et al. 1998). Xu et al. (1997), comparing different mapping populations in rice, found that a DH population consistently gave more segregation distortion than the F2 for the same cross, suggesting something inherent in the creation of the DH population causing an increase in segregation distortion. In one case on linkage group 10 (Xu et al. 1997) an RFLP marker in the F2 with excess heterozygosity corresponded to a region of excess indica alleles in the doubled-haploid population. This could come about by selection for the indica allele in the male meiotic products and for the japonica allele in the female meiotic products, which may generate excess heterozygotes as a mechanism independent of heterozygote advantage.

Sex-dependent effects: Disparity between male and female recombination frequencies and chiasma count expectations may also contribute to map length difference, as linkage maps represent an average of crossovers from male and female meioses. Male and female chiasma frequency and distribution behavior have been studied in a range of plant taxa (Koul et al. 2000), differences being species specific. In humans, female recombination events, for a subset of markers with extreme segregation distortion and shows the RI lines and linkage groups involved. Presumed close double-crossover events can come about in various ways. Knox and Ellis (2001) suggested that changes in PstI methylation states were responsible for some apparent close double crossovers, assessed by AFLP analysis, and that this led to map expansion. However, in the present study excess heterozygosity and crossovers in earlier generations are also seen to contribute. RI lines 69, 82, and 91 each have >35 crossovers for all seven linkage groups in both the A and B maps. Table 4 shows that these three RI lines have close double recombination for more than one marker in both A and B maps (these are not scoring errors). Additionally on all linkage groups of the core mapping data there are markers, codominant and dominant, other than those in Table 4, which display close double recombination events in these three RI lines (data not shown). The most extreme is RI line 91, having seven independent close double-recombination events in linkage group III. As these events are potentially a consequence of excess heterozygosity in early generations, aiming to minimize them would reduce map length at the expense of marker order. Map length is important but correct marker order is of more benefit for genome analysis and marker-assisted breeding. In A. thaliana it was found that the order for a set of markers on chromosome 4 from physical mapping disproved the suggested genetic map order even with a high likelihood value from the mapping program (Lister and Dean 1993).

Both BC and DH mapping populations, raised to obtain a 1:1 ratio in a shorter time than it takes to obtain an RI population, are also susceptible to map length variation from markers with segregation distortion (Xu et al. 1997; Virk et al. 1998). Xu et al. (1997), comparing different mapping populations in rice, found that a DH population consistently gave more segregation distortion than the F2 for the same cross, suggesting something inherent in the creation of the DH population causing an increase in segregation distortion. In one case on linkage group 10 (Xu et al. 1997) an RFLP marker in the F2 with excess heterozygosity corresponded to a region of excess indica alleles in the doubled-haploid population. This could come about by selection for the indica allele in the male meiotic products and for the japonica allele in the female meiotic products, which may generate excess heterozygotes as a mechanism independent of heterozygote advantage.
Map Expansion in Pea

chiasma counts were almost double that for male (Hultén 1998), corresponding to an approximate doubling of map length; this is in concordance with the human female genetic map that was found to be 66% longer than the male map (Broman et al. 1998). Segregation distortion was also reported in plant studies (De Vincente and Tanksley 1991; Van Ooijen et al. 1994; Busso et al. 1995; Kearsey et al. 1996; Faris et al. 1998), with markers in both female- and male-derived populations showing excess heterozygosity. It is not known whether there are major differences in recombination frequency in male vs. female meiosis in pea.

The genetics of distortion: The 36 markers distorted in the F2 are distributed among the seven linkage groups of the RI map. Fifteen of the 18 markers with extreme distortion show a tendency to cluster, in particular on linkage groups II, V, and VII. Distorted segregation ratios have been reported in relation to the esterase (Est) isozyme loci in barley (Konishi et al. 1990), rice (Lu et al. 2000), pea (Weeden and Marx 1987), alfalfa (Kiss et al. 1993), and for lentil and tomato (Zamir and Tadmor 1986). In addition, the Est loci have been shown to have strong association with gametophyte factor or sterility genes loci (Konishi et al. 1990, 1992; Lu et al. 2000).

In pea Est loci have also been identified on linkage groups II, V, and VII (Weeden et al. 1998) and so are associated with the extreme segregation distortion described in this study. The region on linkage group II with 10 markers exhibiting distortion, 5 of them extreme, spans an Est locus. Similarly for linkage group V an Est locus maps close to the 6 distorted markers; 4 of these are extreme. In linkage group VII the 8 markers with distortion, 6 extreme, are spread in three tight clusters distally and centrally with an Est locus between two of the clusters. These regions may well include loci in relation to gametophytic competition, for example, early pollen tube emergence (Konishi et al. 1990), in particular for linkage groups II and V that have above average representation of nonrecombinant linkage groups. In addition to the Est loci the mapping of other isozymes in a pea F2 study (Weeden and Marx 1987) resulted in 4 of 6 markers with a heterozygote deficiency and 7 of 7 with heterozygote excess mapping to linkage groups II, V, and VII.

CONCLUSION

Segregation distortion is a characteristic of most marker and mapping systems and has a major effect on map length and marker order. We have shown that excess heterozygosity can lead to genetic map expansion especially in RI mapping populations. For our F2 and RI map comparison, 137 of the dominant markers used in this study segregated with a Mendelian ratio in both populations and did not have segregation distortion. In F2 maps markers with excess heterozygosity, but normal allele ratios, will tend not to be used because of their distortion. Thus segments of the genetic map that exhibit this type of distortion may be missing from F2 maps. This will tend to inflate RI map lengths with respect to F2 maps as the segregation ratios are normal in the RI and the markers will be included. In RI mapping, algorithms that minimize predicted close double recombinants might misrepresent the order of markers in regions with excess heterozygosity because these regions contain an unexpectedly high frequency of recombinants. This is because the opportunities for observed recombination in the generation of RI populations are underestimated in regions of excess heterozygosity. Excess map length is discomforting, but the excess is from an expectation that assumes that markers are immune from segregation distortion, unexpected heterozygosity, and nongenetic effects such as DNA methylation.

The authors thank D. Laurie, G. Moore, J. Snape, and G. Hewitt for useful discussions and comments on the manuscript. The research reported in this article is part of a Ph.D. thesis at the John Innes Centre, which is supported by a grant in aid from the Biotechnology and Biological Sciences Research Council, United Kingdom. We also acknowledge the support of the European Union funded projects MEDICAGO (QLR2-CT-2000-30676) and TEGERM (QLKCT-2000-10152).

LITERATURE CITED


HALDANE, J. B. S., 1919 The probable errors of calculated linkage values, and the most accurate method of determining gametic from certain zygotic series. J. Genet. 8: 291–297.


JIN, W., R. G. PALMER, H. T. HORNER and R. C. SHOEMAKER, 1999 Fr1 (root fluorescence) locus is located in a segregation distortion region on linkage group k of soybean genetic map. J. Hered. 90: 553–556.


Map Expansion in Pea


Communicating editor: J. A. Birchler