Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers

MATS HANSEN1*, THOMAS KRAFT1, SARA GANESTAM1, TORBJÖRN SÄLL2 AND NILS-OTTO NILSSON1

1 Novartis Seeds AB, PO Box 302, S-261 23 Landskrona, Sweden
2 Department of Genetics, Lund University, Sölvegatan 29, S-223 62 Lund, Sweden

(Received 7 April 2000 and in revised form 11 August 2000)

Summary

The possibility of using linkage disequilibrium mapping in natural plant populations was assessed. In studying linkage disequilibrium among 137 mapped AFLP markers in four populations of sea beet (Beta vulgaris ssp. maritima (L.) Arcang.) it was shown that tightly linked loci could be detected by screening for associations. It was hypothesized that the short distances spanned by linkage disequilibrium enable markers that are very tightly linked to a target gene to be identified. The hypothesis was tested by whole-genome screening of AFLP markers for association with the gene for the annual growth habit, the B gene, in a sample of 106 sea beets. Despite the dominant nature of AFLP, two markers showing significant linkage disequilibrium with the B gene were detected. The results indicate the potential use of linkage disequilibrium for gene mapping in natural plant populations.

1. Introduction

The most common method for mapping plant genes involves the crossing of two lines to produce an F1, which is further crossed in any of a number of different ways to produce segregating offspring. This segregating population is used to estimate the recombination frequencies between the marker loci and the gene of interest. However, such a strategy is time-consuming since it requires the growth of several generations before linkage analysis can be performed. In addition, the map resolution is dependent upon the number of segregating offspring produced. For small recombination frequencies in a mapping population of average size, very few recombinants are expected. This reduces the accuracy of recombination frequency estimate for short map distances.

In contrast, in linkage disequilibrium mapping, natural populations rather than segregating offspring are sampled, markers being screened for associations with the gene of interest. Such a strategy allows recombination events from all generations in the genealogical history of the sample that have occurred since the linkage disequilibrium was created to be sampled (Hästbacka et al., 1992; Houwen et al. 1994). Since significant linkage disequilibrium is unlikely to remain for many generations, except in the case of tightly linked markers, the power to accurately resolve the order of tightly linked loci is thus greatly improved compared with linkage mapping involving the sampling of recombination events from one generation only. The method just described is widely used for the identification of markers that are tightly linked to human disease genes (de la Chapelle, 1993; Lander & Schork, 1994; Jorde, 1995; Peltonen & Uusitalo, 1997; Sheffield et al., 1998). Although such high resolution would appear useful in many applications, linkage disequilibrium mapping in plants has not been explored to a great extent.

Linkage disequilibrium mapping is preferentially used for traits involving simple inheritance, i.e. traits with a small environmental variance that are determined by single genes of high penetrance. Traits for which inheritance is more complex are difficult to map by use of this method (Lander & Schork, 1994; Peltonen & Uusitalo, 1997; Barcellos et al., 1997; Kruglyak, 1997; Chapman & Wijsman, 1998). A suitable trait for testing linkage disequilibrium mapping in plants is that of the growth habit in beet (Beta...
vulgaris L.), determining whether the beet plant requires a vernalization period at low temperature prior to bolting, flowering and seed setting. Plants requiring no vernalization are annual ones in which bolting and flowering occur the same year as germination, whereas plants that require vernalization are biennial ones that start to bolt and flower in their second year. The term annual and biennial, although commonly employed, are not strictly correct, since both types, if they survive, can flower over a period of several years. The polymorphism of the growth habit is determined by a single gene, known as the bolting gene or the B gene (van Dijk & Boudry, 1991; Abe et al., 1993, 1997a, b; Boudry et al., 1994). Since the allele for the annual growth habit, B, is dominant, biennial plants are homozygous for the recessive allele b (Boudry et al., 1994; Abe et al., 1997a).

Sea beet (Beta vulgaris ssp. maritima (L.) Arcang.), the wild relative of sugar beet (B. vulgaris ssp. vulgaris L.), is distributed along the coasts of the Mediterranean and the Atlantic coast of Europe. In the Mediterranean area most beets are annual, whereas populations in Great Britain and northern Europe are completely biennial. On the Atlantic coast of France, mixed populations consisting of both annual and biennial plants can be found (van Dijk & Boudry, 1991; van Dijk et al., 1997). This cline was probably created by natural selection, since summers in the north are too short for plants to germinate, bolt, flower and set seed during the same year (van Dijk & Boudry, 1991; van Dijk et al., 1997).

The present study represents a first step towards linkage disequilibrium mapping of a plant gene. The B gene was chosen as a model trait because of its relatively simple inheritance and the presence of polymorphic populations in sea beet. AFLP markers were used for studying the general map distances over which linkage disequilibrium is present and to screen the genome for associations with the B gene.

2. Materials and methods
(i) Plant materials
Seeds were collected from 106 individual plants from four natural populations of sea beet on the Atlantic coast of France (Table 1). A single plant from each seed parent was grown in a greenhouse with 22 h of daylight and a minimum temperature of 19 °C. Plants flowering after 18 weeks were considered to be annuals, not requiring vernalization. Plants showing no signs of flowering or bolting after 18 weeks were considered to be biennial, requiring vernalization before flowering (Table 1). Some plants bolted without showing any signs of flowering. Although van Dijk et al. (1997) classified such plants as biennials, two classification systems were tested here, the bolting, non-flowering plants being considered in the one system as annuals and in the other as biennials. DNA was isolated from freeze-dried leaves, using the method described by Halldén et al. (1996).

(ii) Linkage disequilibrium mapping
For the investigations reported, 440 polymorphic AFLP markers amplified by 17 primer combinations were scored in the four natural sea beet populations. The AFLP procedure used has been described by Nilsson et al. (1999). Three different analyses were performed. First, a subset of 137 markers that had previously been mapped in sugar beet (Nilsson et al., 1999) was used to assess the relation between map distance and linkage disequilibrium in sea beet. Second, all markers were tested for association with the B gene in the sea beet populations. Third, four markers located within a 1-5 cM segment containing the B gene were analysed for association with B in each of the sea beet populations. Linkage to the B gene for the four markers was identified by screening 192 AFLP primer combinations by means of bulked segregant analysis (BSA) (Michelmore et al., 1991). A sugar beet F2 population of 194 individuals was used for the construction of one bulk of nine annual plants and two bulks of five biennial plants each. Candidate markers were confirmed in 11 bulks of two biennial plants each and subsequently mapped in the F2 population using the JoinMap software (Stam & Van Ooijen, 1995).

(iii) Data analysis
The data presented here do not permit estimates of linkage disequilibrium in a strict sense, due to the dominant nature of the data, both for the markers and for the B gene. Instead, associations between phenotypes at pairs of markers or between markers and B were tested, using Fisher’s exact test (Lewontin, 1995; Weir, 1996). Since associations between a recessive allele at one locus and a dominant allele at another are very difficult to detect, one-sided P values were calculated for the hypothesis of a positive association being found between dominant alleles at both loci.

3. Results
(i) Relation between linkage and linkage disequilibrium
The map distances spanned by linkage disequilibria in the sea beet populations were investigated using 137 AFLP markers previously mapped in sugar beet (Nilsson et al., 1999). All markers were first tested for the homogeneity of allele frequencies over the populations. The four populations were found to be similar
Table 1. Location, number of plants analysed and proportion of biennial plants (including bolting, non-flowering plants) for the four sea beet populations used in the study

| Population | A          | B          | C          | D          |
|------------|------------|------------|------------|------------|
| Location   | Talmont    | La Rochelle| La Baule   | Les Sables-d’Olonne |
|            | 45° 32' 32" N | 46° 18' 10" N | 47° 16' 14" N | 46° 25' 72" N |
|            | 0° 54' 34" W | 1° 07' 74" W | 2° 27' 96" W | 1° 37' 78" W |
| No. of plants | 48         | 21         | 22         | 15         |
| Frequency of biennial plants | 67%        | 71%        | 55%        | 47%        |

Table 2. Tests for linkage disequilibrium between pairs of AFLP marker loci

| P value | Linkage (cM) |
|---------|--------------|
| < 1 x 10^-3 | 0            |
| < 1 x 10^-3 | 0            |
| < 1 x 10^-3 | 0            |
| < 1 x 10^-3 | 0            |
| < 1 x 10^-3 | Unlinked      |
| 1.4 x 10^-3 | 0            |
| 6.1 x 10^-3 | Unlinked      |
| 7.2 x 10^-3 | Unlinked      |
| 1.2 x 10^-4 | 3.4           |
| 1.3 x 10^-4 | 0             |
| 4.4 x 10^-4 | 2.7           |
| 6.0 x 10^-4 | Unlinked      |
| 7.4 x 10^-4 | Unlinked      |
| 1.0 x 10^-3 | Unlinked      |

P values for the hypothesis of no association between loci, as well as distances between loci in the sugar beet linkage map of Nilsson et al. (1999), are shown.

for most markers (data not shown), suggesting that they can be considered as a single, large population. The data from all populations were thus pooled, except for four markers with significant heterogeneity (P < 0.01) in allele frequencies among the populations. These markers were discarded from the data-set to prevent spurious associations between markers due to any substructuring of the population. The presence of linkage disequilibrium was then tested for all pairs of the remaining 133 AFLP markers, resulting in 8778 tests. The vast majority of tests were between unlinked loci, only 459 tests being between markers located within 5 cM of each other as based on the sugar beet map (Nilsson et al., 1999). For tests having P values of less than 1 x 10^-3 the situation was the opposite, the majority (8 of 13) of them involving markers closely linked in sugar beet (Table 2). Pairs of tightly linked markers are thus overrepresented among the cases of strong allelic association (P = 6.8 x 10^-8 that at least 8 of 13 pairs are linked by chance, given that in total 459 or 8778 are linked). These significant linkage disequilibria are not due to population admixture, since the P values for single populations were also low (data not shown).

The very strong linkage disequilibria detected for a few pairs of markers unlinked in sugar beet is highly surprising, even when the large number of tests is considered. A likely explanation is that at least one of the bands in each marker pair is non-homologous between sea beet and sugar beet. It has been shown that AFLP bands of equal size are not always homologous, even within the sugar beet (Hansen et al., 1999). Hence, these scored bands most likely represent loci that are linked in sea beet. In conclusion, the present results indicate that screening for linkage disequilibria can be used to detect linkage between closely linked loci.

(ii) Screen of anonymous markers for linkage disequilibrium with B

The utilization of linkage disequilibrium for de novo mapping of disease genes in humans has been described previously (Houwen et al., 1994), although the number of studies actually employing that strategy is still rather small. In the present experiment, the possibility of performing linkage disequilibrium mapping directly in natural plant populations was assessed. Screening of 440 AFLP loci in the four sea beet populations was performed in order to identify associations between markers and the B allele. P values were calculated for all marker-B pairs for the two different phenotype classification systems. Testing for heterogeneity among populations in allele frequencies at the B gene resulted in high P values: 0.70 and 0.79 for phenotype classification 1 and 2, respectively. The data from all populations were therefore pooled, which identified two markers with very low P values (Table 3). Marker 1 (E16M16:38) had P values of 9.7 x 10^-5 for phenotype classification 1 (all the bolting plants are annuals) and 8.6 x 10^-4 for phenotype classification 2 (only the flowering plants are annuals) whereas marker 2 (E11M02:554) had P values of 0.25 and 8.7 x 10^-5 for phenotype classi-
markers were tested for association with polymorphic in the sea beet populations. These
surrounding the 1
B
be to test for associations with identifying the markers most tightly linked to 672300004857
When associations were studied between the two markers and the B gene separately in the different populations, the P values varied. However, some P values were below 0.05, with one case as low as 0.0012, and even when the P values were higher the association was always in the same direction, with an excess of annuals with the marker band and biennials lacking the marker band. Thus, the lower P values resulting from pooling the populations is not due to heterogeneity among the populations, but rather due to increased statistical power because of larger sample size.
Permutations were used to determine the significance of the observed P values (Long et al., 1998; Long & Langley, 1999). First, all individuals were randomized with respect to the phenotype data. Then, P values between all markers and the B gene were calculated. This process was repeated 10000 times, the lowest P value being noted for each replicate. The frequency of minimum P values below a certain value was then taken as the probability of observing a P value as low as this when no association exists. The actually observed lowest P values for E16M16:38 and E11M02-554 correspond to probabilities of 0.020 and 0.022, respectively.

Data for the phenotypic classification giving the lowest P values are shown.

| E16M16:38 in population | E11M02-554 in population |
|------------------------|--------------------------|
| A  B  C  D  Total      | A  B  C  D  Total      |
| 1–B 5 12 6 7 30       | 1–B 8 4 9 3 24       |
| 1–b 0 0 2 2           | 3 2 4 2 11           |
| 0–B 16 4 8 0 28       | 1 2 3 4 10           |
| 0–b 14 1 6 2 23       | 13 12 3 4 32         |

4. Discussion

The present study is an attempt to use linkage disequilibrium for the mapping of a plant gene. We have shown that it is possible to identify closely linked markers by screening for associations in natural populations. Considering the easy access to natural populations of many species, linkage disequilibrium mapping should be an interesting alternative to conventional linkage mapping. Due to its high efficiency in marker generation, AFLP could provide a suitable system for this type of analysis. One drawback of it, however, is the dominant nature of the markers. Our analysis of linkage disequilibrium between mapped marker loci shows that significant associations between markers are correlated with tight linkage. This demonstrates the feasibility of detecting linkage between AFLP markers by studying linkage disequilibrium in natural populations.

In employing a limited number of AFLP markers (440) in the natural sea beet populations of 106 plants altogether, we were able to identify two markers highly associated with the allele for the annual growth habit. This is an example of the de novo targeting of a plant gene by use of linkage disequilibrium mapping. Use of linkage disequilibrium for the fine-mapping of markers already known to be linked to human disease...
genes has been described (de la Chapelle, 1993; Lander & Schork, 1994; Jorde, 1995; Peltonen & Uusitalo, 1997; Sheffield et al., 1998). We employed this approach in sea beet by testing for associations between the B gene and four markers that were identified by bulked segregant analysis in sugar beet. Given the accuracy of the map, all markers were located within 1-5 cM of the B gene. Three markers, all mapping to the same position in sugar beet, were polymorphic in the sea beet populations, whereas the fourth marker was uninformative in all populations. For one of the markers, the association with B is much stronger than for the other two, indicating there to be a tighter linkage between this marker and B. Taken together, these results are encouraging for future attempts to use AFLP and linkage disequilibrium for the mapping of plant genes.

One application that can directly benefit from the improved resolution that linkage disequilibrium mapping provides is the map-based cloning of genes by use of a chromosome-landing strategy (Tanksley et al., 1995; Cnops et al., 1996; Schwarz et al., 1999). A large number of segregating progeny is needed for a linkage map resolution compatible with the average insert size of a BAC library, for example. This is impractical in many cases and may even be impossible to obtain in some species. Screening of natural populations takes advantage of the recombination events that have accumulated since the origin of the alleles of interest. The number of samples that needs to be screened for a given map resolution is thus greatly reduced.

In the present study, we employed AFLP markers on a large number of individual plants from natural populations. This strategy offers the advantage that even weak linkage disequilibria can be detected. On the other hand, in using large sample sizes, fewer marker loci can be screened for a given input of resources. An alternative approach would be to use bulks of plants for the screening of markers. This strategy has been shown to be very effective for linkage disequilibrium mapping in humans (Arnhem et al., 1985; Barcellos et al., 1997). In those studies, allele frequencies were quantified in the bulks. Ranking of the markers based on differences in intensity between the bulks allowed a set of candidates to be selected that were then further analysed in individual samples. AFLP markers may not be ideal for the detection of allele frequencies as a function of band intensity. However, use of a highly multiplex marker system such as AFLP in combination with bulked samples could be very effective for the detection of absolute linkage disequilibria. If successful, this strategy could possibly improve map resolution further.

Linkage disequilibrium mapping has mostly been applied to the mapping of rare human disease genes in young populations, such as the Finnish (Hästbacka et al., 1992; de la Chapelle, 1993; Lehesjoki et al., 1993; Pekkarinen et al., 1998). In such cases, all copies of the disease allele probably originate from a single chromosome in the founding population. It is much more difficult to apply the method when each phenotype class is likely to represent several different alleles. This is similar to our case, since there is no reason to assume that the B gene polymorphism is young and since the two phenotype classes probably represent several alleles each. However, it has been shown that drift-generated linkage disequilibrium can be used effectively to map such genes in small, stable populations (Laan & Paabo, 1998; Terwilliger et al., 1998). Using larger populations should have the effect of narrowing the region spanned by linkage disequilibrium. Since our populations are not small, it was rather unexpected to find significant linkage disequilibrium between the B gene and any of the markers, given the small number of markers screened. The reason for this can probably be attributed to the position of the B gene. It is well known that the distribution of recombination in the sugar beet genome is very uneven (Halléden et al., 1996; Nilsson et al., 1997, 1999). The B gene was mapped to a region of suppressed recombination (data not shown). For this reason, linkage disequilibrium around the B gene can be expected to span a relatively large proportion of the physical map in sea beet as well. This increases the probability of detecting significant linkage disequilibrium between a random marker and the B gene.

The research was supported by the Swedish Strategic Network for Plant Biotechnology.

References

Abe, J., Guan, G.-P. & Shimamoto, Y. (1993). Linkage maps for nine isozyme and four marker loci in sugar beet (Beta vulgaris L.). Euphytica 66, 117–126.

Abe, J., Guan, G.-P. & Shimamoto, Y. (1997a). A gene complex for annual growth habit in sugar beet (Beta vulgaris L.). Euphytica 94, 129–135.

Abe, J., Guan, G.-P. & Shimamoto, Y. (1997b). A marker assisted analysis of bolting tendency in sugar beet (Beta vulgaris L.). Euphytica 94, 137–144.

Arnhem, N., Strange, C. & Erlich, H. (1985). Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. Proceedings of the National Academy of Sciences of the USA 82, 6970–6974.

Barcellos, L. F., Klitz, W., Field, L. L., Tobias, R., Bowcock, A. M., Wilson, R., Nelson, M. P., Nagatomi, J. & Thomson, G. (1997). Association mapping of disease loci, by use of a pooled DNA genomic screen. American Journal of Human Genetics 61, 734–747.

Boudry, P., Wieber, R., Saumitou-Laprade, P., Pillen, K., Van Dijk, H. & Jung, C. (1994). Identification of RFLP markers closely linked to the bolting gene B and their significance for the study of the annual habit in beets (Beta vulgaris L.). Theoretical and Applied Genetics 88, 852–858.
de la Chapelle, A. (1993). Disease gene mapping in isolated human populations: the example of Finland. *Journal of Medical Genetics* 30, 857–865.

Chapman, N. H. & Wijsman, E. M. (1998). Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. *American Journal of Human Genetics* 63, 1872–1885.

Cnops, G., den Boer, B., Gerarts, A., Van Montagu, M., Van Lijsbettems, M. (1996). Chromosome landing at the Arabidopsis *TORNADO* locus using an AFLP-based strategy. *Molecular and General Genetics* 253, 32–41.

Hästbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. & Lander, E. (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics* 2, 204–211.

Halden, C., Hjerdn, A., Rading, I. M., Säll, T., Fridlundh, B., Johannsdottir, G., Tuveson, S., Åkesson, C. & Nilsson, N.-O. (1996). A high-density linkage map of sugarbeet. *Genome* 39, 634–645.

Hansen, M., Kraft, T., Christiansson, M. & Nilsson, N.-O. (1999). Evaluation of AFLP in Beta. *Theoretical and Applied Genetcis* 98, 845–852.

Houwen, R. H. J., Bahrarloo, S., Blankenship, K., Raeymaeker, P., Juyn, J., Sandkuilj, L. A. & Freimer, N. B. (1994). Genome screening by searching for shared segments: mapping a gene for benign recurrent intra-hepatic cholestasis. *Nature Genetics* 8, 380–386.

Jorde, L. B. (1995). Linkage disequilibrium as a gene mapping tool. *American Journal of Human Genetics* 56, 11–14.

Kruglyak, L. (1997). What is significant in whole-genome linkage disequilibrium studies? *American Journal of Human Genetics* 61, 810–812.

Laan, M. & Pääbo, S. (1998). Mapping genes by drift-generated linkage disequilibrium. *American Journal of Human Genetics* 63, 654–656.

Lander, E. S. & Schork, N. J. (1994). Genetic dissection of complex traits. *Science* 265, 2037–2048.

Lehesjoki, A.-E., Koskinemi, M., Norio, R., Tirrito, S., Sistonen, P., Lander, E. & de la Chapelle, A. (1993). Localization of the *EPM1* gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Human Molecular Genetics* 2, 1229–1234.

Lewontin, R. C. (1995). The detection of linkage disequilibrium in molecular sequence data. *Genetics* 140, 377–388.

Long, A. D. & Langley, C. H. (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research* 9, 720–731.

Long, A. D., Lyman, R. F., Langley, C. H. & Mackay, T. F. C. (1998). Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* 149, 999–1017.

Michelmore, R. W., Paran, I. & Kesseli, R. V. (1991). Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the USA* 88, 9828–9832.

Nilsson, N.-O., Halldén, C., Hansen, M., Hjerdin, A. & Säll, T. (1997). Comparing the distribution of RAPD and RFLP markers in high density linkage map of sugar beet. *Genome* 40, 644–651.

Nilsson, N.-O., Hansen, M., Panagopoulos, A. H., Tuveson, S., Ehilde, M., Christiansson, M., Rading, I. M., Rissler, M. & Kraft, T. (1999). QTL analysis of *Cercospora* leaf spot resistance in sugar beet. *Plant Breeding* 118, 327–334.

Pekkarinen, P., Kestila, M., Paloneva, J., Terwilliger, J., Varilo, T., Järvi, O., Hakola, P. & Peltonen, L. (1998). Fine-scale mapping of a novel dementia gene, PLOSL, by linkage disequilibrium. *Genomics* 54, 307–315.

Peltonen, L. & Uusitalo, A. (1997). Rare disease genes: lessons and challenges. *Genome Research* 7, 765–767.

Schwarz, G., Michalek, W., Mohler, V., Wenzel, G. & Jahoor, A. (1999). Chromosome banding at the Mla locus in barley (*Hordeum vulgare*) by means of high-resolution mapping with AFLP markers. *Theoretical Applied Genetics* 98, 521–530.

Sheffield, V. C., Stone, E. M. & Carmi, R. (1998). Use of isolated inbred human populations for identification of disease genes. *Trends in Genetics* 14, 391–395.

Stam, P. & Van Ooijen, J. W. (1995). JoinMap version 2.0: software for the calculation of genetic linkage maps. *CPRO-DLO, Wageningen.*

Tanksley, S. D., Ganal, M. W. & Martin, G. B. (1995). Chromosome banding: a paradigm for map-based gene cloning in plants with large genomes. *Trends in Genetics* 11, 63–68.

Tevrilliger, J. D., Zöllner, S., Laan, M. & Pääbo, S. (1998). Mapping genes through the use of linkage disequilibrium generated by genetic drift: ‘drift mapping’ in small populations with no demographic expansion. *Human Heredity* 48, 138–154.

Van Dijk, H. & Boudry, P. (2001). Genetic variation for life histories in *Beta maritima*. *International Board of Plant Genetic Resources* 7, 44–55.

Van Dijk, H., Boudry, P., McCombie, H. & Vernet, P. (1997). Flowering time in wild beet (*Beta vulgaris ssp. maritima*) along a latitudinal cline. *Acta Oecologica* 18, 47–60.