Pedigree Genotyping: A New Pedigree-based Approach of QTL Identification and Allele Mining by Exploiting Breeding Material

W.E. van de Weg, R.E. Voorrips, R. Finkers, L.P. Kodde and E.J. Meulenbroek
Plant Research International, P.O. Box 16 6700 AA Wageningen
The Netherlands

J. Jansen and M.C.A.M. Bink
Biometris, P.O. Box 100, 6700 AC Wageningen
The Netherlands

Keywords: Identity by Descent, mapping, molecular marker, Linkage Disequilibrium, apple, strawberry, Marker-Assisted Breeding

Abstract

To date, molecular markers have been made available for many economically important traits. Unfortunately, lack of knowledge of their allelic variation hampers their full exploitation in commercial breeding programs. These markers have usually been identified in one single cross. Consequently, only one or two favourable alleles of the related QTL are identified and may be exploited for marker-assisted breeding (MAB), while a breeding program may include several alleles. Selection for only these alleles means that many favourable genotypes are ignored, which decreases efficiency and leads to genetic erosion.

A new approach, called pedigree genotyping, allows the identification and exploitation of the majority of alleles present in an ongoing breeding program. This is achieved by including breeding material itself in QTL detection, so covering multiple generations and linking many crosses through their common ancestors in the pedigree. The principle of Identity by Descent (IBD) is utilised to express the identity of an allele of a modern selection in terms of alleles of founding cultivars. These founder alleles are used as factors in statistical analysis. Co-dominant markers, like SSR (= microsatellite) markers, are essential in this approach since they are able to connect cultivars, breeding selections and progenies at the molecular marker level by monitoring specific chromosomal segments along family trees.

Additional advantages of the use of breeding genetic material are (1) a major reduction in experimental costs since plant material is already available and phenotyped by default (2) continuity over generations within breeding programs with regard to marker research (3) the testing of QTL-alleles against a wide range of genetic backgrounds, making results generally applicable (4) possibility to explore intra- as well as inter-QTL interactions. Fruit firmness in apple is used as an example to illustrate the principles of this powerful approach to detect QTLs and estimate their allelic variation. Prospects for strawberry are also indicated.

INTRODUCTION

To date, molecular markers have been identified for many loci governing important horticultural traits. These markers have usually been identified in one single cross. As a consequence, only one or two favourable alleles of a locus are identified, whereas a breeding program usually includes many favourable alleles. If a breeder focuses selection on these alleles, many favourable genotypes would be unnecessarily discarded. This reduces the efficiency of the breeding program. Moreover, the genetic diversity of the material is unnecessarily narrowed. A new approach called 'Pedigree Genotyping' makes it possible to find markers for all favourable alleles present in a breeding program. The costs of this approach are low compared to traditional marker research because it utilises data from the ongoing breeding program.

Principle of Pedigree Genotyping

The principle of pedigree genotyping is illustrated here by way of fruit firmness in apple. Apple is a diploid, outcrossing, vegetatively propagated species. Chromosome 10 is interesting in relation to *fruit firmness* (King et al., 2000; Maliepaard et al., 2001). Figure 1A shows a linkage map of this chromosome containing five SSR markers. SSR markers show generally co-dominant segregation, often having different marker alleles on their homologous chromosomes (Fig. 1B). This makes these markers very suitable for following the inheritance of their alleles through a breeding program. Figure 2 shows the pedigree of selection 81015-045, which is based on four different founder cultivars: Golden Delicious, Jonathan, Cox and Ingrid Marie. This pedigree is genotyped with the five SSR markers. We can now follow the transmission of these markers from one generation to the next, putting Pedigree Genotyping to work.

An example: 81015-045 has two alleles for SSR-5: '232' and '0'. Using the pedigree we can show that these two alleles are derived from ancestors 'Golden Delicious' and 'Ingrid Marie', respectively. This is called an "Identity by Descent" (IBD) analysis. The identity of an allele of a modern selection can now be expressed in terms of alleles of founding cultivars. These founder alleles are used as factors in a statistical analysis.

Marker-Allele Associations

One major locus for fruit firmness is located close to marker SSR-5. Some cultivars and related breeding selections, including those of Figure 2, were phenotyped and genotyped. Firmness was measured by penetrometer; values around 8 are desired, while 4 corresponds to apples that can be squeezed by hand. Results are presented in Figure 3. The '232' allele of 'Golden Delicious' (GD) and 'Wagnerapfel' (Wa) appear to be associated with good firmness. The average firmness of genotypes having this allele was around 8.3. This favourable linkage seems to be absent for the '232' marker of 'Jonathan' (Jo), which has an average value of 6 and is thus associated with soft fruit. The same SSR allele can thus be associated with different phenotypic effects depending on the origin of the marker. Consequently, it is important that the origin of the allele is taken into account in a statistical analysis in which traits are related to marker alleles.

Interactions Between Alleles

Sometimes a trait is not determined by the alleles separately, but by a combination of alleles at one locus or between alleles of different loci. Such specific combinations may be more favourable than expected from the average effects of the alleles. Specific combinations within a locus are exploited in F_1-hybrid cultivars and in vegetatively propagated crops, and are automatically identified by Pedigree Genotyping. For example, with regard to fruit firmness, genotypes with the allele combination '0, 230' have soft fruit (Fig. 4). However, genotypes that are homozygous for one of these SSR alleles (i.e. '0,0'and '230,230') may have good firmness. This indicates that these SSR alleles are not necessarily associated with inferior firmness alleles. Only in specific combinations of alleles does it result in the undesired phenotype. Consequently, crosses between '0,0' and '230,230' genotypes should be avoided. Such knowledge is of great advantage to breeders. Interactions can also occur between alleles of different loci, as in the case of complementary genes. With Pedigree Genotyping much more allele combinations can be evaluated than in a single test progeny, thus improving the prospects of identifying interactions both within and between loci. Indeed, recent simulation studies at Plant Research International confirmed the power of this approach in demonstrating the presence of complementary genes and in modelling their contribution to the phenotype.

Starting Points

Pedigree Genotyping can start from zero, when no marker-locus associations are known for the trait of interest. It can also start of from an already known locus, as in our example of fruit firmness. Starting from a known locus, new alleles for this locus can be identified. When no locus is known, Pedigree Genotyping can be used to identify loci for

a trait once sufficient numbers of genotypes have been evaluated. Compared to a single cross, a larger number of genotypes is required because of the larger number of alleles that have to be accounted for. However, once incorporated in ongoing breeding programmes the number of individuals will steadily grow over the years, and may soon exceed the size of any single cross.

Cost-effective

While breeding, Pedigree Genotyping searches for new marker-trait associations thereby making this approach cost-effective. It avoids the costs of growing and phenotyping specially designed 'scientific' progenies. Besides, SSR markers are cost effective for genotyping. The SSR markers of our example can be tested simultaneously (multiplexed). A chromosome can thus be genotyped by a single PCR reaction and a single lane of a gel. Once a genomic region of interest is identified, testing additional markers will improve resolution, leading to more tightly linked markers.

Requirements

Pedigree Genotyping requires genetically related breeding material, a set of multiallelic markers (like SSRs or sets of SNPs) that cover the genome segment of interest, software to calculate the genetic value of different marker alleles (IBD) as well as their effect on phenotype (QTL analysis), and in the longer term, a database to store all phenotypic and genotypic data. To date, the availability of co-dominant markers varies between species. For apple, completion of a genome spanning set of SSR-markers is under way (Liebhard et al., 2002; Gianfranceschi and Soglio, 2004).

Scientific Context

Interest in the exploitation of pedigree information in genetic analysis in plants is booming. For example, apple and pear pedigrees were explored to estimate heritabilities for various agronomic traits using, in the absence of molecular marker data, genome wide co-ancestries as factors in a statistical analysis (Durel et al., 1998, 2004).

The IBD approach has recently proven its value in human and animal genetics (Lynch and Walsh, 1998; Balding et al., 2001). However, it can even be more powerful in plant species like apple thanks to availability of plant material of some six generations and the possibility to examine plant material of all generations in the same year and at the same site. Vegetative propagation makes it possible to test genotypes simultaneously at various locations.

In plants, simulation studies (Jansen et al., 2003; Pérez-Enciso et al., 2003) as well as studies with real phenotypic data are used to show the efficiency of Pedigree Genotyping in particular cases. For example, Bink et al. (2002) employed the IBD approach in the diploid potato to identify QTLs, and linked molecular markers using six genetically related crosses. To date, no data are available on an integrated analysis of multiple crosses with a complex pedigree, cultivars and breeding selections.

Another approach to identify marker-trait associations in breeding and wild germplasm that has recently received much attention is 'Linkage Disequilibrium (LD)-mapping' (Gaut and Long, 2003; Gebhardt et al., 2004). This approach is, however, less efficient in cases where pedigree information is available, having a lower statistical power and requiring a very high density of molecular markers. The Pedigree Genotyping approach is more effective in cases where pedigree information is available, having a better statistical power thereby increasing the chances of success, and requiring a much lower density of molecular markers, thereby saving expenses on genotyping.

Statistics and Software

The current statistical tools and software applied in human genetics need adaptations to be applied to the plant system mainly because of the complex pedigree structure (many inbreeding loops), and to the expected high level of allelic variation. Besides, software packages employed in human and animal genetics are still limited in the sense

that interactions between alleles of the same locus as well as among those of different loci cannot yet be unravelled. Plant Research International and Biometris are developing software that meets the requirements. The software package FlexQTL™ (Bink, 2002; Bink et al., 2002) calculates IBD probabilities, and numbers and positions of QTLs as well as contributions of QTL alleles. The package PediMap supports the graphical presentation of the FlexQTL™ output.

Ongoing Applications

To date, this approach is followed in the HIDRAS EU-project (Gianfranceschi and Soglio, 2004), aiming to identify genes and linked molecular markers for fruit quality by an integrated analysis of over 300 cultivars and advanced breeding selections, and 1400 seedlings from 25 crosses.

Crops To Go For

Pedigree Genotyping offers great prospects for any crop in any breeding system. The greatest advantages are obtained if genotypes from past breeding programs and their phenotypes are available or easy to produce e.g. many vegetatively propagated crops or inbred lines and their F_1 hybrids; when it takes a long time to construct and evaluate mapping populations; when many loci are already known from special mapping populations; when the trait of interest is oligogenic; when phenotypic assessments are recorded routinely; when individual genotypes are relatively expensive.

Strawberry has several characteristics that give substantial advantages to Pedigree Genotyping. The species is vegetatively propagated, modern cultivars have well documented pedigrees, and many generations are still available. Also much phenotypic data from cultivar trials and breeding programs have been recorded.

Some disadvantages may arise from the octoploid nature of the cultivated strawberry. In theory, octoploidy would allow the simultaneous presence of eight different alleles for a single gene. However, segregation patterns of co-dominant molecular markers (Viruell et al., 2002), isozymes (Arulsekar et al., 1981) and closely linked (repulsion phase) AFLP makers (Van de Weg, unpublished) as well as cytological observations (Byrne and Jelenkovic, 1976) indicate that the cultivated strawberry is highly diploidised. This amphidiploid (allo-octoploid) nature of the cultivated strawberry allows the use of standard mapping and gene (QTL) identification procedures. Gene identification will therefore not only be feasible for major genes but also for quantitative traits.

The current lack of a genome spanning set of co-dominant markers for the cultivated strawberry will be rapidly reversed in the future, in the light of current efforts on SSR development (Viruell et al., 2002; Sargent et al., 2003; Van de Weg et al., unpublished).

CONCLUSION

Pedigree Genotyping is a powerful approach to marker-assisted breeding. Its advantages in summary are: (1) markers are found for most alleles that are relevant to the breeder since they are part of his own breeding material (2) alleles that show interactions are identified (3) Pedigree Genotyping can be fully performed on existing pedigrees thus reducing costs and time-to-market.

Pedigree Genotyping will change the way breeders work with their material. Within a Pedigree Genotyping context, the breeding material is not only a source of new varieties, but also a source of information. The value of this information will grow as more molecular data and phenotypic characterisations accumulate over generations. This requires a long-term view of its value. But after all, a long-term view is what breeders are famous for.

ACKNOWLEDGEMENTS

This paper has been written with financial support from the Commission of the European Communities, specific research programme “Quality of Life and Management

of Living Resources", QLK5-CT-2002-01492 "High-quality Disease Resistant Apples for a Sustainable Agriculture". It does not necessarily reflect its views and in no way anticipates the Commission's future policy in this area."

Literature Cited

Arulsekar, S., Bringhurst, R.S. and Voth, V. 1981. Inheritance of PGI and LAP isozymes in octoploid cultivated strawberry. J Am Soc Hortic Sci 106:679-683.

Balding, D.J., Bishop, M. and Cannings, C. 2001. Handbook of Statistical Genetics. Wiley, New York.

Bink, M.C.A.M. 2002. On flexible finite polygenic models for multiple-trait evaluation. Genet Res 80:245–256.

Bink, M.C.A.M., Uimari, P., Sillanpää, M.J., Janss, L.L.G. and Jansen, R.C. 2002. Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor. Appl. Genet. 104:751-762.

Byrne, D. and Jelenkovic, G. 1976. Cytological diploidization in the cultivated octoploid strawberry *Fragaria × ananassa*. Can J Genet Cytol 18:653-659.

Durel, C.E., Laurens, F., Fouillet, A. and Lespinasse, Y. 1998. Utilization of pedigree information to estimate genetic parameters from large unbalanced data sets in apple. Theor. Appl. Genet. 96:1077-1085.

Durel, C.E., Guérif, P., Belouin, A. and LeLezec, M. 2004. Estimation of fire blight resistance heritability in the French pear breeding program using a pedigree-based approach. Acta Hort. 663:251-256.

Gaut, B.S. and Long, A.D. 2003. The lowdown on Linkage Disequilibrium. The Plant Cell 15:1502-1506.

Gebhardt, C., Ballvora, A., Walkemeier, B., Oberhagemann, P. and Schuler, K. 2004. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol Breed 13:93-102.

Gianfranceschi, L. and Soglio, V. 2004. The European project HIDRAS: Innovative and Multidisciplinary approach to breed high quality disease resistant apples. Acta Hort. 663:327-330.

Jansen, R.C., Jannick, J.L. and Beavis, W.D. 2003. Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. Crop Sci 2003 43:829-834.

King, G.J., Maliepaard, C., Lynn, J.R., Alston, F.H., Durel, C.E., Evans, K.M., Griffon, B., Laurens, F., Manganaris, A.G., Schrevens, E., Tartarini, S. and Verhaegh, J. 2000. Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.). Theor Appl Genet 100: 1074-1084.

Liebhard, R., Gianfranceschi, L., Koller, B., Ryder, C.D., Tarchini, R., Van de Weg, W.E. and Gessler, C. 2002. Development and characterisation of 140 new microsatellites in apple (*Malus × domestica Borkh*.). Molecular Breeding 10:217-241.

Lynch, M. and Walsh, B. 1998. Genetics and analysis of quantitative traits. Sinauer Associates Inc, Sunderland.

Maliepaard, C., Silanpää, M.J., Van Ooijen, J.W., Jansen, R.C. and Arjas, E. 2001. Bayesian versus frequentitst analysis of multiple quantitative trait loci with an application to an outbred apple cross. Theor Appl Genet 103:1243-1253.

Perez-Enciso, M., Toro, M.A., Tenenhaus, M. and Gianola, D. 2003. Combining gene expression and molecular marker information for mapping complex trait genes: a simulation study. Genetics 164:1597-1606.

Sargent, D.J., Hadonou, A.M. and Simpson, D.W. 2003. Development and characterization of polymorphic microsatellite markers from *Fragaria viridis*, a wild diploid strawberry. Molecular-Ecology-Notes. 3:550-552.

Viruel1, M.A., Sánchez, D. and Arús, P. 2002. An SSR and RFLP linkage map for the octoploid strawberry (*Fragaria × ananassa*). Int Pl Anim Genomes Conf. http://www.intl-pag.org/pag/10/abstracts/PAGX_P660.html

Figures

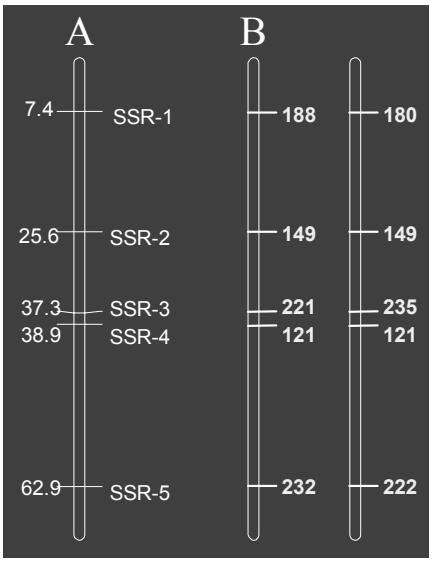

Fig. 1 **A**. Molecular marker map of chromosome 10 of apple.
B: allelic composition of the two individual homologous chromosomes of 'Golden Delicious' for the five SSR markers of Fig. 1A.

Alleles of Founders	# genotypes	Average Firmness
0 - IM	12	7.2
0 - JG	1	5.1
222 - GD	10	7.5
230 - Co	5	7.7
230 - Jo	3	6.8
230 - JG	1	5.1
232 - GD	8	8.3
232 - Wa	5	8,4
232 - Jo	3	6.6
234 - Pr	5	7.9

Fig. 3. Founder alleles of SSR-5 (distinguished by size and colour), and the average firmness of genotypes in which they occur.

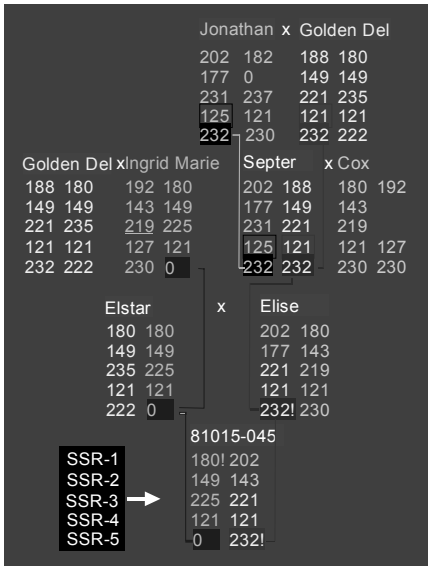

Fig. 2. Pedigree of the breeding selection '81015-045' and the allelic composition of each genotype for the five SSR markers of Fig. 1.

Data are used to assess allele flows over generations. For example, 'Elise' has two alleles for SSR-5, '230' and '232' that unambiguously descend from 'Cox' and 'Septer' respectively. 'Septer' is homozygous for '232', one allele coming from 'Jonathan', and one from 'Golden Delicious'. The linked marker SSR-4-121 of 'Elise' is present in 'Golden' only, indicating that Elise's 232 allele originated from this grandparent. However, a small chance remains that Jonathan was the actual source, since a recombination event may have occurred in 'Septer', linking SSR4-121 of 'Golden' to SSR-5- 232 of 'Jonathan'. The chance of recombination depends on the distance between the SSR markers. IBD values are therefore *probabilities*, which are enhanced by denser linkage maps.
Another example: 'Ingrid Marie' shows a single marker for SSR5-230 that cannot be homozygous, due to its lacking of '230' in 'Ingrid's' offspring, 'Elstar'. 'Ingrid Marie' must thus have a null allele for SSR5. Next, this null allele is passed to selection '81015-045'

Genotype	Allele		Firmness
James Grieve	0	230	5.1
73001-041	0	230	5.8
86015-111	0	0	7.4
Cox	230	230	8.3

Fig. 4. Firmness and allelic constitution of SSR-5 for four apple genotypes.