Molecular genetic studies in flax (*Linum usitatissimum* L.)

Moleculair genetische studies in vlas (*Linum usitatissimum* L.)

Promotor:

Prof. dr. ir. P. Stam
Hoogleraar in de Plantenveredeling

Copromotor:

Dr. ir. H.J. van Eck
Universitair docent, Laboratorium voor plantenveredeling

Promotiecommissie:

Prof. dr. A.G.M. Gerats (Radboud Universiteit Nijmegen)
Prof. dr. R.F. Hoekstra (Wageningen Universiteit)
Dr. R.G. van den Berg (Wageningen Universiteit)
Dr. ir. J.W. van Ooijen (Kyazma B.V., Wageningen)

Dit onderzoek is uitgevoerd binnen de onderzoeksscholen ‘Experimental Plant Sciences’ en ‘Production Ecology and Resource Conservation’.

Molecular genetic studies in flax (*Linum usitatissimum* L.)

Jaap Vromans

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. Dr. M.J. Kropff,
in het openbaar te verdedigen
op woensdag 15 maart 2006
des namiddags te vier uur in de Aula

CIP-DTA Koninklijke Bibliotheek, Den Haag

Molecular genetic studies in flax (*Linum usitatissimum* L.)
Jaap Vromans

PhD thesis, Wageningen University, The Netherlands
With references – with summaries in Dutch and English

ISBN 90-8504-374-3

Contents

Chapter 1

General introduction

History of flax breeding

Centre of origin

The origin of flax (*Linum usitatissimum* L.) is uncertain. Some authors consider *L. bienne* as the progenitor of small seeded flax, originating from Kurdistan and Iran, whereas others consider *L. angustifolium* containing high oil content and seed weight, as progenitor, originating from the Mediterranean region (Murre 1955; Zeven and de Wet 1975). Other authors suggest that *L. bienne* and *L. angustifolium* are the same species, and are widely distributed over Western Europe, the Mediterranean basin, North Africa, the Near East, Iran and Caucasus (Tutin *et al.* 1968; Zohary and Hopf 1993). Recently, a study with molecular markers suggested that the three species originate from one common ancestor, *L. angustifolium* being most ancient (Muravenko *et al.* 2003). While *L. usitatissimum* is an annual crop species, the wild forms can also be biannual or perennial. All species are predominantly self-pollinated (Zohary and Hopf 1993). Cross pollination may occur via honey bees (Williams 1988) or by artificial means.

Domestication

Flax was already grown 6000 – 8000 years ago in Egypt and Sumaria, and belongs (together with barley and wheat) to the oldest of cultivated plants. The distribution of flax from the Near East into Europe is well documented (Zohary and Hopf 1993). It is considered that flax cultivation in Western Europe (the Netherlands, Northern France, Belgium and Switzerland) started about 5000-3000 BC when semi-nomads, originating from the Middle East settled in Flanders and introduced flax cultivation (Dewilde 1983). Since the domestication of flax, there has been a preference for growing flax either for its fiber or oil. In the Western region of Eurasia, flax is mainly grown for its fiber, whereas in the Eastern region of Eurasia flax is grown for its oil (Gill 1987). Fiber flax has a long unbranched growth habit, whereas linseed (oil flax) is much shorter and highly branched. Throughout this thesis the distinction between fiber flax and linseed is made.

Modern plant breeding

Similar to most agricultural crops, commercial breeding of flax started at the end of the 19th century. However, already in 1816 Gelf Jensma developed the landrace Friesche Witbloei. First, he selected white flowering plants from blue flowering flax from Russia. Then after several rounds of selection in the white flowering plants, the landrace Friesche Witbloei originated with long stems (De Jonge 1942). Traditional flax breeding in the Netherlands was based on mass selection. Around 1900, pedigree selection of flax was introduced by prof. Broekema which later resulted in the high performing high yield or well performing cultivars Concurrent (Dorst), Wiera (Wiersema) en Resistenta (Hylkema) (De

Haan 1952; Murre 1955). Nowadays, several breeding methods are available, but the pedigree method is the most common one used in flax breeding (Salas and Friedt 1995).

Taxonomic status

The Linaceae family comprises of 22 genera of which genus *Linum* is the most well-known. The more than 200 species present in the genus *Linum* are divided in five subsections (Tutin *et al.* 1968), of which subsection *Linum* contains the cultivated species *L. usitatissimum* L. and the ornamentals *L. grandiflorum* and *L. perenne*. However, the latter two species are of little economic importance. The number of chromosomes of the *Linum* species show a wide range varying from 2n = 16 to 2n = 72 (Fedorov 1974). *L. usitatissimum* and its wild relatives contain 2n = 30 chromosomes (Muravenko *et al.* 2003). The genome size (1C) of cultivated flax is 686 Mbp (Bennett and Leitch 2004).

Utilization of flax as a crop species

The dual purpose of flax was already known in ancient times. In ancient Egypt, linen (derived from the fiber) was used for wrapping the royal mummies and additionally linseed oil was used to embalm the bodies of deceased Pharaohs (Dewilde 1983). For a long time flax has been cultivated as a dual-purpose crop, but nowadays fiber flax and linseed represent different gene pools. Fiber flax has been cultivated in the Netherlands and most likely in Belgium and Northern France since ancient times. The quality and fineness of the linen has been proven ever since (Bostock and Riley 1856; Stokkers *et al.* 2004).

The application of flax is not restricted to the production of linen yarn. In fact almost the whole plant is used, justifying the name given by Linnaeus, *L. usitatissimum*, which means useful flax. The short fibers are used in paper, isolation material, matrix composites and linen painting textile. The wooden shives which are released during the scutching of flax can serve as an energy source, litter in cattle farming or as source material for pressurized wooden bricks (Stokkers *et al.* 2004). The seeds of fiber flax are mainly used as sowing seeds for the next year.

Canada, China, the Russian federation and the United States in America are responsible for more than 65% of the world wide production of linseed. Besides the extraction of linseed oil, the seeds are also used in some food products, e.g. as an ingredient in bread. Linseed oil is high in linolenic fatty acid content (45–60%), making it a very effective drying agent. Although the oil is edible, it is used primarily for industrial purposes, such as the production of paints and oil-based coverings and the manufacture of linoleum flooring (Rowland 1998).

Linseed oil also offers important nutritional benefits because of the high levels of omega-3 fatty acids. Animal experiments and clinical intervention studies indicate that omega-3 fatty acids have anti-inflammatory properties, and therefore, might be useful in

the management of inflammatory and autoimmune diseases, including coronary heart disease, major depression, aging, rheumatoid arthritis, Chrohn's disease and cancer (Simopoulos 2002). Recent breeding achievements are in the development of a new flax type called 'Solin'. This name is used for flax cultivars with low (<5%) linolenic acid content in the oil (Dribnenki and Green 1995). Solin, which is agronomically not different from regular linseed, is being developed for the edible oil market.

Agronomical aspects (Cultivation and processing of flax)

Fiber flax and linseed perform best in different regions. Fiber flax is mainly grown in climates with a relative low temperature and high air humidity, which is characteristic for northern temperate regions. The subtropical regions and highlands are ideal locations for linseed cultivation and therefore linseed should be more tolerant to prolonged periods of drought (Bunting 1951). Although the soil type is not the most important factor in flax cultivation, the sandy clay soils of the Netherlands, Belgium and Northern France are very suitable for fiber flax cultivation.

Flax requires a wide crop rotation of about seven years. Also the preceding crop is important for growing flax to prevent the occurrence of diseases and lodging. Flax does not perform well after potatoes and sugar beets as the soil may be too loose and *Rhizoctonia* disease could be a problem. On the contrary, cereals and maize are good preceding crops. Flax also does well after legume crops, but *Rhizoctonia* might be a problem as well (Rowland 1998; Stokkers *et al.* 2004).

As a rule of thumb, flax is sown at day 100 of the year and harvested at day 200, which is a growing period of 100 days. However, this depends somewhat on the cultivar and environmental conditions. The high sowing density of fiber flax of 110-130 kg/ha (Bonthuis *et al.* 2005) results in plant elongation due to the competition for light. This is important to obtain long high quality unbranched fibers. Linseed is sown with a lower density, 25-55 kg/ha (Rowland 1998), to stimulate branching in order to obtain higher numbers of flowers and an increased seed yield. Flax starts to flower approximately 11 to 14 weeks after sowing. The flowers are open for only a couple of hours in the morning, after which the sepals fall off and petals close. Ten to 14 days after flowering the fruit reaches its final size, after which the weight remains stable until it decreases as a consequence of the ripening process. At the end of the development the flax plant hardens, turns yellow (senescence) and loses its leafs.

At a certain point the plants are ready for the retting process, although the seeds might not be fully ripened. In the Netherlands the seeds are rippled while pulling, which implies that synchronization of fiber and grain maturation is important to reduce harvesting risks (Keijzer 1988). The retting process is the most crucial phase of flax cultivation, because it determines the yield and quality of the fiber. In the Netherlands dew retting is

used, which is a natural rotting process where the fibers are released from the wood by bacterial enzymes. As dew retting depends on the weather conditions, this process is difficult to control (Stokkers *et al.* 2004). After the retting process, the straw is harvested and the fibers are released from the wood by scutching and smoothened by hackling. The fiber lint will be send to spinning companies that produce the final linen yarn. Unlike fiber flax, linseed is harvested by straight combining or cutting with a swather and treshing later with a combine, depending on the maturity and dryness of the seeds. Traditionally, the straw of linseed is processed by companies that extract the flax fiber for the production of specialty papers (i.e. paper for cigarettes, currency, bibles and artwork etc.) (Rowland 1998).

Importance of flax cultivation in the Netherlands

Around 1950 approximately 25,000-30,000 ha flax was grown in the Netherlands (Murre 1955). The situation in the Netherlands changed under influence of low price Russian flax which was introduced on the Western Europe linen market in the late 1950's. Furthermore, the introduction of synthetic fibers, in combination with strongly reduced cotton prices, led to a collapse of the linen market. Consequently, the flax acreage in the Netherlands decreased dramatically to about 3,000 ha in 1980 (Kozolowski 1997). In the last 20 years the flax production in the Netherlands has been relatively stable around 4,000 ha (Stokkers *et al.* 2004).

Recently, China has developed into one of the most important producers of high-quality linen clothing in the world and therefore need to import large quantities of flax fibers. Hence, an increased production area of flax was expected in Western Europe. The growth of the Dutch flax production area from 1999 to 2003 is however small (29%) compared to France (56%) and Belgium (58%). Traditionally, the Netherlands could rely on a strong position in breeding and seed production, but France has improved the level of self-sufficiency in sowing seeds and weakening the position of the Dutch seed-growers. A second threat to flax cultivation in the Netherlands is the decrease of direct financial support for flax cultivation from the Dutch government after 2005 due to the reform of the Common Agricultural Policy, whereas both French and Belgian flax growers obtain governmental financial support by means of subsidies. Also the significantly higher labor and land costs in the Netherlands as compared to Belgium and France has weakened the position of Dutch flax cultivation (Stokkers *et al.* 2004).

Flax belongs to the Dutch national heritage and contributes to rural diversity. Besides, flax has a positive image because of its environmental friendly cultivation and the utility as green raw material. These characteristics provide possibilities to bring flax cultivation under attention to policy makers and politicians. In the future this may result in

more governmental support. In this respect the flax industry in France and Belgium has already achieved successes (Stokkers *et al.* 2004).

Current breeding goals

Breeding for quantitative traits

Breeding for fiber yield can be divided into two components, straw yield and fiber content (Popescu *et al.* 1998). The low heritability found for fiber yield suggests a considerable environmental influence. Contrary, fiber content (ratio of fiber weight on the stem) is more heritable and easier to determine. However, in order to calculate fiber content, large numbers of stems are necessary and consequently large trial plots and big quantities of seeds (Fouilloux 1988). Furthermore, it is known that both additive and dominant effects of genes are involved in the heredity of fiber content and both effects are influenced by environmental conditions (Popescu *et al.* 1998). However, it is still uncertain how many genes are involved in the heredity of fiber content. Likewise, the inheritance of straw yield is poorly understood. Straw yield is supposed to be controlled by several genes as well. A modest gain is to be expected from breeding for straw yield as heritabilities were shown to be low (Mourad and Abo-Kaied 2003). Considering the genetic basis of both straw yield and fiber content, breeding for fiber yield should be mainly focused on the more heritable fiber content. Oil yield is the most important quantitative trait in linseed. Oil yield is dependent on the seed yield and linseed oil content. Low heritabilities were observed for seed yield in early generations. Contrary, selection on oil content, a character with comparatively high heritability, in an early stage, should be feasible and successful in linseed (Salas and Friedt 1995).

Resistance breeding

Three major diseases posed a threat to flax cultivation in the Netherlands over the past 100 years, namely rust, scorch and wilting. Most research focused on rust resistance, where the differential interaction between the pathogen (*Melamspora lini*) races and the resistant host genotypes, resulted in the widely known gene-for-gene theory (Flor 1956). Nowadays, the Dutch soils are no longer contaminated with rust and all cultivars in the Netherlands are resistant to flax rust (Bonthuis *et al.* 2005). Therefore, resistance breeding is now focused on scorch and wilting.

At this moment the soil-borne disease scorch is one of the major problems in Western European flax cultivation. In France 20 percent of the acreage is infected and depending on the cultivar and disease pressure, losses in harvest can vary from 10 up to 90 percent. To avoid these yield losses, either flax should be grown on soils free of scorch or resistant cultivars should be used (Cariou *et al.* 2003).

Breeding for scorch resistance is difficult, because little is known about the inheritance of scorch resistance and the causative pathogen is unclear. *Chalara Elegans, Pythium megalacanthum* and *Pythium buismaniae* have been suggested as the most likely pathogen of flax scorch (Cariou *et al.* 2003; Delon and Kiffer 1978; Wiersema 1955). Breeding for scorch resistance has a long history in the Netherlands! In 1893 Prof. L. Broekema selected the first resistant plants. Nevertheless, several of the current Dutch cultivars are still partly or fully susceptible to scorch (Bonthuis *et al.* 2005).

Another important disease in flax is wilting caused by *Fusarium oxysporum* f.sp. *lini*. Contrary to the scorch pathogen much more is known about the fungus *Fusarium*. Flax wilt occurs across all main flax and linseed growing countries and may cause severe losses. The main route of infection is through the roots. The symptoms may show up throughout the whole growing season. To select resistant cultivars and breeding lines field wilt nurseries are used. Recently, two in vitro screening methods were developed which proved to be useful in resistance screening, for evaluating race specificity of resistance and to study pathogenesis (Kroes *et al.* 1998). Little is known about the inheritance of resistance to *Fusarium*. The results concerning the inheritance and durability reported so far are contradictory, but the majority of the reports mention the quantitative nature of the inheritance (Goray *et al.* 1987; Kamthan *et al.* 1981; Knowles and Houston 1955; Knowles *et al.* 1956; Kommendahl *et al.* 1970; Pavelek 1983). Recently Spielmeyer *et al.* (1999) identified two QTLs involved in the inheritance of Fusarium resistance explaining 38 percent and 26 percent of the phenotypic variance. Unfortunately, the AFLP marker loci and the linkage groups have been described in a way that does not allow identification of the same linkage groups and the same loci in other mapping populations.

Breeding for quality traits

Fiber quality is the most important trait related to quality. The flax fiber is an irregular strand of varying number of individual cells, the so-called elementary fiber cells. These cells also vary in diameter and in cell wall thickness. The cell wall consists mainly of cellulose. The quality of the fiber after retting and scutching can be defined by different physical and chemical parameters, i.e. strength, fineness and degree of polymerization of cellulose (D.P.-value) (Keijzer and Metz 1992). Recently near infrared (NIR) spectroscopy has been used for the prediction of fiber fineness and strength in flax NIR spectroscopy is a useful tool (Faughey and Sharma 2000), and is currently used in flax breeding. To identify genes involved in fiber quality a large number of Expressed Sequence Tags (ESTs) have been generated from genes expressed during flax fiber cell formation (Day *et al.* 2005). This resulted in the identification of a number of highly expressed genes involved in the synthesis of cell wall polymers such as celluloses, hemicelluloses, pectins and lignins. Apparently, the inheritance of fiber quality could be a complex trait, but it

should be feasible to test several candidate genes for the presence or absence of polymorphisms in expression level or DNA sequence.

The quality of the linseed oil is determined by the fatty acid composition (especially the linoleic and linolenic acid content). Wild type linseed oil contains a high level (45-65%) of α-linolenic acid. This polyunsaturated fatty acid is highly susceptible to oxidation and polymerization and therefore well suited for industrial purposes. Contrary, for edible purposes, linseed oil is not sufficiently stable, because autoxidation during storage will result in off flavors (Green 1986a). To improve linseed oil for edible purposes two different mutant genotypes flax were developed that contain low levels of linolenic acid in its seed oil (Green 1986b; Rowland 1991). These mutants resulted in the first low linolenic acid cultivar LINOLA^{TM}-947 of which the oil has great potential in food industry (Dribnenki and Green 1995).

Marker Assisted Selection (MAS)

Molecular markers are widely used in plant and animal research to understand the genetic basis of monogenic and polygenic complex traits (quantitative trait loci, QTL). Molecular markers are applied in breeding programs, where marker assisted selection can replace or complement the conventional phenotypic selection. Marker assisted selection is especially powerful for complex traits with low heritabilities (Knapp 1998; Lande and Thompson 1990). MAS is also a valuable tool in breeding for disease resistance, where reliable inoculation and scoring methods are often lacking (Young 2000). Another motivation for using MAS is to achieve a reduction in costs by replacing more expensive methods of phenotypic selection (Knapp 1998; Peleman and Van der Voort 2003; Stam 1994). The greatest benefits of MAS will be achieved in perennial crops because many traits are expressed only after several years of costly field maintenance (Liebhard *et al.* 2003). Furthermore, MAS showed to be very successful in the construction of Backcross Introgression Libraries (BILs) which serve as QTL detection and breeding material (Eshed *et al.* 1992; Howell *et al.* 1996; Jeuken and Lindhout 2004).

To efficiently apply MAS, a few practical considerations concerning the limitations of its utility should be given proper attention. First, the marker loci and the trait loci should be tightly linked, which requires mapping studies with high density linkage maps (Lande and Thompson 1990). With the introduction of multiplex PCR marker system such as AFLP (Vos *et al.* 1995) the construction of dense linkage maps have proven to be successful in almost all important agricultural crops such as potato, barley, rice, maize and soybeans etc. (Castiglioni *et al.* 1999; Eck *et al.* 1995; Lin *et al.* 1996; Maheswaran *et al.* 1997; Qi *et al.* 1998). Secondly, the association between marker alleles and trait genes should not only occur in a single experimental full sib population, but in a wider range of cultivated germplasm. This extrapolation will rely on linkage disequilibria, which are continuously

eroded by recombination. Therefore, MAS is optimally suited for the development of new varieties, especially in plants, where breeders routinely cross different pre-existing elite varieties to start a new cycle of selection (Lande and Thompson 1990). To detect QTLs for traits with a low heritability the offspring size of a mapping population should be very large (Lande *et al.* 1990). With QTL mapping simulation studies, Beavis *et al.* (1994) showed that only a small percentage of all QTLs could be discovered when small population sizes of 100-200 progeny individuals were used. He further indicated that approximately 500 individuals are required, to obtain more reliable and comprehensive information on the presence of quantitative trait loci (Beavis *et al.* 1994).

Scope of the thesis

The research objectives

The main objective is to investigate the hereditary basis of important traits in flax, related to disease resistance and fiber quality. This is achieved by DNA fingerprinting in mapping populations that segregate for these traits. Eventually this could result in applicable results. In flax breeding, MAS can speed up the breeding process and for some traits MAS could be more cost effective as compared to field trials. The second objective is the analysis of the amount of genetic diversity in the flax gene pool. Experts dealing with testing plant cultivars for Distinctness, Uniformity and Stability (DUS) criteria or Plant Breeders' right have noticed that the morphological variation between new cultivars has decreased (Everaert *et al.* 2001). Dutch flax breeders also arrived to this conclusion as they experience increasing difficulties to develop new cultivars with improved fiber content, fiber quality and other important traits over the existing cultivars (personal communication, flax breeders). The results on genetic diversity may provide knowledge and tools to increase the diversity for future flax breeding. Finally, some of the research questions focus on the utility of the AFLP technology in flax.

Outline of the thesis

In **chapter 2** we investigate the genetic variation in the gene pool of fiber flax and linseed. The low amount of genetic variation observed in this chapter prompts us to investigate the amount of genetic diversity among wild species related to *L. usitatissimum*. The observed grouping of wild species and the possible use of wild species in flax introgression breeding or BIL development are discussed in **chapter 3**. As flax is a very homogeneous crop the number of polymorphic markers in AFLP fingerprints appeared to be very low. To improve the efficiency of AFLP fingerprinting of mapping populations, **Chapter 4** describes the results of screening a large number of AFLP primer combinations. We discuss the efficiency of AFLP fingerprinting in flax and the benefit of a preliminary screen. Consecutively, the most polymorphic primer combinations were selected to construct an

integrated genetic linkage map. The approach to construct linkage maps using two mapping populations by combining the JoinMap and RECORD marker ordering algorithms, and the approach to obtain an integrated map of flax is discussed in **chapter 5**. Finally, in **chapter 6** we try to identify the genetic loci involved in several qualitative and quantitative traits of flax. This resulted in the localization of a large number of QTLs on the genetic linkage maps of flax. We describe the results of four traits in more detail, because of their relevance in fiber flax and linseed breeding.

References

Beavis, W.D., Smith, O.S., Grant, D. and Fincher, R. (1994) Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. *Crop Science* 34: 882-896

Bennett, M.D. and Leitch, I.J. (2004) Plant DNA C-values database (release 3.0) http://www.rbgkew.org.uk/cval/homepage.html

Bonthuis, H., Donner, D.A. and van Vliegen, A. (2005) 80e Rassenlijst voor Landbouwgewassen 2005. 214 pp

Bostock, J. and Riley, H.T. (1856) The history of Pliny. Vol. IV. London: 523 pp

Bunting, A.H. (1951) Linseed. *Paint Manufacturing* 21: 444-447

Cariou, E., Brochard, M. and Raynal, G. (2003) La brulure du lin, due a *Chalara elegans* et plusieurs especes de Pythium. *Phytoma* 566: 6-9

Castiglioni, P., Ajmone-Marsan, P., van Wijk, R. and Motto, M. (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. *Theoretical and Applied Genetics* 99: 425-431

Day, A., Addi, M., Kim, W., David, H., Bert, F., Mesnage, P., Rolando, C., Chabbert, B., Neutelings, G. and Hawkins, S. (2005) ESTs from the fibre-bearing stem tissues of flax (*Linum usitatissimum* L.): Expression analyses of sequences related to cell wall development. *Plant Biology* 7: 23-32

De Haan, H. (1952) Flax breeding and flax varieties in the Netherlands. *Euphytica* 1: 212-218

De Jonge, L.J.A. (1942) Het rassenvraagstuk bij vlas. H. Veerman & zoonen, Wageningen, The Netherlands: 175 pp

Delon, R. and Kiffer, E. (1978) *Chalara elegans* (*Thielaviopsis basicola*) and allied species, 1: Generalities and pathology. *Annales du Tabac* 2: 159-186

Dewilde, B. (1983) 20 eeuwen vlas in Vlaanderen: 439 pp

Dribnenki, J.C.P. and Green, A.G. (1995) LinolaTM '947' low linolenic acid flax. *Canadian Journal of Plant Science* 75: 201-202

Eshed, Y., Abu Abied, M., Saranga, Y. and Zamir D (1992) *Lycopersicon esculentum* lines containing small overlapping introgressions from *L. pennellii*. *Theoretical and Applied Genetics* 83: 1027-1034

Everaert, I., de Riek, J., de Loose, M., van Waes J., and van Bockstaele, E. (2001) Most similar variety grouping for distinctness evaluation of flax and linseed (*Linum usitatissimum* L.) varieties by means of AFLP and morphological data. *Plant Varieties and Seeds* 14: 69-87

Faughey, G.J. and Sharma, H.S.S. (2000) A preliminary evaluation of near infrared spectroscopy for assessing physical and chemical characteristics of flax fibre. *Journal of Near Infrared Spectroscopy* 8: 61-69

Fedorov A.A. (1974) Chromosome numbers of flowering plants: 412-414

Flor, H.H. (1956) The complementary genic systems in flax and flax rust. *Advances in Genetics Incorporating Molecular Genetic Medicine* 8: 29-54

Fouilloux, G. (1988) Breeding flax methods. In Flax: Breeding and Utilisation: 14-25

Gill, K.S. (1987) Linseed. Indian Council of Agricultural Research, New Dehli: 186 pp

Goray, S.C., Khosla, H.K., Upadhyaya, Y.M., Naik, S.L. and Mandloi, S.C. (1987) Inheritance of wilt resistance in linseed. *Indian Journal of Agricultural Sciences* 57: 625-627

Green, A.G. (1986a) Genetic conversion of linseed oil from industrial to edible quality. *Journal of the American Oil Chemists Society* 63: 464-464

Green, A.G. (1986b) A mutant genotype of flax (*Linum usitatissimum* L.) containing very low levels of linolenic acid in its seed oil. *Canadian Journal of Plant Science* 66: 499-503

Howell, P.M., Marshall, D.F. and Lydiate, D.J. (1996) Towards developing intervarietal substitution lines in *Brassica napus* using marker-assisted selection. *Genome* 39: 348-358

Jeuken, M.J.W. and Lindhout, P. (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. *Theoretical and Applied Genetics* 2004; 109: 394-401

Kamthan, K.P., Misra, D.P. and Shukla, A.K. (1981) Independent genetic resistance to wilt and rust in linseed. *Indian Journal of Agricultural Sciences* 51: 556-558

Keijzer, P. (1988) Synchronization of fibe and grain maturation. In Flax: Breeding and Utilization: 26-36

Keijzer, P. and Metz, P.L.J. (1992) Breeding for fibre production in Western Europe. In: The Biology and Processing of Flax: 33-66

Knapp, S.J. (1998) Marker-assisted selection as a strategy for increasing the probability of selecting superior genotypes. *Crop Science* 38: 1164-1174

Knowles, P.F. and Houston, B.R. (1955) Inheritance of resistance to Fusarium wilt of flax in Dakota Selection 48-94. *Agronomy Journal* 47: 131-137

Knowles, P.F., Houston, B.R. and McOnie, J.B. (1956) Inheritance of resistance to Fusarium wilt of flax in Punjab 53. *Agronomy Journal* 48: 135-137

Kommendahl, T., Christensen, J.J. and Frederiksen, J.B. (1970) A half century of research in Minnesota on flax wilt caused by *Fusarium oxysporum*. Minnesota Agricultural Experiment Station. *Technical Bulletin* 273: 35 pp

Kozolowski, R. (1997) Euroflax Newsletter. *Information bulletin of the FAO european cooperative research network on flax and other bast plants* 1: 38 pp

Kroes, G., Sommers, E. and Lange, W. (1998) Two in vitro assays to evaluate resistance in *Linum usitatissimum* to Fusarium wilt disease. *European Journal of Plant Pathology* 104: 561-568

Lande, R. and Thompson, R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124: 743-756

Liebhard, R., Koller, B., Gianfranceschi, L. and Gessler, C. (2003) Creating a saturated reference map for the apple (*Malus* x *domestica* Borkh.) genome. *Theoretical and Applied Genetics* 106: 1497-1508

Lin, J.J., Kuo, J., Ma, J., Saunders, J.A., Beard, H.S., MacDonald, M.H., Kenworthy, W., Ude, G.N. and Matthews, B.F. (1996) Identification of molecular markers in soybean comparing RFLP, RAPD and AFLP DNA mapping techniques. *Plant Molecular Biology Reporter* 14: 156-169

Maheswaran, M., Subudhi, P.K., Nandi, S., Xu, J.C., Parco, A., Yang, D.C. and Huang, N. (1997) Polymorphism, distribution, and segregation of AFLP markers in a doubled haploid rice population. *Theoretical and Applied Genetics* 94: 39-45

Mourad, N.K.M. and Abo-Kaied, H.M.H. (2003) Genetic analysis of yield and yield components in flax (*Linum usitatissimum* L.). *Egyptian Journal of Agricultural Research* 81: 595-608

Muravenko, O.V., Lemesh, V.A., Samatadze, T.E., Amosova, A.V., Grushetskaya, Z.E., Popov, K.V., Semenova, O.Y., Khotyuleva, L.V. and Zelenin, A.V. (2003) Genome comparisons with chromosomal and molecular markers for three closely related flax species and their hybrids. *Russian Journal of Genetics* 39: 414-421

Murre, M. (1955) Vezelvlas. Uitgeverij Ceres, Meppel, The Netherlands: 112 pp

Pavelek, M. (1983) Inheritance of horizontal resistance of fibre flax to fusariosis. *Len a Konopi* 19: 47-64

Peleman, J.D. and van der Voort, J.R. (2003) Breeding by Design. *Trends in Plant Science* 8: 330-334

Popescu, F., Marinescu, I. and Vasile, I. (1998) Heredity and stability of flax fiber content. *Romanian Agricultural Research* 9: 15-24

Qi, X., Stam, P. and Lindhout, P. (1998) Use of locus-specific AFLP markers to construct a high-density molecular map in barley. *Theoretical and Applied Genetics* 96: 376-384

Rowland, G.G. (1998) Growing flax: Production, management and diagnostic guide. Flax Council of Canada and Saskatchewan Flax Development Commission

Rowland, G.G. (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). *Canadian Journal of Plant Science* 71: 393-396

Salas, G. and Friedt, W. (1995) Comparison of pedigree selection and single seed descent for oil yield in linseed (*Linum usitatissimum* L). *Euphytica* 83: 25-32

Simopoulos, A.P. (2002) Omega-3 fatty acids in inflammation and autoimmune diseases. *Journal of the American College of Nutrition* 21: 495-505

Spielmeyer, W., Green, A.G., Bittisnich, D., Mendhem, N. and Lagudah, E.S. (1998) Identification of quantitative trait loci contributing to Fusarium wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). *Theoretical and Applied Genetics* 97: 633-641

Stam, P. (1994) Marker-assisted breeding. In: Biometrics in plant breeding: applications of molecular markers: 32-44

Stokkers, R., Janssens, S.R.M., Vvan der Waal, B.H.C. (2004) Vezelvlassector in opmars. Lei, Wageningen University, Den Haag: 66 pp

Tutin, T.G., Heywood, V.H., Burges, N.A., Moore, D.M., Valentine, D.H., Walters, S.M. and Webb, D.A. (1968) Flora Europaea Rosaceae to Umbelliferae 2: 206-211

Van Eck, H.J., van der Voort, J.R., Draaistra, J., van Zandvoort, P., van Enckevort, E., Segers, B., Peleman, J., Jacobsen, E., Helder, J. and Bakker, J. (1995) The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. *Molecular Breeding* 1: 397-410

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414

Wiersema, H.T. (1955) Flax scorch. *Euphytica* 4: 197

Williams, I.H. (1988) The pollination of linseed and flax. *Bee World* 69: 145-152

Young, N.D. (2000) A cautiously optimistic vision for marker-assisted breeding. *Molecular Breeding* 5: 505-510

Zeven, A.C. and de Wet, J.M.J. (1975) Dictionary of cultivated plants and their regions of diversity. Pudoc, Centre fro Agricultural Publishing and Documentation, Wageningen, The Netherlands: 263 pp

Zohary, D. and Hopf, M. (2000) Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Oxford University Press, Oxford: 316 p

Chapter 2

Molecular analysis of flax (*Linum usitatissimum* L.) reveals a narrow genetic basis of fiber flax

Jaap Vromans, Eugenie van de Bilt, Bjorn Pieper, Piet Stam and Herman J. van Eck

Abstract

A set of 110 flax (*Linum usitatissimum*) accessions (cultivars and breeding lines) and one wild species (*L. bienne*) were evaluated with AFLP markers to study the diversity within and between the two morphological groups of flax, linseed and fiber flax, respectively. Neighbor Joining cluster analysis showed a clear separation of the *L. bienne* accession from the other accessions, supported by a high bootstrap value. Next, the fiber flax cultivars cluster separately from the linseed and the unknown purpose accessions, though not supported by a high bootstrap value. Contrary, principle component analysis (PCA) showed a clear differentiation between fiber flax and linseed cultivars, but *L. bienne* could not be distinguished from the linseed cultivars. Furthermore, PCA showed a clear differentiation between Dutch, French and fiber flax cultivars from other countries, and it also revealed a further differentiation of Dutch fiber flax cultivars from linseed cultivars. The observed difference between linseed and fiber flax cultivars is mainly caused by significant linseed or fiber flax specific markers. This implies the presence of population structure in the flax genepool. Finally, analysis of molecular variance (AMOVA) showed that fiber flax cultivars are less variable than linseed cultivars and the unknown purpose cultivars. For the future introduction of new favorable alleles into the fiber flax gene pool, the more variable linseed cultivars and the wild species *L. bienne* are potential donors.

Introduction

Flax (*Linum usitatissimum* L.) is a self-pollinating annual crop species, which is grown commercially for two products, fiber and seed. Due to selective breeding two distinguishable cultivar groups have developed and are grown as different crops, fiber flax and linseed. Fiber flax is bred for its long stem containing long fibers and is mainly grown in Russia, China, Egypt and near the Northwestern European coast, whereas linseed was deliberately bred for short and highly branched plants to increase the number of flowers (enhanced seed production). The main purpose of the fiber is the production high quality linen whereas linseed oil originating from the seeds has many industrial applications in e.g. paints, linoleum carpet and ink. Fiber flax breeding in the Netherlands started around 1900 when L. Broekema at Wageningen obtained cultivars through pedigree selection (De Haan 1952). In recent years it has become more difficult to breed new fiber flax cultivars with a better fiber quality, increased fiber yield and the required resistances, due to repeated use of modern cultivars as crossing parents. Also experts dealing with testing plant cultivars for DUS criteria or Plant Breeders' right have noticed that the morphological variation among new cultivars has decreased (Everaert *et al.* 2001).

Molecular markers have proved to be powerful tools in the assessment of genetic variation both within and between plant populations by analyzing large numbers of loci distributed throughout the genome (Powell *et al.* 1995). Compared with other molecular marker systems, AFLPs proved to be very efficient in germplasm analysis because of the high number of markers (20-150 loci per assay) that can be generated per analysis and no prior sequence knowledge is required (Powell *et al.* 1996; Vos *et al.* 1995). Genetic diversity studies with the use of AFLP markers have successfully been performed on several important crops like *Brassica rapa*, rice, potato, durum wheat and tomato (Kardolus *et al.* 1998; Park *et al.* 2004; Soleimani *et al.* 2002; Zhao *et al.* 2005; Zhu *et al.* 1998).

Flax germplasm has recently been investigated with different marker systems. The first molecular markers used to study the genetic diversity of flax were isozymes (Månsby *et al.* 2000). These authors clearly defined five groups, all with little variation within groups. They also observed an unexpected high genetic diversity within accessions. RAPD markers were used to assess the genetic variation within and among flax cultivars and landraces, to estimate the rate of change in genetic diversity in Canadian flax breeding programs over the last 50 years and to examine the geographic distribution of molecular diversity providing insight into flax domestication (Fu *et al.* 2002; Fu *et al.* 2003; Fu 2005). Fu (*l.c.*) showed that fiber flax cultivars are closely related (or similar to each other) and were classified as a homogeneous group. All the linseed cultivars were clustered in diverse groups together with the nine landrace accessions (Fu *et al.* 2002). They also observed that Canadian breeding programs resulted in a greater loss of diversity, as

compared to United States breeding programs. These conclusions were based on the higher proportion of fixed loci in the Canadian breeding programs (Fu *et al.* 2003; Fu *et al.* 2002). Analyzing a worldwide collection they concluded that the majority of the variation (84.2%) resided within accessions of each country and only 15.8% was present among accessions of different countries. Furthermore, accessions from East Asia and European regions were most diverse, but those from Africa and the Indian Subcontinent were more distinct (Fu 2005).

AFLPs can assist in the removal of redundant accessions from the flax collections in gene banks (Van Treuren *et al.* 2001). Van Treuren *et al.* (2001) also observed an unexpected high proportion of within accession variation, which is comparable with the results of Månsby *et al.* (2000). Also Everaert *et al.* (2001) used AFLPs and concluded that molecular markers are useful, as additional descriptors for the determination of genetic diversity. In addition he observed that a fairly large part of the genetic variation (13%) could be explained by the division between new and old cultivars and only 3% by the division between linseed and fiber flax. Finally, a clustering into eight groups based on ISSR showed to be significantly correlated with thousand seed mass.

In order to study the effect of plant breeding more closely a detailed examination of the variation between the two morphological groups and within fiber flax is desirable to study the effect of plant breeding more closely. In this report we used AFLP technology to analyze the relationships among 110 *L. usitatissimum* and one *L. bienne* accessions collected in different parts of the world. Special emphasis is put on the comparison of the two morphological groups within cultivated flax and the effect of breeding on the genetic diversity within fiber flax.

Material and Methods

Plant material

We analyzed 110 *L. usitatissimum* cultivars and breeding lines and one *L. bienne* accession with AFLP. The cultivars and breeding lines have been provided mainly by three breeding companies from the Netherlands: Dobbelaar Breeding, Van de Bilt Zaden en Vlas and Cebeco seeds. Four *L. usitatissimum* accessions and one *L. bienne* accession were obtained from the Dutch Centre of Genetic Resources, CGN (Wageningen, the Netherlands). The remaining cultivars have been supplied by USDA-ARS Plant Introduction Station Iowa State University (USA), CSIRO Plant Industry in Canberra (Australia) and Dr. Steve Knapp (Oregon State University, Canada). In Table 1 all accessions are listed with their provider, place of origin and pedigree information and type of cultivar. The collection includes fiber flax cultivars, linseed cultivars, some unknown purpose types and breeding material originating from different locations.

Table 1 List of accessions used in this study.

Cultivar name	Parents	Origin	Year of introduction	Code [1]/ Purpose	Source [2]
Abyssinian		Ethiopia		U1	VDB
AC Emerson	Noralta x Vimy	Canada	1994	L1	DOB
AC McDuff	(Redwood 65 x Linott) x Dufferin) x (McGregor x (Redwood 65 x a high oil line of unknown origin)	Canada	1993	L2	CEB
Adin		Rumania	1988	L3	VDB
Agatha	Nanda x Viking	The Netherlands	2001	F1	CEB
Angelin	(NYNKE x VIKING) x (73H30-51-2 x VIKING)	The Netherlands	1997	F2	VDB
Antares		France	1973	L4	CEB
Areco		Argentina		L5	CEB
Argos	Fany x sel.(Natasja x Tomskij 10)	France	1991	F3	CEB
Atalante		France	1981	L6	CEB
Avant Garde		Russia	1968	L7	VDB
Barbara		Hungary	1989	L8	CEB
Belinka	REINA x (ENGELUM 476 x FIBRA)	The Netherlands	1977	F4	CEB
Berber	ROP L 579 x NATASJA	The Netherlands	1982	F5	DOB
Bertelin				F6	VDB
Beryl		Hungary		L9	CSIRO
Blue Chip		Hungary		L10	CEB
Caesar Augustus	Saskia x Viking	The Netherlands	2001	F7	VDB
Capricorn	Linda x McGregor	The Nethelands		L11	CEB
CEB 9601	Marina x Viking	The Netherlands		F8	CEB
Chrystal		Hungary		L12	CEB
Coniston		Canada		L13	CEB
Culbert	Wisdom x Bison 70	USA	1978	L14	CEB
Datcha		France	1976	F9	VDB
Diana	Hercules x sel. Russian seed x Hollandia	The Netherlands	1954	F10	VDB
Diane	[BELINKA x (68-2)97] x FANYx(72-11)48]/(Belinka x (Natasja x T10)) x (Fany x (Silva x Natasja))	France	1995	F11	CEB
Diva		France	1999	F12	CEB
DOB 1987-040-001	McGREGOR X REGINA	The Netherlands		F13	DOB
Electra	ARIANE x Giselle	The Netherlands	1997	F14	CEB
Elise	FANYxROP W1419	The Netherlands		F15	DOB
Engelum 476		The Netherlands		F16	VDB
Escalina	BELINKA x ARIANE	The Netherlands	1995	F17	CEB
Ethiopian		Ethiopia		U2	DOB
Exel	Viking x (Currong x Eva)	France	1996	F18	CEB
Fany	Natasja x Tomsky 10	France	1979	F19	VDB
Flanders	McGregor x Dufferin	Canada		L15	CEB
Formosa	F 6 x Texala	The Netherlands	1949	F20	VDB
Geria	Azur x Olayozon	Rumania	1991	L16	VDB
Giselle	(Hera x Reina)x Natasja	The Netherlands	1982	F21	CEB
Glenelg		Australia		L17	CSIRO
Hella		Germany		L18	CEB
Hera	((WIERA x ST.491) x DIANA) x WIERA	The Netherlands	1968	F22	DOB
Herkules	Selection from Russian landrace	Sweden	1927	F23	VDB
Hermes	FANY x (NATASJA x T10)	France	1991	F24	VDB
Hollandia	F 6 x Texala	The Netherlands	1947	F25	VDB
Ilona	sel.(CEB 7201 x Taiga) x Ariane	The Netherlands	1996	F26	CEB
Klasse		Canada		L19	CEB
Korean		Korea		U3	USDA-ARS (PI522555)
L. bienne		Portugal	n/a	W1	CGN (CGN21296)
L. usitatissimum-A			n/a	U4	CGN (CGN21182)
L. usitatissimum-B			n/a	U5	CGN (CGN21293)
L. usitatissimum-C			n/a	U6	CGN (CGN21295)
L. usitatissimum-D			n/a	U7	CGN (CGN964910)
Laura	BELINKA x NATASJA	The Netherlands	1989	F27	CEB
Linola 989	(McGregor x Zero) x CPI 84495) x 3*McGregor	Canada	1995	L20	CEB
Linton	CI2506 M3M3 P3P3 x Culbert	USA	1986	L21	CEB
Lirina		Germany		L22	CEB
Liviola	GISELLE x NATASJA	Germany		F28	VDB
Lola	E1747 x Somme	Canada	1998	L23	CEB
McGregor	(Rockett x Raja) x Redwood 65	Canada	1981	L24	DOB
Melina	Viking x Nike	The Netherlands	2005	F29	CEB
Nanda	BELINKA x NATASJA	The Netherlands		F32	CEB
Natasja	WIERA x MAPUM	The Netherlands	1972	F33	DOB
Neche	(CI2204 x Foster) x Culbert 79	USA	1988	L27	CEB
Nike		Poland		F34	VDB
Norman	(Noralta x 3 x Redwood) x (Valuta x Raja x 4 x Linott)	Canada	1989	L29	USDA-ARS (PI522366)
Nusis		Germany		L30	CEB

Table 1 Continued

Cultivar name	Parents	Origin	Year of introduction	Code [1]/ Purpose	Source [2]
Midin		Rumania	1977	L25	CEB
Mikael	Tape Parana x(Ocean x Valuta)	France		L26	CEB
Milas		Turkey		U8	USDA-ARS (PI522823)
Modram		Chech Republic		F30	VDB
Mogilevski 2		Russia		F31	VDB
Nynke	((R491xENGELUM 51-7)xDIANA)xWIERA	The Netherlands	1974	F35	DOB
Odin		Rumania		L31	CEB
Omega	CI3036(=L-581) x Flor((BisonM3M3 x Linott) x (BisonP3P3 x Linott))	USA	1990	L32	CEB
Opaline	(CI687xFIBRA) x TOMSKY 10/(60-2)1 x Tomsky 10	France	1984	F36	CEB
Orshankski 2		Russia		F37	VDB
Oscar		France	1997	L33	CEB
Percello	Concurrent x Texala	The Netherlands	1944	F38	VDB
Raisa		Russia		F39	VDB
Raluca	(Tape Parana x Iris) x (Culbert x 5017)	Rumania	1993	L34	VDB
Reconquista		Argentina		L35	USDA-ARS (PI468067)
Regina	REINA x (REINA x FIBRA)	The Netherlands	1976	F40	CEB
Reina	WIERA x (WIERA x (FORMOSA x (FORMOSA x CASCADE)))	The Netherlands	1962	F41	CEB
Saidabad		Iran		U9	USDA-ARS (PI523054)
Saskia	ROP L 574 x NATASJA	The Netherlands	1982	F42	DOB
Silva		France	1977	F43	VDB
Solon	McGregor mutant	Canada		L36	OSU
Somme	(Redwood 65xFoster) x Norlin	Canada	1989	L37	VDB
Start		Russia		L38	VDB
Stormont Zephyr		Northern Ireland	1982	F44	VDB
Super	LUTCH x NATASJA	Chech Republic	1991	F45	DOB
Symphonia		France	1998	L39	CEB
Tape Parana		Argentina		L40	USDA-ARS (PI468069)
Tash		Afganistan		U10	USDA-ARS (PI522823)
Taurus	McGregor x Norlin	The Netherlands	1999	L41	CEB
Texa		Chech Republic	1991	F46	VDB
Torshokski 4		Russia		F47	VDB
USA Golden		USA		L42	DOB
Ussurisk		China		U11	USDA-ARS (PI522577)
VDB 00-01		The Netherlands		F48	VDB
Venus	Viking x (Currong x Fany)	France	1996	F49	CEB
Veralin	TORSHOK x LIDIA	The Netherlands	1999	F50	CEB
Vimy	Kubanskii x Linott	Canada	1986	L43	CEB
Viola	FANY x DATCHA	Northern Ireland	1996	F51	VDB
Virgo	Flanders x Areco	The Nethelands		L44	CEB
W10				L45	CSIRO
Windermere		Canada	1993	L46	CEB
Zarja 87		Russia		F52	VDB
Zoltan		Hungary	1993	L47	CEB

[1] This code is used in PCA bi-plot analysis, where U stands for unknown type, L for linseed, F for fiber and W for wild species

[2] VDB=Van De Bilt Zaden en Vlas
CEB=Cebeco Seeds
DOB=Dobbelaar
OSU=Oregon State University
CGN=Centre for Genetic Resources, the Netherlands
CSIRO=Commonwealth Scientific and Industrial Research Organisation
USDA-ARS=United States Department of Agriculture-Agricultural Research Service

DNA isolation and AFLP analysis

Per genotype leaf material was collected from a single two-week old seedling, grown in the greenhouse. Genomic DNA was extracted from fresh leaves according to the procedure described by Van der Beek *et al.* (1992), adjusted for 96-well format using 1 ml tubes of Micronics (Micronic BV, Lelystad, The Netherlands). Leaf tissue was ground using a Retsch 300 mm shaker at maximum speed (Retsch BV, Ochten, The Netherlands). After

hooking the precipitated DNA from the isopropanol mixture, the DNA was washed overnight in 76% ethanol and 10 mM NH4Ac, dried and dissolved in 100 µl of mQ. The AFLP procedure was performed as described by Vos *et al.* (1995). Total genomic DNA (250ng) was digested using two restriction enzyme combinations, *Pst*I/*Mse*I and *Eco*RI/*Mse*I, and ligated to adaptors. Pre-amplifications were performed with *Pst*I+0/*Mse*I+C and *Eco*RI+A/*Mse*I+C primers. Five µl of the twenty fold diluted pre-amplification product was used as template for the selective amplification using five primer combinations (eAAA/mCTG, pCA/mCTC, pCT/mCTA, pGA/mCCT and pGC/mCTT). Only the *Pst*I and *Eco*RI primers were labeled with IRD-700 or IRD-800 fluorescent dyes at 5´ ends for the selective amplification. After the selective amplification, the reaction products were mixed with an equal volume of formamide-loading buffer (98% formamide, 10 mM EDTA pH 8.0 and 0.1% Bromo Phenol Blue), denatured for 5 minutes at 94°C, cooled on ice and run on a 5.5% denaturing polyacrylamide gel with a NEN® Global Edition IR2 DNA Analyzer (LI-COR® Biosciences, Lincoln, NE).

Data analysis

The software package AFLP-Quantar™ was used to collect data from the AFLP gel image and all clearly distinguishable AFLP bands ranging from 60 bp to 500 bp were treated as dominant markers and scored as either present (1) or absent (0). Phylogenetic trees were constructed using TREECON software (Van der Peer and Wachter, 1994). Genetic distances were calculated according to (Nei and Li 1979). Cluster analysis was performed using the un-weighted pair group method with arithmetic averages (UPGMA, Sneath and Sokal 1973) and Neighbor Joining (NJ, Saitou and Nei 1987). Bootstrap values were calculated on the basis of 1000 permutations and presented as percentages if higher than 50%. A principal coordinate analysis (PCA) was done to complement the information generated from cluster analysis, because cluster analysis is more sensitive to closely related individuals, whereas PCA is more informative regarding distances among major groups (Hauser and Crovello 1982; Sun *et al.* 2001). The results of the PCA were displayed in a biplot

A χ^2 test was used to determine heterogeneity in AFLP band frequencies between the two morphological types of flax using the linkage disequilibrium component in the ARLEQUIN 2.000 software (Schneider *et al.*, 2000). The presence of allelic codominant markers in flax was observed by Vromans *et al.* (Chapter 5) when analyzing two mapping populations. He assigned two polymorphic bands as one allelic codominant marker if the two polymorphic bands comply with the two criteria made by Alonso Blanco *et al.* (1998) (*i*) Two AFLP bands might be allelic when they are derived from different parents, with the same primer combination. (*ii*) The two AFLP markers show a complementary pattern (Alonso Blanco *et al. 1998*).

Analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) was used to partition the total genetic variation to among- and within-population variance components and to calculate the genetic variation (sum of square values) in each population. The populations were defined as fiber, linseed and unknown population. The phi-statistic (Φst), which was used to describe the interpopulation distance (Huff 1997), and its level of significance was also calculated by the AMOVA component in ARLEQUIN 2.

Results

AFLP fingerprints

The five AFLP primer combinations generated a total of 365 scorable bands among the 111 accessions studied, 199 (54.5%) of which were polymorphic. The information generated for each primer combination is shown in Table 2. The primer combinations show variation in complexity and number of polymorphic bands. The pCT/mCTA primer combination produced the largest number of total bands as well as the largest number of polymorphic bands, which is probably due to the lower CG-content in the selective nucleotides of this primer combination, as compared to the other *Pst*I/*Mse*I primer combination. However, this primer combination did not have the highest percentage of polymorphisms. This implies that the complexity of the fingerprint rather than the percentage of polymorphic bands determines the discriminative power of a primer combination. Every accession could be uniquely genotyped by the polymorphic bands observed in this study. In total 49 singletons (i.e. a fragment present or absent in only one accession) were observed, 25 of which originated from only one *L bienne* accession, whereas the remaining 24 derived from the other accessions. A high number of unique bands are expected for the *L bienne* accession, as it is the only wild species included in this study.

Table 2 Number of bands and corresponding polymorphic rate generated by different primer combinations and the presence of singletons

Primer combination	Total # of bands	# of polymorphic bands	% Polymorphic bands	# of *L. bienne* singletons	# of singletons from other accessions	# of informative bands
eAAA/mCTG	66	40	61	7	5	28
pCA/mCTC	84	50	60	6	6	38
pCT/mCTA	103	57	55	7	8	42
pGA/mCCT	54	24	44	2	3	19
pGC/mCTT	58	28	48	3	2	23
Total	365	199	55	25	24	150

Comparison of AFLP patterns between related accessions

Five fiber flax cultivars and three linseed cultivars have both their parents included in the analysis. AFLP scores of these eight cultivars were compared with those of their parents, and in six cases ambiguities arose. Either AFLP markers were not present in the two

parents but present in the offspring, or AFLP markers were present in both parents but not in the offspring. The latter ambiguity could have resulted from crossing two heterozygous parents, although this is not likely for an inbreeder. Only the AFLP patterns of cultivars Laura and Virgo agreed with pedigree information while in Nanda, Lola and Viola only a few markers were problematic (one, one and three markers respectively). However, breeding line dob1987040001 (nine), cultivar Taurus (six) and especially Liviola (17) showed high numbers of inexplicable markers.

Genetic relationship among the accessions

The dendrograms constructed by the UPGMA and NJ clustering analysis were statistically compared by calculating the cophenetic correlation coefficients. A high correlation (r=0.95 and r=0.93 for NJ and UPGMA respectively) was observed between the calculated distance matrices and both distance matrices obtained from the dendrograms (Figure 1, UPGMA not shown), which indicated a minimal distortion caused by the clustering procedure and suggesting a good representation of the real topology (Mohammadi and Prasanna 2003). As the higher correlation coefficient for NJ over UPGMA implies that NJ represents the calculated distances slightly better, NJ clustering was subsequently used in this study.

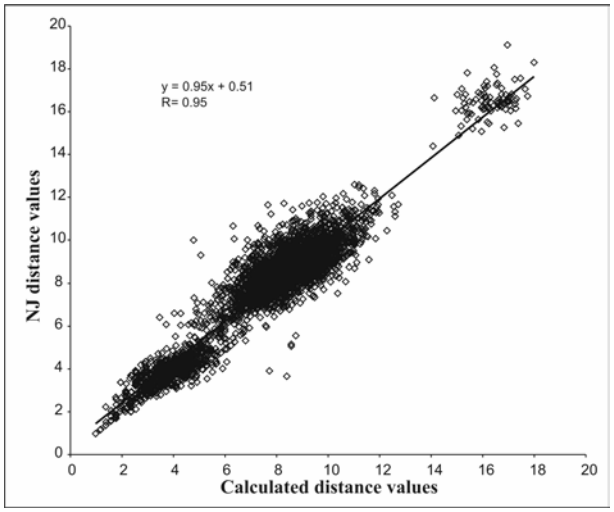

Figure 1 Cophenetic correlation between the NJ distance matrix and the distance matrix obtained by calculated values from the dendrograms

Cluster analysis showed a clear separation of *L. bienne* accession from the other accessions, supported by a high bootstrap value (Figure 2). A close inspection of the dendrogram reveals that some branches are small groups supported by a significant bootstrap value (Figure 2). In general these groups contain only two accessions, which share pedigree information (Table 1). However, some larger significant supported groups could be identified as well. An example is a group of three Dutch fiber flax cultivars

(Berber, Laura and Nanda) all sharing Natasja as the father (two of them even sharing both parents; Table 1).

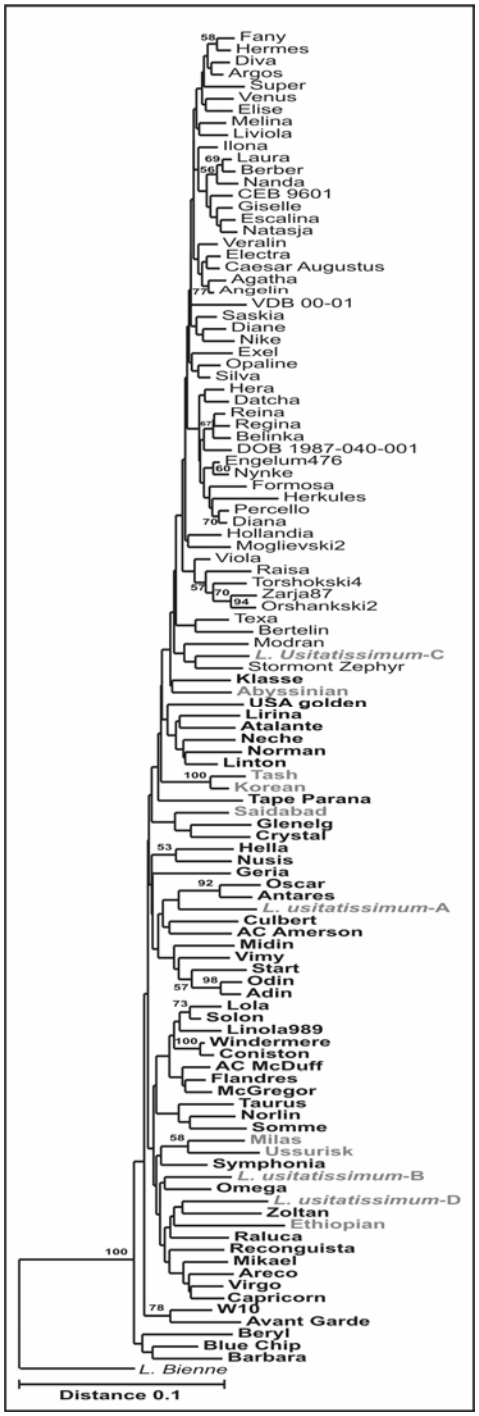

Figure 2 Neighbor Joining phenetic tree of relationships between the 111 *Linum* accessions based on Nei and Li's genetic distance. Only bootstrap values above 50% are presented.

Similarly, another three Dutch fiber flax cultivars (Reina, Regina and Belinka, Regina and Belinka sharing Reina as mother) nicely cluster together. Also a group of three East-European linseed cultivars (Odin, Adin and Start) comprise such a small group. Because the genotypes in these small fiber flax clusters share part of the breeding history their

grouping is not surprising. This is not a general rule however: cultivars with an overlapping breeding history do not always group together, for example Fany and Laura who share Natasja as parent. Apparently selection for a desirable phenotype (the phenotype of the common parent) is the main reason for flax cultivars, with an overlapping breeding history, to group together. Obviously the fiber flax cultivars cluster separately from the linseed and the unknown purpose accessions, although not supported by a high bootstrap value. The fiber flax cluster seems to be a subset of the population of linseed cultivars (Figure 2).

Another way to visualize relationships between accessions and differentiation between the two morphological groups of flax is to perform a PCA. The PCA analysis showed that except for a few genotypes a clear differentiation between fiber flax and linseed can be observed indeed (Figure 3). The majority of the variation (16.24%) is explained by the first principle component and in total this biplot explained approximately 22% of the variation.

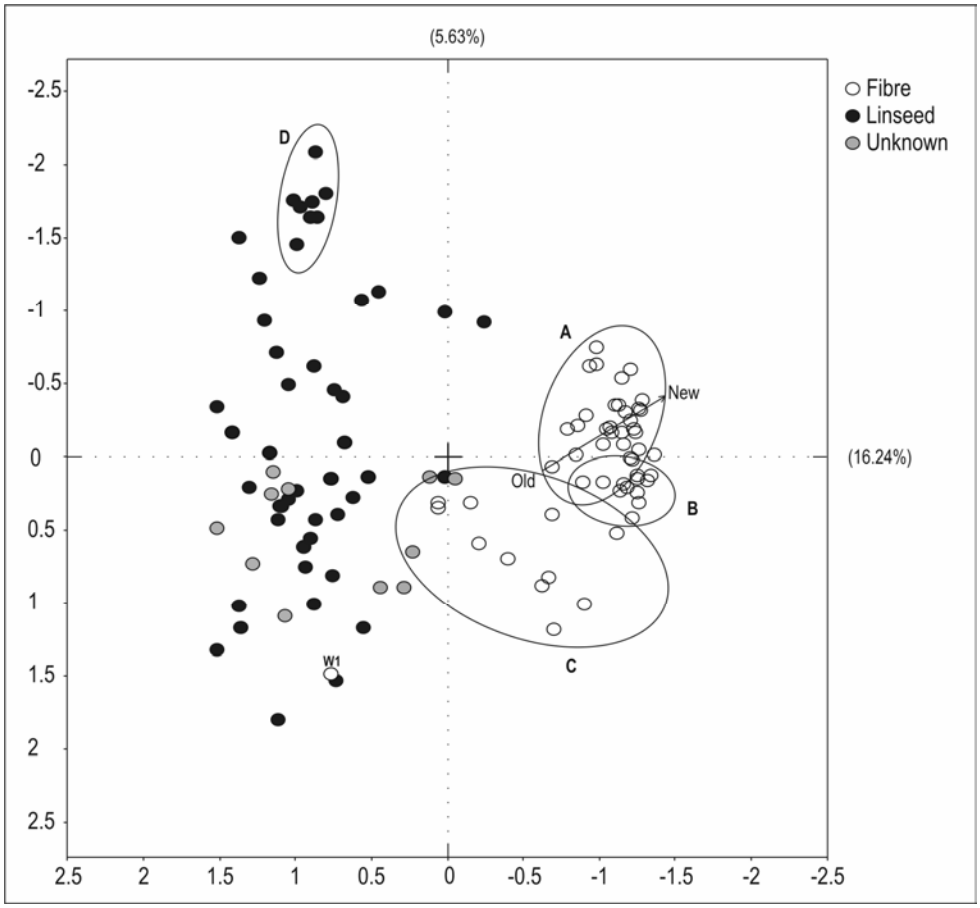

Figure 3 Biplot of a PCA analysis. A represent Dutch fiber flax cultivars, B French fiber flax cultivars, C fiber flax cultivars from other countries and D are Canadian linseed cultivars. W1 is *L. bienne*. Further explanation in the text.

On the other hand *L. bienne* seemed to be related to linseed and not really distinguishable from linseed as revealed by the dendrogram. However, if the third or fourth principle component is plotted, *L. bienne* is clearly separated from both linseed and fiber flax, but

still more related to linseed cultivars (data not shown). It also showed that the Dutch fiber flax varieties are slightly differentiated from the French cultivars and clearly different from the fiber flax cultivars from other countries (A, B and C in Figure 3). Except for a small group of Canadian cultivars (D in Figure 3) the effect of geographic origin on linseed was very small, as linseed cultivars of different origin were distributed over the whole range of linseed cultivars. Finally, the analysis indicates that fiber flax breeding activities in the Netherlands during the last 50 years have led to a further differentiation of fiber flax from linseed, which is indicated by the arrow in Figure 3.

Both the dendrogram and the bi-plot showed that all unknown purpose accessions group in the cluster of linseed cultivars. Inter-population distance estimates (Φ_{st}) among the three described populations of accessions confirmed that the population comprised of unknown purpose accessions is much closer to linseed cultivars as compared to fiber flax cultivars (Table 3). Therefore it can be assumed that both the linseed and the unknown purpose population represent the same genepool.

Table 3 Inter-population distance (phi-statistic = Φ_{st}) estimates among the three populations detected by AMOVA from AFLP data. Significance of each Φ_{st} value is based on a permutation test (n=1000)

Comparison	Inter-population distance Φ_{st}	P value
Fiber vs. Linseed	0.259	***
Fiber vs. Unknown	0.284	***
Linseed vs. Unknown	0.044	***

*** significant at P < 0.001

Markers associated with the difference between linseed and fiber flax

Although we have not observed AFLP markers that were absolutely specific for linseed or fiber flax, some markers showed strong association with the two morphological groups (Table 4). Out of the 163 polymorphic markers (polymorphic in the 99 fiber and linseed cultivars), 42 markers were significantly associated (after Bonferroni correction for multiple testing) with the morphological classes (Table 4). As expected these markers also had the highest influence on the differentiation between linseed and fiber cultivars in the PCA (Data not shown). In general these markers showed association among themselves as well (data not shown), which means that linkage disequilibrium (LD) prevails among these markers. Of the 42 markers associated with morphology five pairs of markers are allelic. Allelism was corroborated by the observation that these marker pairs were from the same primer combination and differing only a few nucleotides in size. Except for codominant marker 1 (pCA/mCTC-295.7 and pCA/mCTC-297.5, Table 4), the summed frequency of the two alleles of all codominant markers either was less than (two cases) or exceeded 1 (four cases).

Table 4 χ^2 test of class heterogeneity of AFLP band frequency, and band frequencies in the morphological classes of flax, based on data from the 52 fiber and 47 linseed genotypes

AFLP-marker	p-value	Fiber	Linseed
pGC/mCTT-219.8	2.85×10^{-04}	0.48	0.83
pCT/mCTA-238.1	2.73×10^{-04}	0.13	0.47
pCT/mCTA-118.3	2.45×10^{-04}	0.98	0.72
pCT/mCTA-84.8	2.45×10^{-04}	0.98	0.72
pGA/mCCT-71.4	2.19×10^{-04}	0.77	0.40
pCT/mCTA-252.5	2.14×10^{-04}	0.00	0.23
eAAA/mCTG-75.6	1.67×10^{-04}	0.06	0.36
pGC/mCTT-122.0	1.43×10^{-04}	0.54	0.17
pCA/mCTC-295.7 (1)	1.12×10^{-04}	0.98	0.70
pCA/mCTC-297.5 (1)	1.12×10^{-04}	0.02	0.30
pCT/mCTA-255.7	1.02×10^{-04}	1.00	0.74
pCT/mCTA-144.0	5.21×10^{-05}	0.98	0.68
pCT/mCTA-183.3	4.70×10^{-05}	1.00	0.72
pCT/mCTA-60.0	2.34×10^{-05}	0.92	0.55
pCA/mCTC-107.0	1.43×10^{-05}	0.83	0.40
pCT/mCTA-441.4	1.03×10^{-05}	0.65	0.21
pGA/mCCT-401.2	9.72×10^{-06}	0.00	0.32
pCT/mCTA-495.0	8.54×10^{-07}	0.48	0.94
eAAA/mCTG-181.8	8.18×10^{-07}	0.00	0.38
pCT/mCTA-418.0 (2)	6.82×10^{-07}	0.92	0.47
pGA/mCCT-100.0	6.26×10^{-07}	0.04	0.47
pCT/mCTA-457.9	4.36×10^{-07}	0.06	0.51
pCT/mCTA-205.0	3.39×10^{-07}	0.00	0.40
pCT/mCTA-420.2 (2)	2.60×10^{-07}	0.08	0.55
pGC/mCTT-277.6	1.67×10^{-07}	0.06	0.53
eAAA/mCTG-300.6	1.41×10^{-07}	0.90	0.40
eAAA/mCTG-206.9	1.29×10^{-07}	0.02	0.47
pCT/mCTA-175.2	9.67×10^{-08}	0.04	0.51
pCT/mCTA-426.7	3.80×10^{-08}	0.33	0.87
pGC/mCTT-436.8	3.71×10^{-08}	0.04	0.53
pGA/mCCT-417.3	2.29×10^{-08}	0.94	0.43
pCA/mCTC-427.0	8.56×10^{-09}	0.00	0.49
pCA/mCTC-348.0	3.74×10^{-09}	0.37	0.94
pGC/mCTT-169.6 (3)	1.18×10^{-09}	1.00	0.47
pGC/mCTT-157.5 (3)	8.97×10^{-10}	0.02	0.57
pCA/mCTC-342.4 (4)	1.37×10^{-11}	0.81	0.13
pGA/mCCT-341.4	4.20×10^{-12}	0.88	0.19
pCA/mCTC-344.3 (4)	7.82×10^{-13}	0.17	0.89
pCA/mCTC-234.8	3.58×10^{-14}	0.94	0.19
pCT/mCTA-154.6	2.70×10^{-14}	0.04	0.79
pCT/mCTA-172.6 (5)	1.23×10^{-17}	0.04	0.89
pCT/mCTA-170.6 (5)	2.19×10^{-18}	0.96	0.09

The nominal significance level was set at $p \leq 0.05$. The significance level after the Bonferroni correction for multiple tests is $p \leq 3.09 \times 10^{-4}$. Markers that are labeled with a common number in brackets are considered to be alleles of the same locus.

Estimation of population genetic variances

Analysis of molecular variance revealed that most of the variation (77.3 %) is due to differences within the 'purpose populations' (Table 5). In accordance with the dendrogram and biplot fiber flax had the smallest amount of variation. The within population variation of the linseed and unknown population was much larger and in the same range (Table 5).

Table 5 Analysis of molecular variance (AMOVA) of all genotypes except *L. bienne*

Source of variation	d.f.	Sum of squares	Variance components	% of variation	Probability
Among populations	2	344.02	4.84	22.7	<0.001
Within populations	107	1761.66	16.46	77.3	<0.001
Within fiber	51	604.57	11.85		
Within linseed	46	929.45	20.21		
Within unknown	10	227.64	22.76		
Total	109	2105.68	21.30		

Discussion

Modern fiber flax cultivars have a narrow genetic basis

Experts dealing with testing plant cultivars for Distinctness, Uniformity and Stability (DUS) criteria or Plant Breeders' right have noticed that the morphological variation between new cultivars has decreased (Everaert *et al.* 2001). This points towards a narrow genetic basis of modern flax cultivars. Dutch flax breeders have arrived at the same conclusion as they experienced an increasing difficulty to develop new cultivars with improved fiber content, fiber quality and other important traits (flax breeders, pers. comm..). In this study a large set of flax genotypes has been evaluated with AFLPs to obtain insight in the genetic diversity between and within both, fiber flax and linseed cultivars, as well as the effect of breeding activities on the genetic diversity in fiber flax. The results of this study, combined with phenotypic characteristics of cultivars, can support the selection of parents in flax breeding programs.

Inconsistent parent-offspring AFLP fingerprints

We observed a number of inconsistencies when comparing the AFLP signatures of cultivars with those of their parents. In case the ambiguities were few (less than four, say) these might be due to artifacts. Another cause of such apparent inconsistencies may be the way flax cultivars are created. A flax cultivar may consist of a number of phenotypically similar F5 or F6 plant derived sister lines (H. de Jong, pers. comm.) that, at the molecular level, still may differ for a small proportion of the genome. When single plants are used for DNA fingerprinting (as we did), such a heterogeneity may, of course, result in non-matching band patterns of parental and offspring cultivars. The large number of unexpected bands in breeding line dob19877040001 (nine) and cultivar Taurus (six) suggests that either the parents were heterogeneous, or the pedigree information is incorrect, or outcrossing occurred during development of the cultivar/breeding line.

The latter phenomenon has been discussed by both Evenaert *et al.* (2001) and Van Treuren *et al.* (2001) who also observed a considerable heterogeneity within cultivars. Van Treuren *et al.* (2001) concluded that this high intra-accession variation is not in line with the low outcrossing rates observed in flax. The rate of outcrossing in flax generally has

been estimated to be less than 6%, although higher values may be observed due to phenotypic variation in flower morphology (Williams 1988). Safety measures have been recommended by plant breeders to prevent outcrossing, but so far these precautionary measures during the propagation of flax have not yet been implemented by gene banks (Van Treuren *et al.* 2001).

Flax breeding and similarity of cultivars

The general topology of the NJ dendrogram showed that the fiber flax cultivars are differentiated from linseed cultivars (Figure 2), but not significantly, taking into account that internal tree branches that have a bootstrap value of 70% and higher are correct at 95% probability (Mohammadi and Prasanna 2003). A study in *Brassica rapa* comprising accessions of several cultivar groups showed that clustering between the morphotypes was possible, but this was not supported by a significant bootstrap value (Zhao *et al.* 2005). However, PCA did show a clear differentiation and also high inter-population distance supported the differentiation between the two morphological groups. The clearly distinguishable groups of fiber and linseed reported in this study with the use of molecular markers is in agreement with the cluster analyses of Fu *et al.* (2002) and Månsby *et al.* (2000), but not with those of Everaert *et al.* (2001) and Wiesnerova and Wiesner (2004) who found no clear separation of linseed and fiber flax cultivars.

Although the analysis did not result in a significant clustering of the fiber flax cultivars (Figure 2), still the fiber flax population seems to be a subset of the population of linseed cultivars. This confirms that the deviation of fiber flax from linseed cultivars is caused by recent breeding activities focused on fiber flax characteristics. This in turn is in agreement with the hypothesis that fiber flax has evolved out of linseed (Zohary and Hopf 2000). The consequence of this hypothesis is that the oldest fiber flax cultivars are more close to linseed than new varieties. The PCA analysis showed that for Dutch fiber flax cultivars this seems to be the case (Figure 3).

Linkage disequilibrium

Although not a single marker was found to be entirely specific for the fiber flax or linseed cultivar groups, several markers were significantly associated with the morphotypes of *L. usitatissimum* (Table 4). Several of the markers used in the present study have recently been located on a linkage map (Chapter 5). Two of the markers associated with morphotype are located on different linkage groups. This indicates that different regions of the genome cause the morphological differences between fiber flax and linseed. This also means that spurious linkage disequilibrium (LD) has been observed, which is indeed caused by population structure of the flax gene pool. This phenomenon has been previously reported in maize and durum wheat (Liu *et al.* 2003; Maccaferri *et al.* 2005; Remington *et al.* 2001).

These authors observed a decrease of LD if subpopulations were taken into account, as compared to considering the germplasm collection as a whole. Maccaferri *et al.* (2005) observed high levels of LD in durum wheat both at locus pairs with an intrachromosomal distance <50 cM as well as at those with distances over 50 cM and unlinked ones, if the germplasm collection is considered as a whole. On the other hand, after re-evaluating LD taking into account the population structure, the level of LD remained high for tightly to moderately linked locus pairs (<20 cM), but greatly reduced for the loosely linked and independent locus pairs. Therefore, in future association mapping studies in flax, the population structure should be duly considered.

After a thorough selection of the most polymorphic primer combination, allelism of AFLP markers is commonly observed in flax (Chapter 4). In mapping populations two markers are bi-allelic codominant markers if an alternated pattern can be seen in the same fingerprint, double absent doesn't occur and heterozygous genotypes contain both bands in half intensity (Alonso Blanco *et al.* 1998). Bi-allelic codominant markers can easily be observed in mapping populations. However in a genetic diversity study some caution should be taken as double absence could occur if several alleles exist in the germplasm but these do not appear in the same fingerprint. Furthermore, heterozygosity of a codominant marker should be duly considered as the frequency of the two alleles of a codominant marker exceeded 1 in four cases.

Broadening the genetic basis of flax

AMOVA showed that although most of the variation (77.3 %) was attributable to differences within the populations, still a fair percentage (21.7 %) could be ascribed to variation among populations (Table 3). This is in contradiction with the results obtained by Eveaert *et al.* (2001), as he found little variation among populations. These different results might be caused by the large within cultivar variation observed by *Eveaert et al.* (2001), which was not calculated in the analysis presented in this paper. Considering the variance components of the different groups, the linseed and unknown purpose groups contain a similar amount of variation. Contrastingly, the fiber flax group contains a must smaller amount of variation. Therefore, we assume that recent plant breeding activities in flax have narrowed the genetic basis of flax, especially that of fiber flax. The narrowed genetic diversity in plants is a common consequence of modern plant breeding activities (Tanksley and McCouch 1998). In fiber flax a further decrease of the variation will take place if fiber flax breeders continue to use modern cultivars as parents in their breeding program. Consequently, the development of distinguishable cultivars with improved characteristics will become increasingly difficult. The use of germplasm from wild flax species may result in the introduction of novel favorable alleles. The molecular genetic tools that are currently available enable a quick and efficient transfer of favorable alleles from wild species.

Especially the backcross introgression line strategy has been shown to be of great value to plant breeding (Eshed *et al.* 1992; Jeuken and Lindhout 2004).

Conclusion

In conclusion, the present study has demonstrated that breeding activities have led to a narrow genetic basis of fiber flax. The development of new cultivars with improved characteristics will hardly be feasible if breeders continue to use modern cultivars as parents in their breeding programs. Linseed cultivars and wild flax species harbor more genetic diversity and should be considered as good genetic resources to enlarge the gene pool and introduce new favorable alleles in fiber flax breeding programs.

Acknowledgements

We are very grateful to Hein de Jong for providing valuable comments on the manuscript. We thank Marcos Malosetti for his statistical assistance. Seeds were kindly provided by the breeding companies ‘Van De Bilt Zaden en Vlas’, Cebeco Seeds and Dobbelaar, the Dutch Centre of Genetic Resources, CGN (Wageningen, the Netherlands, USDA-ARS Plant Introduction Station Iowa State University (USA), CSIRO Plant Industry in Canberra (Australia) and Dr. Steve Knapp (Oregon State University, Canada). We gratefully acknowledge the Dutch breeding campanies Cebeco Seeds, ‘Van de Bilt Zaden en Vlas’ and Dobbelaar Breeding and the Dutch provinces Zeeland for their financial contribution.

References

Alonso Blanco, C., Peeters, A.J.M., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., Pot, J. and Kuiper, M.T.R. (1998) Development of an AFLP based linkage map of *Ler*, *Col* and *Cvi Arabidopsis thaliana* ecotypes and construction of a *Ler*/*Cvi* recombinant inbred line population. *Plant Journal* 14: 259-271

De Haan, H. (1952) Flax breeding and flax varieties in the Netherlands. *Euphytica* 1: 212-218

Eshed, Y., Abu Abied, M., Saranga, Y. and Zamir, D. (1992) *Lycopersicon esculentum* lines containing small overlapping introgressions from *L. pennellii*. *Theoretical and Applied Genetics* 83: 1027-1034

Everaert, I., de Riek, J., de Loose, M., van Waes J., and van Bockstaele, E. (2001) Most similar variety grouping for distinctness evaluation of flax and linseed (*Linum usitatissimum* L.) varieties by means of AFLP and morphological data. *Plant Varieties and Seeds* 14: 69-87

Excoffier, L., Smouse, P.E. and Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491

Fu, Y.B., Diederichsen, A., Richards, K.W. and Peterson, G. (2002) Genetic diversity within a range of cultivars and landraces of flax (*Linum usitatissimum* L.) as revealed by RAPDs. *Genetic Resources and Crop Evolution* 49: 167-174

Fu, Y.B., Rowland, G.G., Duguid, S.D. and Richards, K.W. (2003) RAPD analysis of 54 North American flax cultivars. *Crop Science* 43: 1510-1515

Fu, Y.B. (2005) Geographic patterns of RAPD variation in cultivated flax. *Crop Science* 45: 1084-1091

Hauser, L.A. and Crovello, T.J. (1982) Numerical analysis of generic relationships in *Thelypodieae* (*Brassicaceae*). *Systematic Botany* 7: 249-268

Huff, D.R. (1997) RAPD characterization of heterogeneous perennial ryegrass cultivars. *Crop Science* 37: 557-564

Jeuken, M.J.W. and Lindhout, P. (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. *Theoretical and Applied Genetics* 2004; 109: 394-401

Kardolus, J.P., van Eck, H.J. and van de Berg, R.G. (1998) The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (*Solanaceae*). *Plant Systematics and Evolution* 210: 87-103

Liu, K.J., Goodman, M., Muse, S., Smith, J.S., Buckler, E. and Doebley, J. (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165: 2117-2128

Maccaferri, M., Sanguineti, M.C., Noli, E. and Tuberosa, R. (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Molecular Breeding* 15: 271-289

Månsby, E., Diaz, O. and von Bothmer, R. (2000) Preliminary study of genetic diversity in Swedish flax (*Linum usitatissimum*). *Genetic Resources and Crop Evolution* 47: 417-424

Mohammadi, S.A. and Prasanna, B.M. (2003) Analysis of genetic diversity in crop plants - salient statistical tools and considerations. *Crop Science* 43: 1235-1248

Nei, M. and Li W.H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* 76: 5269-5273

Park, Y.H., West, M.A.L. and St Clair, D.A. (2004) Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum* L.). Genome 47: 510-518

Powell, W., Orozcocastillo, C., Chalmers, K.J., Provan, J. and Waugh, R. (1995) Polymerase chain reaction-based assays for the characterization of plant genetic resources. *Electrophoresis* 16: 1726-1730

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. and Rafalski, A. (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2: 225-238

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doeblay, J., Kresovich, S., Goodman, M.M. and Buckler, E.S. (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America* 98: 11479-11484

Saitou, N. and Nei, M. (1987) The neighbor-joining method a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution 4*: 406-425.

Schneider, S., Roessli, D. and Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Sneath, P.H.A. and Sokal, R.R. (1973) Numerical taxonomy: the principles and practice of numerical classification. Freeman, San Francisco: 573 p

Soleimani, V.D., Baum, B.R. and Johnson, D.A. (2002) AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat (*Triticum turgidum* L. *subsp. durum* (Desf.) Husn.). *Theoretical and Applied Genetics* 104: 350-357.

Sun, G.L., William, M., Liu, J., Kasha, K.J. and Pauls, K.P. (2001) Microsatellite and RAPD polymorphisms in Ontario corn hybrids are related to the commercial sources and maturity ratings. *Molecular Breeding* 7: 13-24

Tanksley, S.D. and McCouch, S.R. (1998) Seed banks and molecular maps: unlocking genetic potential from the wild. *Crop Genetic Resources*: 49-52

Van de Peer Y. and de Wachter, R. (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Computer Applications in the Bioscience* 10, 569-570

Van der Beek, J.G., Verkerk, R., Zabel, P. and Lindhout, P. (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. *Theoretical and Applied Genetics* 84: 1-2

Van Treuren, R., van Soest, L.J.M. and van Hintum, T.J.L. (2001) Marker-assisted rationalisation of genetic resource collections: a case study in flax using AFLPs. *Theoretical and Applied Genetics* 103: 144-152

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414

Wiesnerova, D. and Wiesner, I. (2004) ISSR-based clustering of cultivated flax germplasm is statistically correlated to thousand seed mass. *Molecular Biotechnology* 26: 207-214

Williams, I.H. (1988) The pollination of linseed and flax. *Bee World* 69: 145-152

Zhao, J.J., Wang, X.W., Deng, B., Lou, P., Wu, J., Sun, R.F., Xu, Z.Y., Vromans, J., Koornneef, M., Bonnema, G. (2005) Genetic relationships within *Brassica rapa* as inferred from AFLP fingerprints. *Theoretical and Applied Genetics* 110: 1301-1314

Zhu, J., Gale, M.D., Quarrie, S., Jackson, M.T. and Bryan, G.J. (1998) AFLP markers for the study of rice biodiversity. *Theoretical and Applied Genetics* 96: 602-611

Zohary, D. and Hopf, M. (2000) Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Oxford University Press, Oxford: 316 p

Chapter 3

The molecular genetic variation in the genus *Linum*

Jaap Vromans, Eugenie van de Bilt, Piet Stam and Herman J. van Eck

Abstract

The molecular genetic diversity among 27 *Linum* spp was investigated using the AFLP method. The analysis included the commercially important species flax (*Linum usitatissimum* L.) and the ornamentals *L. grandiflorum* and *L. perenne*. Based on visual AFLP fingerprint similarity ten groups could be identified. The Neighbor Joining cluster analysis also showed clear significant groups, which corresponded very well with the grouping based on visual fingerprint similarity. The results showed that AFLP markers could be used to group species and to differentiate between groups within the genus *Linum*, but not at the level of sections of the genus. In three groups the differentiation of multi-accession species (population A) compared to the other species (population B) present in the group was studied in more detail by means of AMOVA. Obviously, the majority of the variation could be assigned to within population variation. However, some among population variation could be observed between the *Linum bienne* population and *L. usitatissimum* population. These results indicate that some species within group A, E, and F can be lumped into fewer species. Finally, the smallest amount of variation was present in the *L. usitatissimum* population. If breeding for new fiber flax and linseed cultivars becomes more difficult because of the lack of favorable alleles in the current flax cultivars, *L. bienne* and *L. angustifolium* will be good candidates for the efficient introduction of wild germplasm.

Introduction

The genus *Linum*, a representative of the Linaceae family, comprises about 300 species (Hickey and King 1988). The taxonomy of the genus is still unclear, although a subdivision in five sections (*Linum*, *Dasylinum*, *Linastrum*, *Syllinum* and *Cathartolinum*) is generally accepted (Ockendon and Walters 1968). Rogers (1982) proposed to divide section *Linopsis* (*Linastrum*) into five subsections. Only some species from the section *Linum* are used for agronomical purposes. Where *Linum grandiflorum* Desf. and *L. perenne* L. serve as ornamentals, cultivated flax (*L. usitatissimum*) is commercially grown for its fiber and linseed oil.

Crossing barriers could help to delineate the species within the genus. In 1922 the first interspecific progeny was successfully obtained from a cross between *L. usitatissimum* and *L. angustifolium* (Tammes 1923). Later, other studies showed varying degrees of pollen and seed sterility of hybrids obtained from crosses between species with either a chromosome number of 30 or 18 (Gill and Yermanos 1967a; Gill and Yermanos 1967b; Seetharam 1972). Interspecific crosses between varying chromosome numbers did not result in a fertile progeny (Seetharam 1972). The fertility of crosses between species of the *L. perenne* group has been investigated extensively by Ockendon (1968). He concluded that discontinuities in chromosome number and breeding system provides more satisfactory characters for classification of members of the *L. perenne* group than some morphological characters used before.

Additional to morphological characters, fatty acid composition of the linseed oil obtained from the seeds is also considered relevant for taxonomical classification (Plessers 1966; Rogers 1972; Yermanos *et al.* 1966). After reviewing the data for fatty acid composition of seed oil of *Linum* species, Rogers (1972) concluded that species of section *Linum*, section *Dasylinum* (Planch.) Juz. and old world species of section *Linastrum* (Planch.) Bentham generally had lower percentages of linoleic acid and higher percentages of linolenic acid than of species of section *Syllinum* Griseb., section *Cathartolinum* (Reichenb.) Griseb., and New World species of section *Linastrum*. Recently other seed traits, e.g. tocopherol and plastochromanol, were shown to have important chemotaxonomic value for the sections *Linum* and *Syllinum* (Velasco and Goffman 2000).

In molecular phylogenetic studies at the genus level and above, several DNA sequences are available, from nuclear to plastid genomes. To study more closely related taxa or species, DNA sequences often do not show sufficient variation. At this lowest taxonomic level AFLP is one of the most commonly used marker techniques (Vos *et al.* 1995). This method is very useful in phylogenetic studies because of its high number of polymorphisms per assay, the high reproducibility and the fact that no prior sequence information is required. However, to be suitable for phylogenetic studies, AFLPs, and other molecular markers, should meet two basic requirements: (1) the fragment must have

evolved independently; and (2) fragments of equal length must be homologous. When AFLPs are used at the interspecific level the fragment homology might be problematic depending on the species (reviewed by Koopman 2005).

AFLPs have been applied to many different taxa to resolve phylogenetic relationships and they proved to be complementary to internal transcribed spacer (ITS) sequences in assessing genetic diversity. A study in *Soldanella* (Zhang *et al.* 2001) showed that ITS sequences proved to be useful in studying subgeneric relationships, whereas AFLPs were suitable for examining relationships among more closely related species. The same conclusions were drawn from a study in sweet potato (*Ipomoea batatas* (L.) Lam.), where ITS sequences could distinguish the species at the subgeneric level and AFLPs were able to discriminate between the species in *Ipomoea ser. Batatas* (Huang *et al.* 2002). A phylogenetic study in *Lactuca* s.l. indicated that AFLPs have sufficient phylogenetic signal, although AFLPs and ITS sequences gave conflicting results in parts of the most parsimonious trees (Koopman 2005). In this paper Koopman describes that AFLPs are indeed a valuable source for phylogenetic information, but a more detailed analysis of the non-independence of AFLP markers and a better weighing scheme for the construction of phylogenetic trees would be valuable.

A small study, comprising seven species of the section *Linum*, has been described by (Fu *et al.* 2002) using RAPDs. They found additional support for the hypothesis that *L. angustifolium* (synonymous to *L. bienne*) is the wild progenitor of cultivated flax. The objective of this study is to understand the relationships of 27 species of the genus *Linum*, representing four sections, using AFLPs.

Material and methods

Plant material

A collection of 52 accessions (Table 1) representing 27 *Linum* spp was examined with AFLP. The majority of the material was obtained from the IPK genebank in Gatersleben (Germany) and some accessions were from the Dutch genebank (CGN). Two accessions were bought in a garden shop and the fiber flax variety Hermes was provided by Cebeco Seeds.

DNA isolation and AFLP analysis

Leaf material was collected from a single two-week old seedling, grown in the greenhouse. Genomic DNA was extracted from fresh leaves according to the procedure described by (Van der Beek *et al.* 1992), adjusted for 96-well format using 1 ml tubes of Micronics (Micronic BV, Lelystad, The Netherlands). Leaf tissue was ground using a Retsch 300 mm shaker at maximum speed (Retsch BV, Ochten, The Netherlands). After hooking the DNA out of the isopropanol mixture, the DNA was washed overnight in 76% ethanol and 10 mM

NH4Ac, dried and dissolved in 100 µl of mQ. The AFLP procedure was performed as described by (Vos *et al.* 1995). Total genomic DNA (250ng) was digested using two restriction enzymes, *Pst*I and *Mse*I and ligated to adaptors. Pre-amplifications were performed in a volume of 20 µL of 1 x PCR buffer, 0.2mM dNTPs, 30ng *Pst*I and *Mse*I+C primer, 0.4 units Taq polymerase and 5µL of a 10x diluted restriction ligation mix, using 24 cycles of 94°C for 30s, 56°C for 30s and 72°C for 60s. Five µl of the diluted (1:20) pre-amplification product was used as template for the selective amplification using two primer combinations (pGC/mCTT and pCA/mCTC). Only *Pst*I primers were labeled with IRD-700 or IRD-800 at 5´ end for the selective amplification. The selective amplification was carried out with a touch down PCR with following cycling parameters: 12 cycles of 30s at 94°C, 30s at 65°C-56°C (decrease 0.7°C each cycle), and 60s at 72°C, followed by 24 cycles of 30s at 94°C, 30s at 56°C, and 60s at 72°C. After the selective amplification, the reaction products were mixed with an equal volume of formamide-loading buffer (98% formamide, 10 mM EDTA pH 8.0 and 0.1% Bromo Phenol Blue). The samples were denatured for 5 minutes at 94°C, cooled on ice and run on a 5.5% denaturing polyacrylamide gel with a NEN® Global Edition IR2 DNA Analyzer (LI-COR® Biosciences, Lincoln, NE).

Table 1 List of accession with accompanying information used in this study.

Accession name	Source	Genebank number	Origin	Section	Chromosome number (2n)
L. alpinum Jac.	IPK	LIN 1905	Austria	*Linum*	18 (Ray 1944)
L. altaicum Ledeb.	IPK	LIN 1638	Unknown	*Linum*	18 (Ray 1944)
L. angustifolium Huds.	CGN	CGN21962	Unknown	*Linum*	30 (Ray 1944)
L. austriacum L.	IPK	LIN 1608	Germany	*Linum*	18 (Ray 1944)
L. austriacum L.	IPK	LIN 1831	Unknown	*Linum*	18 (Ray 1944)
L. austriacum L.	IPK	LIN 1874	France	*Linum*	18 (Ray 1944)
L. bienne Mill.	CGN	CGN21296	Portugal	*Linum*	30 (Muravenko *et al.* 2003)
L. bienne Mill.	CGN	CGN21294	Belgium	*Linum*	30 (Muravenko *et al.* 2003)
L. bienne Mill.	CGN	CGN21961	Italy	*Linum*	30 (Muravenko *et al.* 2003)
L. campanulatum L.	IPK	LIN 1760	Unknown	*Syllinum*	28 (Ray 1944)
L. capitatum Kit. ex Schulte	IPK	LIN 1903	Unknown	*Syllinum*	28 (Ray 1944), 34 (Chennaveeraiah and Joshi 1983)
L. decumbens Desf.	CGN	CGN981206	Unknown	*Linum*	30 (Gill and Yermanos 1967a)
L. decumbens Desf.	IPK	LIN 1754	Italy	*Linum*	30 (Gill and Yermanos 1967a)
L. flavum L.	IPK	LIN 1603	Hungary	*Syllinum*	28 (Chennaveeraiah and Joshi 1983), 30 (Ray 1944)
L. flavum L.	IPK	LIN 99	Unknown	*Syllinum*	28 (Chennaveeraiah and Joshi 1983), 30 (Ray 1944)
L. flavum L.	IPK	LIN 97	Unknown	*Syllinum*	28 (Chennaveeraiah and Joshi 1983), 30 (Ray 1944)
L. flavum L.	IPK	LIN 98	Unknown	*Syllinum*	28 (Chennaveeraiah and Joshi 1983), 30 (Ray 1944)
L. grandiflorum Desf. cv. Rubrum	Shop		Unknown	*Linum*	16 (Ray 1944)
L. grandiflorum Desf.	CGN	CGN21183	Unknown	*Linum*	16 (Ray 1944)
L. grandiflorum Desf.	IPK	LIN 10	Unknown	*Linum*	16 (Ray 1944)
L. grandiflorum Desf.	IPK	LIN 4	Unknown	*Linum*	16 (Ray 1944)
L. grandiflorum Desf.	IPK	LIN 973	Unknown	*Linum*	16 (Ray 1944)
L. grandiflorum Desf.	IPK	LIN 974	Unknown	*Linum*	16 (Ray 1944)
L. hirsutum L. *subsp. hirsutum*	IPK	LIN 1649	Romania	*Dasylinum*	16 (Ray 1944)
L. humile Mill.	CGN	CGN21184	Unknown	*Linum*	30 (Plessers 1966)
L. komarovii Juss.	IPK	LIN 1716	Unknown	*Linum*	n.a.
L. leonii F. W. Schultz	IPK	LIN 1672	Germany	*Linum*	18 (Ockendon 1968)
L. lewisii Purch	IPK	LIN 1648	Unknown	*Linum*	18 (Ray 1944)
L. lewisii Purch	IPK	LIN 1550	Unknown	*Linum*	18 (Ray 1944)
L. macrorhizum Juz.	IPK	LIN 1876	Kirgistan	*Linum*	n.a.
L. mesostylum Juz.	IPK	LIN 1662	Unknown	*Linum*	18 (Ockendon 1968)
L. narbonense L.	IPK	LIN 1653	France	*Linum*	18 (Gill and Yermanos 1967b), 28 (Ray 1944)
L. narbonense L.	IPK	LIN 2002	Unknown	*Linum*	18 (Gill and Yermanos 1967b), 28 (Ray 1944)
L. nodiflorum L.	IPK	LIN 1877	Israel	*Syllinum*	26 (Chennaveeraiah and Joshi 1983)
L. pallescens Bunge	IPK	LIN 1645	Russia	*Linum*	18 (Ockendon 1968)

Table 1 Continued

Accession name	Source	Genebank number	Origin	Section	Chromosome number (2n)
L. perenne L. cv. Blue Vivace	Shop		Unknown	*Linum*	18 (Ray 1944)
L. perenne L.	IPK	LIN 1807	Russia	*Linum*	18 (Ray 1944)
L. perenne L. *subsp. Extraaxillare* (Kit.) Nym.	IPK	LIN 1651	Poland	*Linum*	18 (Ockendon 1968)
L. perenne L. *subsp. Extraaxillare* (Kit.) Nym.	IPK	LIN 1773	Slovakia	*Linum*	18 (Ockendon 1968)
L. perenne L. *subsp. Anglicum* (Mill.) Ockendon	IPK	LIN 1551	Unknown	*Linum*	36 (Ockendon 1968)
L. suffruticosum L. *subsp. saisoloides* (Lam.) Rouy	IPK	LIN 1552	France	*Linastrum*	36 (Nicholls 1986)
L. tauricum Wild.	IPK	LIN 1658	Unknown	*Syllinum*	n.a.
L. tauricum Wild.	IPK	LIN 1604	Unknown	*Syllinum*	n.a.
L. tenuifolium L.	IPK	LIN 1759	Unknown	*Linastrum*	18 (Ray 1944)
L. tenuifolium L.	IPK	LIN 1528	Unknown	*Linastrum*	18 (Ray 1944)
L. tenuifolium L.	IPK	LIN 1650	France	*Linastrum*	18 (Ray 1944)
L. tenuifolium L.	IPK	LIN 1657	France	*Linastrum*	18 (Ray 1944)
L. tenuifolium L.	IPK	LIN 1675	Hungary	*Linastrum*	18 (Ray 1944)
L. thracicum Degen	IPK	LIN 1553	Unknown	*Syllinum*	28 (Plessers 1966)
L. thracicum Degen	IPK	LIN 1764	Slovenia	*Syllinum*	28 (Plessers 1966)
L. trigynum L.	IPK	LIN 1554	France	*Linastrum*	20 (Rogers *et al.* 1972)
L. usitatissimum L. cv. Hermes	Cebeco		France	*Linum*	30 (Ray 1944)

n.a.= not available

Data analysis

In the experiment the AFLP gel images were analyzed with the software package AFLP-Quantar™ Pro. All AFLP bands were treated as dominant markers and scored as either present (1) or absent (0). Clearly distinguishable bands ranging from 50 bp to 500 bp were used in the data matrix and phylogenetic diversity analysis. The phenetic tree was constructed using TREECON software (Van der Peer and Wachter, 1994). Genetic distances were calculated according to (Nei and Li 1979). Cluster analysis was performed using the Neighbour Joining (NJ, Saitou and Nei 1987) method and bootstrap values were based on 1000 permutations and presented as percentages if higher than 70%.

Based on AFLP fingerprint similarity ten groups were identified. Within group A, D and E, multi-accession species (*L. bienne, L. perenne and L. flavum*) were present. To detect the differentiation of these multi-accession species in these groups, the groups were divided in two populations (Table 2). One population (A) contained the species with the most accessions of a group and the other population (B) consisted of the remaining accessions of that group. As in literature *L. angustifolium* often is described as a synonym to *L. bienne*, *L. angustifolium* is analyzed as part of population A of group A. Next, the Analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) was used to partition the total genetic variation to among- and within-population components of group A, D and E and to calculate the genetic variation (sum of square values) in each population using the AMOVA component in the ARLEQUIN 2.000 software (Schneider *et al.*, 2000).

Table 2 Assignment of accessions to the populations

Group	Population A	Population B
A	*L. angustifolium* Huds. (CGN21962) *L. bienne* Mill. (CGN21296) *L. bienne* Mill. (CGN21294) *L. bienne* Mill. (CGN21961)	*L. grandiflorum* Desf. (CGN21183) *L. usitatissimum* L. cv. Hermes *L. humile* Mill. (CGN21184)
D	*L. perenne* L. cv. Blue Vivace *L. perenne* L. (LIN 1807) *L. perenne* L. *subp. Extraaxillare* (Kit.) Nym. (LIN 1651) *L. perenne* L. *subp. Extraaxillare* (Kit.) Nym. (LIN 1773) *L. perenne* L. *subsp. Anglicum* (Mill.) Ockendon (LIN 1551)	*L. alpinum* Jac. (LIN 1905) *L. altaicum* Ledeb. (LIN 1638) *L. austriacum* L. (LIN 1608) *L. austriacum* L. (LIN 1831) *L. austriacum* L. (LIN 1874) *L. komarovii* Juss. (LIN 1716) *L. leonii* F. W. Schultz (LIN 1672) *L. lewisii* Purch (LIN 1648) *L. lewisii* Purch (LIN 1550) *L. macrorhizum* Juz. (LIN 1876) *L. mesostylum* Juz. (LIN 1662) *L. pallescens* Bunge (LIN 1645)
E	*L. flavum* L. (LIN 1603) *L. flavum* L. (LIN 99) *L. flavum* L. (LIN 97) *L. flavum* L. (LIN 98)	*L. campanulatum* L. (LIN 1760) *L. capitatum* Kit. ex Schulte (LIN 1903) *L. tauricum* Wild. (LIN 1658) *L. tauricum* Wild. (LIN 1604) *L. thracicum* Degen (LIN 1553) *L. thracicum* Degen (LIN 1764)

Results

Analysis of the AFLP fingerprints

Initially AFLP analysis was performed on all accessions randomly loaded on the gel to discover groups of accessions, which have a similar pattern. The next step was to group accessions with a similar pattern in adjacent lanes on the gel (see Figure 1). This resulted in ten groups with a high visual fingerprint similarity within the group and clear dissimilarities between groups. The rearrangement of similar accessions in adjacent lanes enabled data collection with higher accuracy. Group specific markers could be recognized easily as constant bands within groups and allowed smiling correction.

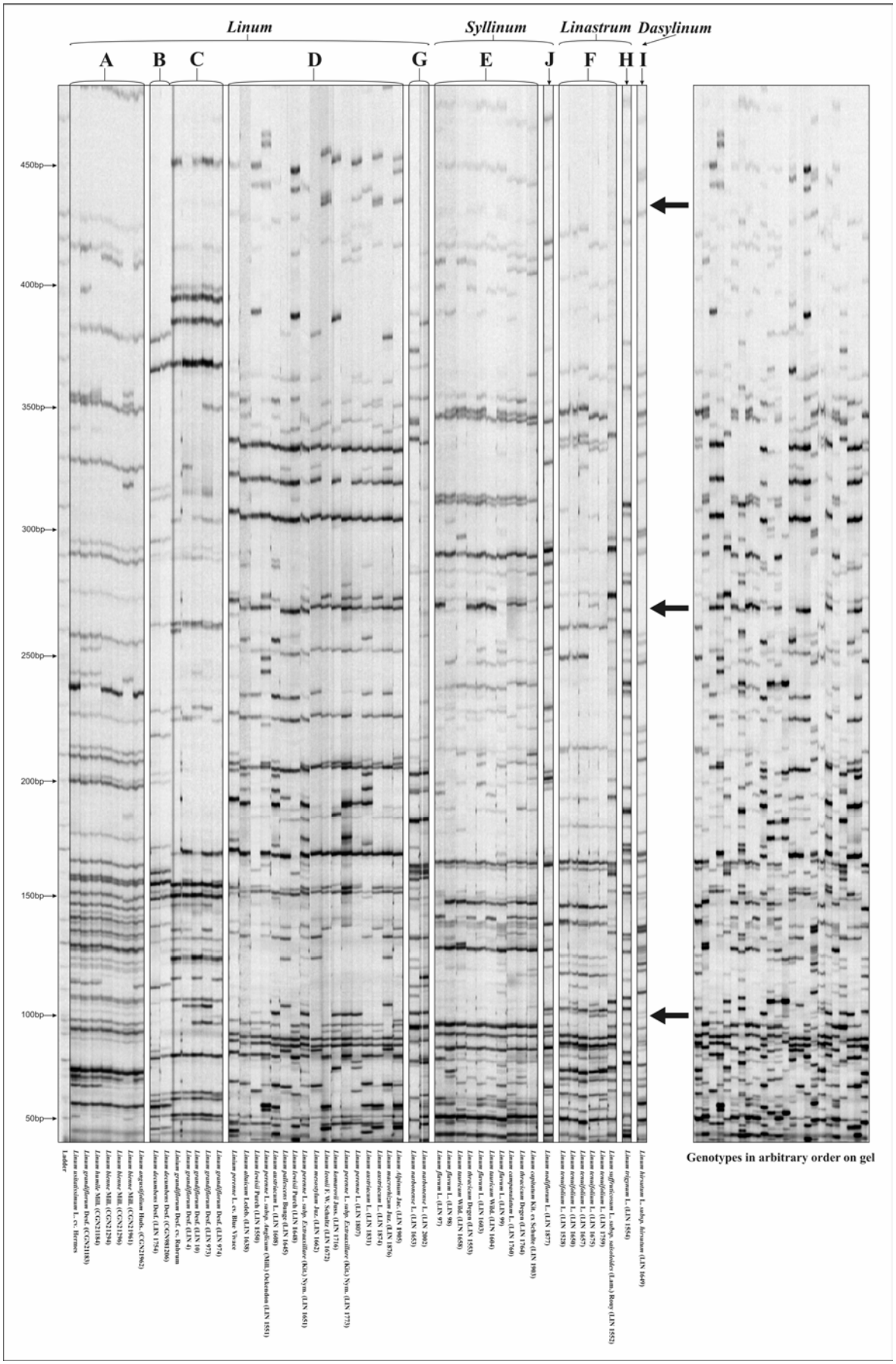

Figure 1 AFLP fingerprints of accessions arbitrarily ordered on gel and specifically ordered on gel based on AFLP pattern similarity

Genetic diversity analysis

A phenetic analysis using Nei and Li's distance estimation and the Neighbour Joining clustering method resulted in a dendrogram shown in Figure 2. Clear groups can be observed which are all supported by a high bootstrap value. The observed groups corresponded very well with the grouping based on the clearly dissimilar AFLP patterns.

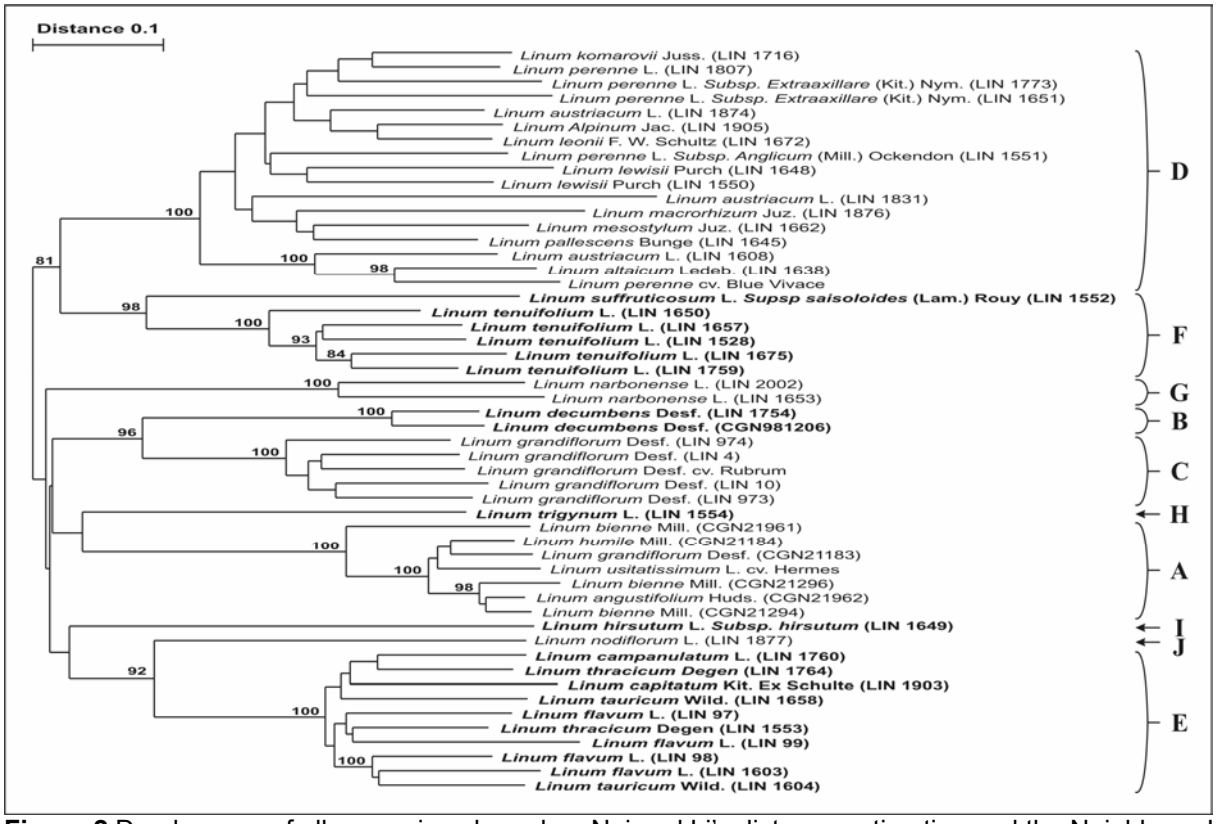

Figure 2 Dendrogram of all accessions based on Nei and Li's distance estimation and the Neighbour Joining clustering method. Only bootstrap values (%) higher than 70 are presented. Bold names visualize the different groups.

The Section *Dasylinum* was represented by *L. hirsutum* only. The separate grouping of *L. hirsutum* justifies its classification in a separate section. Group A, B, C, D and G represent species that have been classified as belonging to section *Linum* but they did not cluster as one monophyletic group. Furthermore, these groups do not show any resemblance in AFLP fingerprint or in position in dendrogram. Group D better known as the *L. perenne* group consisted of ten different species, where *L. perenne*, *L. austriacum* and *L. lewisii* are represented more than once. In group D the accessions share well resembling fingerprint patterns, but no clear differentiation between species could be observed. Only both accessions of species *L. lewisii* grouped together, although not supported by a high bootstrap value. Partitioning the variation within and between populations using an AMOVA showed that the all genetic variability in group D (102.7%) could be assigned to within population variation (Table). Negative variance components

can sometimes occur (Table 3), because they are rather covariances, and their associated F-statistics can be considered as correlation coefficients, which can take negative values. Usually, slightly negative variance components can occur in absence of genetic structure, because the true value of the parameter to be estimated is zero. Thus, if the expectation of the estimator is zero, one may obtain, by chance, slightly positive or slightly negative variance components (Excoffier, pers. comm.). These results confirmed the observation in the dendrogram, that there is no clear differentiation between species. In group A accession *L. grandiflorum* from the Dutch gene bank clearly resembled the fingerprint pattern of group A. Apparently, this accession has been wrongly identified, as this accession is not clustered with other accessions of *L. grandiflorum* in group C. This was confirmed by the Dutch gene bank that reclassified accession CGN221183 from *L. grandiflorum* into *L. usitatissimum* (personal communication).Within cluster A no clear differentiations of species are observed. Furthermore, *L. bienne* (CGN21961) seemed to be the most divergent type in this cluster. The other *L. bienne* accessions together with *L. angustifolium* formed a small cluster within group A. AMOVA showed that part of the genetic variability (14.6%) in group A could be assigned to among population variation (Table 3) which implies a substantial difference between the populations of this group. Groups B and C had slightly similar AFLP (Figure 1) patterns and they grouped together with a significant bootstrap value (Figure 2). Contrary, the species from group B and C are significantly differentiated as two different groups. The two *L. narbonense* accessions clustered significantly as a single group G.

Table 3 Analysis of molecular variance (AMOVA) of group A and D of section *Linum*

Group	Source variation	d.f.	Sum of squares	Variance components	% variation
	Among populations	1	30.3	3.3	14.6
	Within populations	5	95.4	19.1	85.4
A	*Population A*	3	62.8	20.9	
	Population B	2	32.7	16.3	
	Total	6	125.7	22.4	
	Among populations	1	28.3	-0.9	-2.7
	Within populations	15	521.5	34.9	102.7
D	*Population A*	4	144.8	36.2	
	Population B	11	376.7	34.2	
	Total	16	549.8	33.8	

Group H and F comprise species that have been classified as belonging to section *Linastrum,* but did not cluster as one monophyletic group, which was observed in section *Linum* as well. In section *Linastrum*, *L. trigynum* (H) did not cluster with group F. The AFLP patterns of both groups clearly differed. Within group F both *L. tenuifolium* accessions and *L. suffriticosum* significantly clustered (Figure 2). However, *L. suffriticosum* had a slightly

different AFLP (Figure 1) pattern and therefore also clustered apart from the *L. tenuifolium* accessions with a highly supported bootstrap value.

All the species of section *Syllinum* clustered together. Although *L. nodiflorum* branched off from group E, and is indicated as a separate group J. In group E the same occurred as in the *L. perenne* group, as several accessions of single species (*L. flavum*, *L. tauricum* and *L. thracicum*) are randomly distributed in the cluster. This means that also in this group no clear differentiation could be observed between the species. This is confirmed by AMOVA, which showed that almost all the genetic variability (95.8%) in group group E could be assigned to within population variation (Table 4). The only clear differentiation in the section *Syllinum* was the separation of *L. nodiflorum* from the other accessions in this section.

Table 4 Analysis of molecular variance (AMOVA) of group E of section *Syllinum*

Group	Source variation	d.f.	Sum of squares	Variance components	% variation
E	Among populations	1	31.2	1.1	4.2
	Within populations	8	205.8	25.7	95.8
	Population A	3	76.5	25.5	
	Population B	5	129.3	25.9	
	Total	9	237.0	26.9	

The variance components shown in Table 3 and Table 4 indicated that group D was most heterogeneous (34.9) followed by group E (25.7) and A (19.1). Remarkably, the group with species with the smallest amount of variation mainly relies on self-fertilization, whereas species from group D and E are in general self-incompatible outbreeders (Murray 1986). The smallest variance component was observed for the species in population B of group A, all belonging to *L. usitatissimum,* as accession CGN221183 was reclassified into *L. usitatissimum* and *L. humile* is synonymous to *L. usitatissimum*. The cultivated germplasm indeed represents an extremely small (16.3) amount of genetic diversity (Table 3).

Numerical overview of AFLP fingerprints

The analysis of the AFLP fingerprints and the cluster analysis justified the classification of the species into ten groups. In total 374 bands were scored of which 214 originated from the pGC/mCTT primer combination and 160 from the pCA/mCTC primer combination. Except for one marker, all bands were polymorphic (Table 5). No relation was detected between the number of chromosomes (genome size) and the average number of bands. Groups differed considerably in their heterogeneity as shown by the percentage of polymorphic bands within a group. Unfortunately this percentage is strongly confounded with the number of accession in a group. The number of monomorphic bands within a group also indicates genetic heterogeneity.

Table 5 Numerical description of AFLP fingerprints

Group	# of accessions	# of chromosomes	Total # of bands	# of monomorphic bands within group	polymorphic bands within group	% polymorphic Bands within group	# of group specific bands	# of group specific monomorphic bands	Avr. # of bands per genotype
A	7	30	165	65	100	60.6	6	3	116
B	2	30	90	68	22	24.4	1	1	79
C	5	16	171	55	116	67.8	1	1	112
D	17	18, 36	284	5	279	98.2	12	0	83
E	10	28	202	4	198	98.0	0	0	87
F	6	18, 36	198	29	169	85.4	8	1	103
G	2	18	99	53	46	46.5	4	0	76
H	1	20	95	n.a.*	n.a.	n.a.	1	n.a.	95
I	1	16	64	n.a.	n.a.	n.a.	1	n.a.	64
J	1	26	88	n.a.	n.a.	n.a.	2	n.a.	88

* not applicable

From the data shown in Table 5 it is obvious that the groups with a high percentage of polymorphic bands (single accession groups H, I and J excluded), contained the highest number of total bands (group D, E and F), but not necessarily harboring genotypes with high number of AFLP bands (Table 5). The differences in average number of bands per genotypes may reflect the different genome sizes of the species in the various groups. Except for Group E all groups contained at least one group-specific marker. These group-specific markers are in general polymorphic but in some cases they are conserved across that specific group (Table 5).

Discussion

Section Linum

Remarkably, not all species of the section *Linum* (see Table 1) were grouped together. Apparently group A, B, C, D and G (all section *Linum*) did not show much similarity when AFLPs are used. In this study 17 species from section *Linum* were used and these clustered in the five groups. The results of the cluster analysis using AFLP markers gives rise to consider the division of taxa A, B, C, D and G of section *Linum* into subsections or species groups. In a similar way section *Linastrum* (*Linopsis*) has already been divided into five subsections (Rogers 1982), using a number of characteristics.

The *L. perenne* group is a group of species, which is clearly distinguishable from all other *Linum* species. Classification within the *L. perenne* group was suggested to become more easily after an extensive study on the taxonomy of the *L. perenne* group showing that new type of characters provide better tools for the classification of the different species (Ockendon 1968). However, the results of the present study show that the classification within the *L. perenne* group is still complicated as neither the *L. perenne* or the *L. austriacum* species form a specific group (Figure 2). This is underlined by the AMOVA of the *L. perenne* group where variation within species was as large as variation

between species (Table 4). Thus, the classification of the *L. perenne* group needs careful considerations.

Our study shows that although the *L. grandiflorum* and *L. decumbens* species form two separate groups the *L. grandiflorum* and *L. decumbens* species might be considered as one taxon, as all accessions of these two species cluster together, supported by a high bootstrap value (Figure 2). It is known that these two species are closely related (Miriam Repplinger, pers. comm.) and differ only in chromosome numbers, 2n=16 for *L. grandiflorum* (Ray 1944) and 2n=32 for *L. decumbens* (Gill and Yermanos 1967b). Although the grouping of *L. grandiflorum* and *L. decumbens* was also observed by Fu *et al.* (2002); our grouping was more obvious because of the significant bootstrap value.

Besides cultivated flax (*L. usitatissimum*) group A contains several other species, which are related to cultivated flax. Genotype *L. grandiflorum* Desf. (CGN21183) is probably wrongly identified and belongs to *L. usitatissimum* which was confirmed by a phenotypic analysis (data not shown). *L. humile* Miller is considered synonymous with *L. usitatissimum* (Ockendon and Walters 1968) and hence, not surprisingly, *L. humile* groups with *L. usitatissimum*. The taxonomy of species closely related to *L. usitatissimum* is rather complex and ambiguous. *L. angustifolium* has been proposed as the wild progenitor of *L. usitatissimum* (Diederichsen and Hammer 1995; Helbaek 1959; Zohary 1999) and hybrids with cultivated flax are easily obtained (Gill and Yermanos 1967b; Seetharam 1972). About *L. bienne*, bearing the same chromosome number as *L. angustifolium* and cultivated flax, some controversy exists as to its taxonomic position (Muravenko *et al.* 2003). Our study suggests that *L. bienne* is not a monophyletic group and more closely related to *L. angustifolium* than to *L. usitatissimum* (Figure 2). Our results are in contradiction with two studies using RAPD markers, which showed *L. bienne* to be more close to *L. usitatissimum* than *L. angustifolium* to *L. usitatissimum* (Muravenko *et al.* 2003; Lemesh *et al.* 2001). A study with more accessions of *L. bienne*, *L. angustifolium* and *L. usitatissimum* might reveal the true classification of cultivated flax and its progenitors. The observed small variance component is in accordance with the variance partitioning reported in a previous genetic diversity study on fiber flax and linseed varieties (Everaert *et al.* 2001).

Section Linastrum

Three species of the section *Linastrum* were present in this study. Similar to section *Linum,* not all species clustered in one group. However, this is not surprising as it has even been proposed to divide the section into five subsections, where *L. suffriticosum* and *L. tenuifolium* were put in the subsection *Dichrolinum* and *L. trigynum* into subsection *Hallinum* (Rogers 1982). The analysis of the fatty acid composition showed that *L. trigynum* has a different fatty acid composition as compared to *L. suffriticosum* and *L. tenuifolium* (Velasco and Goffman 2000). Next, the chromosome number (Table 1) of *L.*

trigynum seems to have no relation to the chromosome numbers of *L. suffriticosum* and *L. tenuifolium* (Rogers *et al.* 1972).

Based on morphological characters *L. suffriticosum* was considered to be a subspecies of *L. tenuifolium* (Nicholls 1985) and measurements on the fatty acid composition of the seed oil (Velasco and Goffman 2000) showed no clear differences. This supports the clustering of *L. suffriticosum* and *L. tenuifolium*. On the other hand, the AFLP pattern of *L. suffriticosum* clearly differs from that of *L. tenuifolium* and an analysis of the plastochromanol-8 level of seed oil of the two species also showed clear differences (Velasco and Goffman 2000).

Section Syllinum

Several species of the section *Syllinum* were included in this study, which clustered in two different groups. One contained the single specie *L. nodiflorum* that had a different AFLP pattern compared to the other species in this section. Furthermore, *L. nodiflorum* has a reported chromosome number (2n=26) (Chennaveeraiah and Joshi 1983) different from the other species of Group E (2n=28). However, some contradictions exist about the chromosome number of *L. flavum* which appeared to possess either 2n= 28 (Chennaveeraiah and Joshi 1983) or 2n=30 (Ray 1944) and of *L. capitatum* 2n=28 (Ray 1944) or 2n=34 (Chennaveeraiah and Joshi 1983). The separation of *L. nodiflorum* from the other species in this section was also supported by the differences in alpha-tocopherol and linolenic acid content of the seeds (Velasco and Goffman 2000).

The species in the large cluster (Group E) mainly belong to the so called *L. flavum* group except for *L. capitatum* (Tutin *et al.* 1968). However, our study shows that *L. capitatum* cannot be differentiated from the *L. flavum* group Figure 1 and Figure 2) and should be considered as part of this group. Furthermore it is questionable whether *L. thacicum* and *L. tauricum* are different from *L. flavum*, as these species integrate with *L. flavum* accessions (Figure 2) and partitioning the variation shows no clear differentiation within this group (Table 4).

Use of this study for flax breeding

It has become increasingly difficult to distinguish new fiber flax varieties from the existing cultivars for granting plant breeders right (Everaert *et al.* 2001). A previous study showed that especially the molecular genetic diversity of fiber flax has decreased rapidly over the last half century (Vromans *et al.* Chapter 2). If breeding for new fiber flax and linseed cultivars becomes more difficult because of the lack of favorable alleles in the current flax cultivars, the introduction of wild germplasm could be a good solution. The introduction of wild germplasm, by means of backcross inbred lines, has proven to be successful in different crops (e.g. Eshed and Zamir 1994; Jeuken and Lindhout 2004). The best possible

wild species to use in flax are primary the most closely related species still crossable with *L. usitatissimum*. Hence, *L. bienne* and *L. angustifolium* are excellent candidates to introduce favorable alleles from, although added value for the cultivated flax, taking shape as new resistances or other traits might not be visible immediately. Therefore, a more detailed study of several accessions of *L. bienne* and *L. angustifolium* for new traits or new alleles is desirable. Other wild species are less interesting for new trait development, as they are not crossable with *L. usitatissimum* (Seetharam 1972) and have a completely different morphology.

Application of fingerprinting in taxonomic studies

AFLPs proved to be useful for determining the phenetic relationship between cultivars of flax (Everaert *et al.* 2001) but whether AFLPs are useful in phylogenetic studies of flax is discussed in this report as well. AFLP bands are classified as identical on the basis of mobility. The total number of different bands analyzed in this study is probably higher as non-homologous but co-migrating markers are treated as identical among accessions. Although this phenomenon is a common source of error in AFLP analysis when used with more distinct species, AFLPs proved to be suitable to study phylogenetic relationships (Kardolus *et al.* 1998; Koopman *et al.* 2001). Recently, Koopman (2005) reviewed tree topologies generated with AFLP and ITS. In general congruence in tree topology could be observed for a wide range of taxonomic groups. Therefore, the topology described here should not widely differ from a topology based on ITS sequences.

This study shows that AFLP markers can be used to group species and to differentiate between groups within the genus *Linum*. However, this could not be confirmed for the five sections into which the genus *Linum* is subdivided. The phenetic analysis did not result in clusters supported by a high bootstrap value for the sections. We did not aim for an entire reconstruction of the phylogeny of *Linum* species, but we intended to study the amount of genetic variation of cultivated germplasm in relation to crossable and non-crossable species to improve breeding programs. On the basis of these results we nevertheless conclude that biosystematicists should aim for a revision of the genus *Linum*. We suggest that species within group A, E, and F can be lumped into fewer species, because of the highly similar fingerprints.

Between groups the AFLP fingerprints were highly dissimilar. The number of identical bands between groups was very low, and therefore co-incidental co-migration of non-homologous bands starts to prevail over co-migration of truly homologous bands. Therefore, the higher structure of the dendrogram may not reflect the phylogenetic relation between groups. Other molecular data, like nuclear (ITS) and chloroplast sequences, have proven to be of great value for the determination of higher hierarchical relationships (Doi *et*

al. 2002; Garcia-Mas *et al.* 2004; Xu and Ban 2004; Zhang *et al.* 2002) and might be the solution for the detection of the different sections in the genus *Linum*.

Conclusion

AFLPs have proven to be useful to study phylogenetic relationships in the genus *Linum* to a certain taxonomic level. For dividing the genus *Linum* into sections, other molecular tools like ITS sequences should be used. AFLPs showed to be very useful to unravel the complex taxonomy of closely related taxa like the species in the *perenne* and *flavum* group. If in the future *L. bienne* and *L. angustifolium* will be use for extending the flax gene pool, backcross strategies can easily be applied as the percentage of polymorphic markers between the wild species and cultivated flax is three to four times higher than between two cultivars.

Acknowledgements

We are very grateful to Miriam Repplinger for providing useful commments on the phenetic tree. Seeds were kindly provided by the Dutch Centre of Genetic Resources, CGN Wageningen (the Netherlands), IPK genebank in Gatersleben (Germany) and Cebeco Seeds. We gratefully acknowledge the Dutch breeding campanies Cebeco Seeds, ‘Van de Bilt Zaden en Vlas’ and Dobbelaar Breeding and the Dutch provinces Zeeland for their financial contribution.

References

Chennaveeraiah, M.S. and Joshi, K.K. (1983) Karyotypes in cultivated and wild species of *Linum*. *Cytologia* 48: 833-841

Diederichsen, A. and Hammer, K. (1995) Variation of cultivated flax (*Linum usitatissimum* L. subsp. *usitatissimum*) and its wild progenitor pale flax (subsp. *angustifolium* (Huds.) Thell.). *Genetic Resources and Crop Evolution* 42: 263-272

Doi, K., Kaga, A., Tomooka, N. and Vaughan, D.A. (2002) Molecular phylogeny of genus *Vigna* subgenus *Ceratotropis* based on rDNA ITS and atpB-rbcL intergenic spacer of cpDNA sequences. *Genetica* 114: 129-145

Eshed, Y. and Zamir, D. (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. *Euphytica* 79: 175-179

Everaert, I., de Riek, J., de Loose, M., van Waes J., and van Bockstaele, E. (2001) Most similar variety grouping for distinctness evaluation of flax and linseed (*Linum usitatissimum* L.) varieties by means of AFLP and morphological data. *Plant Varieties and Seeds* 14: 69-87

Excoffier, L., Smouse, P.E. and Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491

Fu, Y.B., Diederichsen, A., Richards, K.W. and Peterson, G. (2002) Genetic diversity within a range of cultivars and landraces of flax (*Linum usitatissimum* L.) as revealed by RAPDs. *Genetic Resources and Crop Evolution* 49: 167-174

Garcia-Mas, J., Monforte, A.J. and Arus, P. (2004) Phylogenetic relationships among *Cucumis* species based on the ribosomal internal transcribed spacer sequence and microsatellite markers. *Plant Systematics and Evolution* 248: 191-203

Gill, K.S. and Yermanos, D.M. (1967a) Cytogenetic studies on genus *Linum*.I. Hybrids among taxa With 15 as haploid chromosome number. *Crop Science* 7: 623-627

Gill, K.S. and Yermanos, D.M. (1967b) Cytogenetic studies on genus *Linum*.2. Hybrids among taxa With 9 as haploid chromosome number. *Crop Science* 7: 627-631

Helbaek, H. (1959) Domestication of food plants in the old world. *Science* 130: 365-372

Hickey, M. and King, C. (1988) 100 families of flowering plants. 2nd edition. University Press, Cambridge

Huang, J., Corke, H. and Sun, M. (2002) Highly polymorphic AFLP markers as a complementary tool to ITS sequences in assessing genetic diversity and phylogenetic relationships of sweetpotato (*Ipomoea batatas* (L.) Lam.) and its wild relatives. *Genetic Resources and Crop Evolution* 49: 541-550

Jeuken, M.J.W. and Lindhout, P. (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. *Theoretical and Applied Genetics* 2004; 109: 394-401

Kardolus, J.P., van Eck, H.J. and van de Berg, R.G. (1998) The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (*Solanaceae*). *Plant Systematics and Evolution* 210: 87-103

Koopman, W.J.M., Zevenbergen, M.J. and van den Berg, R.G. (2001) Species relationships in Lactuca s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. *American Journal of Botany* 88: 1881-1887

Koopman, W.J.M. (2005) Phylogenetic signal in AFLP data sets. *Systematic Biology* 54: 197-217

Lemesh, V.A., Malyshev, S.V., Grushetskaya, Z.E. and Khotyleva, L.V. (2001) DNA polymorphism and phylogeny of flax. *Proceedings of the Second Global Workshop; Bast Plants in the New Millennium*: p 139-145

Muravenko, O.V., Lemesh, V.A., Samatadze, T.E., Amosova, A.V., Grushetskaya, Z.E., Popov, K.V., Semenova, O.Y., Khotyuleva, L.V. and Zelenin, A.V. (2003) Genome comparisons with chromosomal and

molecular markers for three closely related flax species and their hybrids. *Russian Journal of Genetics* 39: 414-421

Murray, B.G. (1986) Floral biology and self-incompatibility in *Linum*. *Botanical Gazette* 147: 327-333

Nei, M. and Li W.H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* 76: 5269-5273

Nicholls, M.S. (1985) A systematic study of the *Linum tenuifolium* group (*Linaceae*). *Botanical Journal of the Linnean Society* 91: 473-490

Nicholls, M.S. (1986) Variation and evolution in *Linum tenuifolium* (*Linaceae*). *Plant Systematics and Evolution* 153: 243-258

Ockendon, D.J. (1968) Biosystematic studies in *Linum perenne* group. *New Phytologist* 67: 787-813

Ockendon, D.J. and Walters, S.M. (1968) *Linum*. *Flora Europaea* 2: 206-211

Plessers, A.G. (1966) Variation in fatty acid composition of seed of *Linum s*pecies. *Canadian Journal of Genetics and Cytology* 8: 328-335

Ray, C. Jr. (1944) Cytological studies on the flax genus, *Linum*. *American Journal of Botany* 31: 241-248

Rogers, C.M., Harris, B.D. and Mildner, R. (1972) Some additional chromosome numbers in *Linaceae*. *Brittonia* 24: 313-316

Rogers, C.M. (1972) Taxonomic significance of fatty acid content of seeds of *Linum*. *Brittonia* 24: 415-419

Rogers, C.M. (1982) The systematics of *Linum s*ect. *Linopsis* (*Linaceae*). *Plant Systematics and Evolution* 140: 225-234

Saitou, N. and Nei, M. (1987) The neighbor-joining method a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution 4*: 406-425.

Seetharam, A. (1972) Interspecific hybridization in *Linum*. *Euphytica* 21: 489-495

Schneider, S., Roessli, D. and Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Tammes, T. (1923) Das genotypische Verhältnis zwischen dem wilden *Linum angustifolium* and dem Kulturlein, *Linum usitatissimum*. *Genetica* 5: 61-76

Tutin, T.G., Heywood, V.H., Burges, N.A., Moore, D.M., Valentine, D.H., Walters, S.M. and Webb, D.A. (1968) Flora Europaea Rosaceae to Umbelliferae 2: 206-211

Van de Peer Y. and de Wachter, R. (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Computer Applications in the Bioscience* 10, 569-570

Van der Beek, J.G., Verkerk, R., Zabel, P. and Lindhout, P. (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. *Theoretical and Applied Genetics* 84: 1-2

Velasco, L. and Goffman, F.D. (2000) Tocopherol, plastochromanol and fatty acid patterns in the genus *Linum*. *Plant Systematics and Evolution* 221: 77-88

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414

Xu, D.H. and Ban, T. (2004) Phylogenetic and evolutionary relationships between *Elymus humidus* and other *Elymus* species based on sequencing of non-coding regions of cpDNA and AFLP of nuclear DNA. *Theoretical and Applied Genetics* 108: 1443-1448

Yermanos, D.M., Beard, B.H., Gill, K.S. and Anderson, M.P. (1966) Fatty acid composition of seed oil of wild species of *Linum*. *Agronomy Journal* 58: 30-32

Zhang, L., Comes, H.P. and Kadereit, J.W. (2001) Phylogeny and quaternary history of the European montane/alpine endemic *Soldanella* (*Primulaceae*) based on ITS and AFLP variation. *American Journal of Botany* 88: 2331-2345

Zhang, W., Qu, L.J., Gu, H., Gao, W., Liu, M., Chen, J. and Chen, Z. (2002) Studies on the origin and evolution of tetraploid wheats based on the internal transcribed spacer (ITS) sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics* 104: 1099-1106

Zohary, D. (1999) Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. *Genetic Resources and Crop Evolution* 46: 133-142

Chapter 4

Selection of highly polymorphic AFLP fingerprints in flax (*Linum usitatissimum* L.) reduces the workload in genetic linkage map construction and results in an elevated level of allelic codominant markers.

Jaap Vromans, Gerben Mennink, Piet Stam and Herman J. van Eck

Abstract

Four crossing parents, used for the development of two mapping populations, were preliminary screened for 1024 AFLP primer combinations, to increase the efficiency of AFLP analysis in flax. For the construction of a linkage map containing 300 markers selecting the most polymorphic primer combinations will reduce the workload by more than 50% as compared to randomly chosen primer combination. The efficiency of selection depends on the genetic distance between two accessions and the stringency of the selection. The effect of AT content in the selective nucleotides on fingerprint complexity and level of polymorphism was studied in more detail. The positive correlation found in this study between the number of bands per primer combination and the overall AT content in the selective nucleotides of the primers can be ascribed to methylation and compositional bias of C/G versus A/T residue. The level of polymorphism is affected by the similarity of the parents and not by the AT content of the selective nucleotides. As a side effect of the preliminary screen an unexpected high percentage of valuable allelic codominant markers were detected during screening of the two mapping populations. The identification of two or three additional selective nucleotides at the *Mse*I primer site confirmed the allelism of 14 codominant markers. Sequencing two allelic codominant markers suggests both microsatelite variation as well as INDEL to be the prevalent cause of allelism. This study showed that pre-screening a window of 1024 AFLP primer combinations is profitable for flax.

Introduction

Marker assisted breeding is increasingly being used in breeding programs. To generate molecular markers, which are associated with traits, several approaches can be used, such as Bulk Segregant Analysis for qualitative, monogenic traits (BSA), linkage desequilibrium (LD) mapping and QTL analysis. In all important agricultural crops at least one of these techniques has been applied. The QTL analysis and LD mapping method require high-density marker linkage maps.

Among DNA fingerprinting techniques the AFLP method is known for its low start-up costs, reproducibility and high multiplex ratio (Jones *et al.* 1997; Vos *et al.* 1995). For these reasons AFLP markers have been widely used in mapping studies, including the construction of highly saturated maps of tomato, potato, papaya, pepper and maize (Haanstra *et al.* 1999; Isidore *et al.* 2003; Ma *et al.* 2004; Paran *et al.* 2004; Vuylsteke *et al.* 1999). Usually the selective nucleotides that can vary to AFLP primer combinations can be chosen more or less randomly, without an expensive and time consuming selection of superior primer combinations (Qi and Lindhout 1997). However, species with a very low genetic diversity may require an extensive search for primer combinations that yield a high proportion of polymorphic markers. In *Beta* it has been shown that a two-fold increase in number of segregating markers per fingerprint can be achieved if the upper two percent selected primer combinations are used, as compared to random primer combinations (Hansen *et al.* 1999). A study in barley also showed that a selection of specific AFLP primer combinations is beneficial (Qi and Lindhout 1997). Furthermore these authors showed that the polymorphism rate is independent of the barley lines used. Thus, the most efficient primer combination observed for a given pair of lines turned out to be highly efficient for other sets of lines as well (Qi and Lindhout 1997).

In flax (*Linum usitatissimum L.*), a diploid (2N=30) self-fertilizing annual plant species, two earlier linkage maps have been reported (Spielmeyer *et al.* 1998). One genetic linkage map, based on AFLPs, resulted in 18 linkage groups covering approximately 1400 cM. This map enabled the identification of two quantitative trait loci (QTLs) conferring resistance against Fusarium wilt. One marker associated with one of the QTLs also showed bulk specificity to Fusarium wilt in a BSA approach (Spielmeyer *et al.* 1998). A second map was composed of 94 RFLP and RAPD markers comprising 15 linkage groups and covering 1000 cM (Oh *et al.* 2000).

From previous studies its known that flax is a crop species with very little genetic variation (Fu *et al.* 2002; Vromans *et al.*, Chapter 2). Considering the two cultivar groups, linseed and fiber flax, the variation within fiber flax is very low compared to the variation within linseed and the genetic distance between linseed and fiber flax cultivars (Vromans *et al.*, chapter 2). Therefore, it is expected that the progeny of a cross between a linseed and fiber flax cultivar or two linseed cultivars will harbor more polymorphic markers than

the progeny of a cross between two fiber flax cultivars. But even the progeny of a cross between a linseed cultivar and a linseed breeding line, turned out to contain low numbers of polymorphic bands (Spielmeyer *et al.* 1998). After selecting the 45 most polymorphic primer combinations out of 160 primer combinations, only 5.7 markers per primer combination on average could be detected (Spielmeyer *et al.* 1998). The number of segregating markers may vary substantially, depending on the genetic diversity of the crop species. Typically self-fertilizing crop species are less polymorphic than outbreeders. Averages of more than 15 polymorphic bands per primer combinations are common in crops like reygrass, potato, melon, barley, *Brassica olerace* and maize, even with random, unselected primer combinations. In flax however, the average number of polymorphic bands per primer combination is much smaller (Bert *et al.* 1999; Van Eck *et al.* 1995; Perin *et al.* 2002; Qi *et al.* 1998; Sebastian *et al.* 2000; Vuylsteke *et al.* 1999).

In this study we describe the screening of 1024 AFLP primer combinations, to increase the efficiency of AFLP analysis in flax. This report also describes the effects of the large preliminary screen, with the main focus on labor efficiency and an additional benefit of the preliminary screen.

Material and methods

Plant material

Four crossing parents, cultivars Viking, Belinka and Hermes and breeding line E1747 were used for screening the number of polymorphic bands per AFLP primer combination. Viking and E1747 were used as parents for the development of the VxE recombinant inbred line (RIL) population and Belinka and Hermes for the BxH RIL population. Viking and E1747 were kindly provided by Dr. Steve Knapp of Oregon State University. The genotype E1747, also known as variety Solon, is an EMS-induced McGregor mutant originating from Canada with a low linolenic acid content (Rowland 1991) and Viking is a French fiber flax variety (Coopérative Linière de Fontaine-Cany, France). Hermes is a French fiber flax variety (Coopérative Linière de Fontaine-Cany, France) and Belinka is a Dutch fiber flax variety (Cebeco Seeds B.V., The Netherlands). Leaf material was collected from two-week old seedlings that were grown in the greenhouse. Genomic DNA was extracted from fresh leaves according to the procedure described by Van der Beek *et al.* (1992) with minor modification: the DNA was washed overnight in 76% ethanol and 10mM NH4Ac, dried in a vacuum dryer and dissolved in 100µl mQ.

AFLP analysis

The AFLP procedure was performed according to Vos *et al.* (1995). Template was produced with two enzyme combinations *Eco*RI/*Mse*I and *Pst*I/*Mse*I. In total 1024 *Eco*RI+3/*Mse*I+3 and *Pst*I+2/*Mse*I+3 primer combinations (Table 1) originating from

*Eco*RI+A/*Mse*I+A/C and *Pst*I+0/*Mse*I+A/C secondary template, were tested. AFLP fragments were labelled with *Eco*RI+3 or *Pst*I+2 primers (IRDyeTM 700/ IRDyeTM 800) combined with *Mse*I+3 primers (Table 1), and loaded on a 6% polyacrylamide gel and fingerprint images were captured on a NEN® Global Edition IR2 DNA Analyser (LiCor® Biosciences, Lincoln, NE).

Table 1 Number of primer combinations evaluated, originating from different secondary templates. Values in parenthesis are the number of successful primer combinations.

	*Mse*I+ANN	*Mse*I+CNN
*Eco*RI+ANN	256 (202)	256 (185)
*Pst*I+NN	256 (179)	256 (193)

Excision, re-amplification and sequencing of AFLP fragments

For the identification of the 4th, 5th and 6th position from the *Mse*I adapter sequence the method described by Brugmans *et al.* (2003) was performed. To reduce the complexity of the fingerprint and to facilitate the excision of fragments fluorescently labeled *Eco*RI/*Pst*I primers were used in combination with the extended *Mse*I +5 and +6. To excise the fragments, AFLP fingerprints were first separated on the NEN, Global Edition IR2 DNA Analyser. After separation, the gel was scanned on a LI-COR® Biosciences Odyssey® Infrared Imaging System (LiCor® Biosciences, Lincoln, NE) along with some marking points to allow careful positioning of the AFLP fragments. Gel pieces containing the fragments were excised using a scalpel and successful fragment excision was verified by re-scanning the gel. After excision, gel pieces were put in 50 µl mQ and vortexed several times during an incubation time of approximately one hour, 2 µl supernatant was used for re-amplification in a total volume of 25 µl containing 1µl *Pst*I+2- or *Eco*RI+3 primer, 1µl *Mse*I+6 primer, 1µl dNTP's, 2.5 µl PCR buffer (SuperTaq buffer from Enzyme Technologies), 18.4 µl mQ and 0.1 µl *Taq* polymerase (SuperTaq). The products were re-amplified in 45 cycles, 30 seconds at a denaturation temperature of 94 ^{0}C, 30 seconds at an annealing temperature of 56 ^{0}C and 1 minute at an amplification temperature of 72 ^{0}C. The PCR products were purified and send for sequencing. For each fragment sequence data were obtained using both 3'and 5' primers to be sure of correct sequencing data.

Codominant markers

The results of the allelic codominant markers were obtained from another study by Vromans (Chapter 5) where the most polymorphic primer combinations (selected in this study) were applied to two mapping populations. In some cases he recognized length polymorphisms in the same primer combination, when screening the RILs. He assigned two polymorphic bands as one allelic codominant marker if the two polymorphic bands comply with the two criteria given by Alonso Blanco *et al.* (1998), *i.e*. *(i)* two AFLP bands might be allelic when they derive from different parents, with the same primer combination

and *(ii*) the two AFLP markers segregate complementary (Alonso Blanco *et al.* 1998). Allelic segregating bands of this type were manually scored as a single codominant marker.

Data collection and analysis
All AFLP fragments between 50-500 base pairs were recorded, per primer combination and per genotype. Markers shorter than 50 base pairs were not used in this analysis, as an AFLP fragment consist of two primers with a length of approximately 20 base pairs each, exceeding already the size of 40 base pairs. The number of polymorphisms between the two crossing parents V and E and between the parents B and H was calculated. The genetic distances between two genotypes were calculated as $1-(n_{11}/(n_{01}+n_{10}+n_{11}))$ where n_{11} are common bands and n_{01} as well as n_{10} occur in only one of the two genotypes. The codominant nature of some makers was not taken into consideration, so both alleles were used to measure the genetic distance. The equation $n_{11}/(n_{01}+n_{10}+n_{11})$ is better known as the Jaccard similarity coefficient (Sneath and Sokal 1973).

Results

Selection of highly polymorphic AFLP primer combinations
Out of the 1024 primer combination 265 failed for all four parents for different reasons (Table 1). We assume that failure is due to the quality of at least one of the primers, where in most cases the decay of the fluorescent label is assumed.
To compare between selected and randomly chosen primer combination, we calculated the number of primer combinations required for screening both mapping populations for the construction of a linkage map containing 300 markers. The optimal choice resulted in 29 primer combination for the VxE and 47 primer combinations for the BxH population respectively. Without this selection a much larger number of primer combinations should be employed to obtain 300 markers, *i.e.*79 (average of 3.8 polymorphisms per primer combination) and 130 (average of 2.3 polymorphisms per primer combination) primer combinations for the VxE and BxH population, respectively. Taking the throughput in our lab and a population size of 180 genotypes, selection of the most polymorphic primer combinations reduces the workload by more than 50% (from > 100 workdays to 50 days, approximately). The efficiency of selection depends on the genetic distance between two accessions and the stringency of the selection. The more similar two accessions and the more stringent the selection is, the larger the effect of the selection. As the efficiency increases with a decreasing distance between both parents of a mapping population (Table 2), a preliminary screen of highly polymorphic AFLP primer combinations is especially efficient in genetically homogeneous crops, like flax.

Table 2 Number of polymorphic bands per lane across subsets of superior AFLP primer combinations. The selected primer combinations represent the 2%, 5%, 10%, 20% and 50% most polymorphic primer combinations for both parental combinations. Also the genetic distance (G.D.= 1-(n_{11}/(n_{01}+n_{10}+n_{11}))) is reflected.

		Selection				
Accession comparisons	G.D.	2%	5%	10%	20%	50%
Belinka x Hermes	0.05	3.3	2.9	2.6	2.2	1.6
Viking x E1747	0.08	2.9	2.7	2.4	2.1	1.6

Fingerprint complexity

From previous studies in beet (Hansen *et al.* 1999) and barley (Qi and Lindhout 1997) it is known that an increase of the AT content of the selective nucleotides in AFLP primer combinations results in an increase of both the total number of bands and the number of polymorphic bands per primer combination. This A/T effect is caused by the nucleotide bias of plant species which are generally AT-rich, 60-66 % (Karlin and Mrazek 1997). With our current data set the effect of the AT content of the selective nucleotides in AFLP primer combinations could be investigated. The total number of bands per primer combination of the different classes with the same AT content in the overall selective bases was in general normally distributed. As expected, a positive correlation was found between the number of bands per primer combination and the overall AT content in the selective bases of the primers (Figure 1).

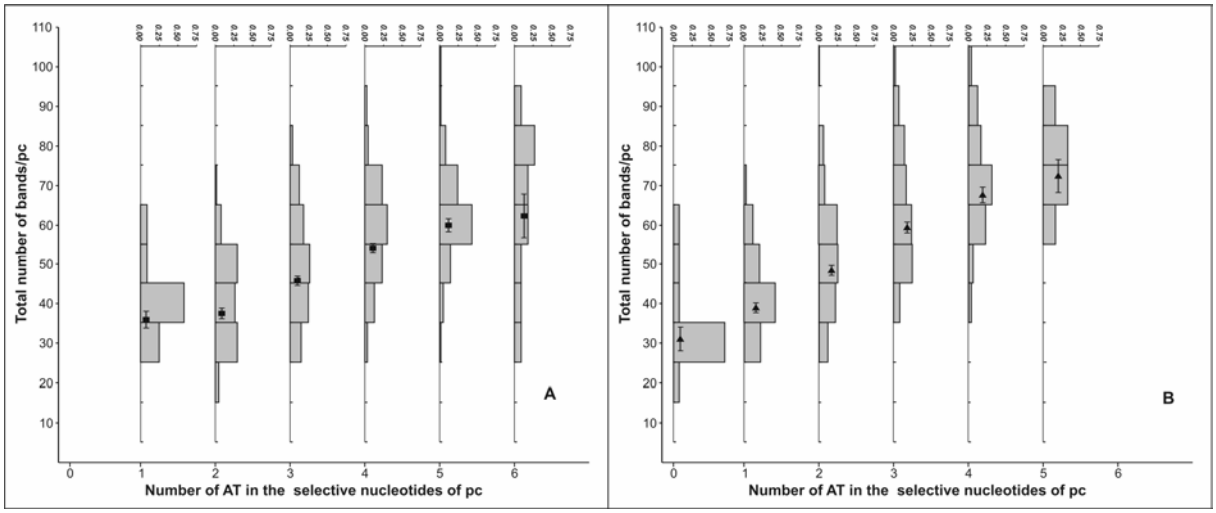

Figure 1 Distributions of the number of bands per primer combination for the different classes of AT content of the selective nucleotides in the primer combinations. Standard errors for all averages are shown with Y Error Bars. **A** *Eco*RI/*Mse*I primer combinations **B** *Pst*I/*Mse*I primer combinations

For both enzyme combinations a similar slope would be expected, describing the increase in additional bands with an increasing number of AT residues in the selective nucleotides. In Figure 1 we observe an approximate linear increase in number of bands as the number of AT residues in the selective nucleotides increases (5.9 and 8.7 additional bands per CG ->AT substitution for *Eco*RI/*Mse*I and *Pst*I/*Mse*I, respectively).

Furthermore the genome size of the organism is reflected by the fingerprint complexity. The number of clear AFLP bands across lanes per primer combination varied from 15 to 100, with an average of 48.5 for the *Eco*RI/*Mse*I primer combinations and 51.3 for the *Pst*I/*Mse*I primer combinations. The complexity of *Eco*RI/*Mse*I fingerprints of flax, which has a relatively small genome size of 1C=686 Mbp (Bennett and Leitch 2004), is in range comparable with plant species with a similar genome size (Table 3).

Table 3 Comparison of average number of bands per primer combination with genome size of crops with similar genome size as flax. Value in parenthesis average number of bands per PstI/*Mse*I primer combination.

Crop	Genome size (1C) in Mbp (Bennett *et al.* 2004)	Average number of bands with +3/+3 *Eco*RI/*Mse*I primer combinations	Reference
Cowpea	588	55	Menendez *et al.* 1997
Melon	931	80	Perin *et al.* 2002
Rice	490	47	Maheswaran *et al.* 1997
Carrot	980	58	Vivek and Simon 1999
Brassca nigra	760	53	Negi *et al.* 2004
Flax	686	49 (51)	This study

Level of polymorphism

A total of 38077 bands across all fingerprints were investigated. The distribution the number of primer combinations that result in a given percentages of polymorphic bands per primer combination of both parental combinations is skewed (Figure 2). It is clear that only a small percentage of bands were polymorphic between parents of a mapping population, 7.7 % for the VxE and 4.8 % for the BxH mapping population respectively.

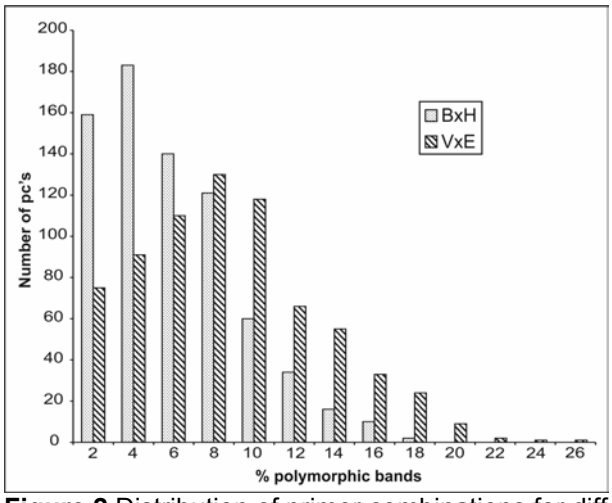

Figure 2 Distribution of primer combinations for different percentages of polymorphic bands per primer combination for the two different parental combinations, Belinka*Hermes (BxH) and Viking*E1747 (VxE)

The percentage of polymorphic bands is not expected to be influenced by the choice of selective nucleotides added to the AFLP primers. Figure 3 shows that this is the case for the BxH population with both enzyme combinations. Although the percentage of

polymorphic bands in the VxE population seems to benefit from an increasing AT content in the selective nucleotides of the *Eco*RI/*Mse*I primer combinations, this increase is not statistically significant.

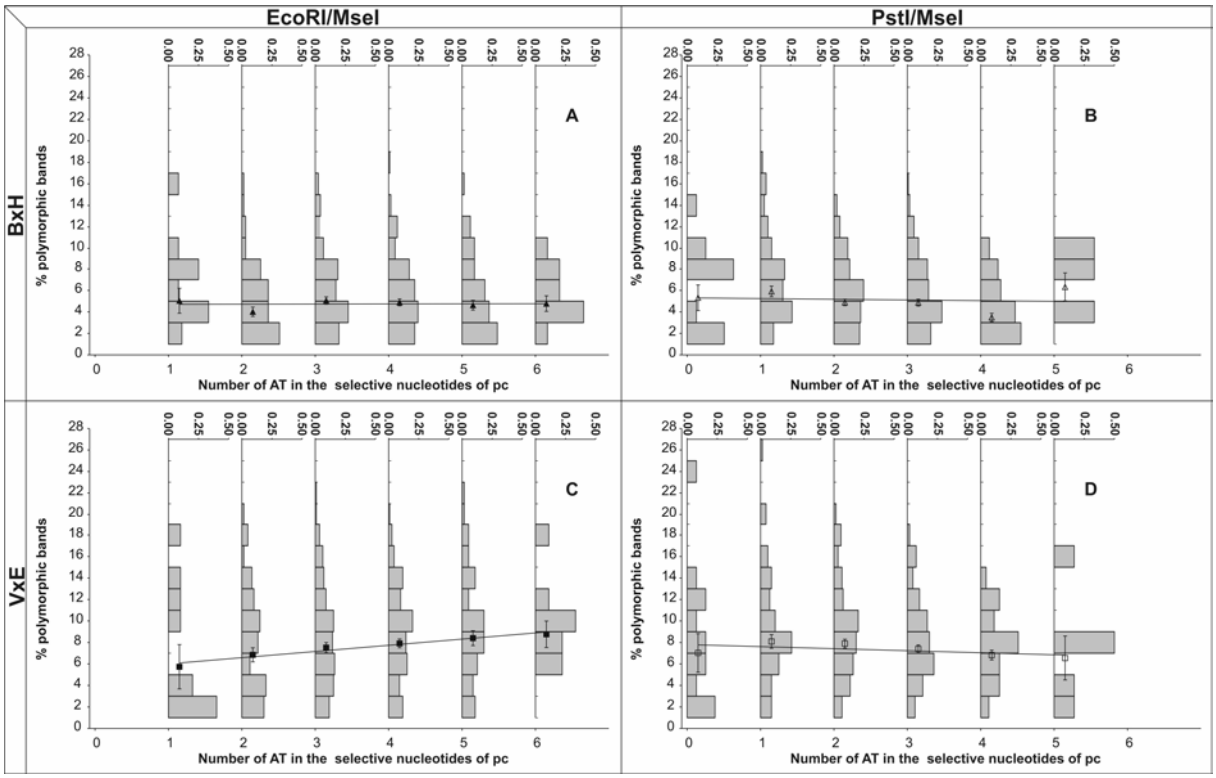

Figure 3 Distribution and correlation of the % polymorphic bands per primer combination for the different AT content classes of the selective nucleotides. Standard errors for all averages are shown with Y Error Bars. **A** Enzyme combination *Eco*RI/*Mse*I and population BxH **B** Enzyme combination *Pst*I/*Mse*I and population BxH **C** Enzyme combination *Eco*RI/*Mse*I and population VxE **D** Enzyme combination *Pst*I/*Mse*I and population VxE

Codominant markers

In addition to the reduction of labor and costs a second advantage was obtained by the selection of highly polymorphic primer combinations. An unexpected high percentage of codominant markers was detected during screening of the two mapping populations. We obtained 58 and 73 codominant markers in the BxH and VxE population, respectively (36% and 38%). AFLP markers are usually considered as dominant “mono allelic” markers, where the alternate allele is not amplified, and thus not observed. Such absence/presence polymorphisms are the most common type of polymorphism in AFLP and based on SNP’s. Codominant AFLPs with both alleles as alternatively segregating bands are usually very infrequent. Such codominant markers are based on insertion/deletion (INDEL) polymorphisms. This can be easily observed when both alleles occur in the same primer combination, usually with a few nucleotides difference.

To determine the molecular basis of the bi-allelic codominant makers, some markers were further analyzed. From the total 131 allelic codominant markers 14 were

analyzed to identify two or three additional selective nucleotides at the *Mse*I primer site using the method described by Brugmans *et al.* (2003). For all 14 codominant markers both allelic bands always had the same internal nucleotides, confirming their allelism. Two codominant markers (pAAmCAT-294/291 and pATmCAA-659/673) were further analyzed by re-amplifying (with new *Mse*I primers) and sequencing both alleles to identify the origin of the length difference. The alleles of both codominant markers showed a high level of sequence homology of more than 99.9 %. Marker pAAmCAT-294/291 contained a small 3 base pair insertion/deletion polymorphic site whereas the alleles of marker pATmCAA-659/673 had a microsatelite variation of 18 base pairs based on the TTC repetition.

Discussion

Selection of highly polymorphic AFLP fingerprints

Molecular markers have shown to be very useful for the construction of genetic linkage maps. With an increasing number of chromosomes, in general larger numbers of polymorphic bands are required for mapping studies. The AFLP method is very suitable to this end, because of its high multiplex ratio (Vos *et al.* 1995). This study showed that selection of the most polymorphic AFLP primer combinations could reduce the workload by more than 50%. The reduction of workload is highly dependent on the genetic distance between the parental genotypes, as shown in Table 2. A similar type of correlation was previously reported by (Hansen *et al.* 1999). In the present study the increase in efficiency resulting from primer combination selection is larger than in Hansen's (1999) case. This is mainly due to the smaller genetic distance between the parents in our study. It is obvious that especially monomorphic crops, linkage mapping benefits from a preliminary screen of primer combinations. Also in more polymorphic crops a large preliminary screen is efficient. Consider the following mapping project: 300 polymorphisms are analyzed to cover a complete linkage map of a specific crop, based on a population size of 180 individuals. Our results indicate that, only when the average number of polymorphic markers per randomly chosen primer combination exceeds 26.5, which is extremely high, a selection procedure is not efficient any more.

Fingerprint complexity

This report shows that the *Eco*RI and *Pst*I fingerprints of flax are of equal complexity (48.5 markers per primer combination for the *Eco*RI/*Mse*I primer combinations and 51.3 markers per primer combination for the *Pst*I/*Mse*I primer combinations) although the *Pst*I primer is extended with only two selective nucleotides. This should result in a four-fold increase of complexity for *Pst*I primer combinations as compared to *Eco*RI based fingerprints. The observation that fingerprint complexity is not affected by the extra selective nucleotide in *Eco*RI/*Mse*I primer combinations can be explained by either (A) methylation of at most

three out of four *Pst*I recognition sites, or (B) compositional bias of C/G versus A/T residues. When assuming a 33%:67% CG versus AT compositional bias, then the overrepresentation of *Eco*RI recognition sites (GAATTC) relative to the under representation of *Pst*I recognition sites (CTGCAG) is also fourfold. Therefore, not only methylation of the *Pst*I recognition sites, but also the CG nucleotide bias may contribute to the reduction of the complexity of *Pst*I/*Mse*I templates. In view of the AT-rich composition of plant genomes, only a modest effect of methylation is expected in plant genome. The positive correlation found in this study between the number of bands per primer combination and the overall AT content in the selective nucleotides of the primers can also be ascribed to compositional bias of C/G versus A/T residues as well.

Level of polymorphisms

The lower percentage of polymorphic bands per primer combination for the cross between the fiber flax cultivars as compared to the cross between a fiber flax and linseed cultivar is not unexpected, since a previous study of Vromans *et al.* (Chapter 2) showed that the genetic distance between fiber flax cultivars is smaller than the distance between a fiber flax and linseed cultivar. The left-skewed distribution found in this study is due to the very small percentage of polymorphisms present in flax. These exceptional low percentages of polymorphic markers are seldom reported between cultivars of other crops. This underlines the importance of a preliminary selective screen of AFLP primer combination in flax.

Codominant markers

The presence of codominant markers in AFLP analysis have been reported in several species, like *Arabidopsis thaliana* (19.2%, Alonso Blanco *et al.* 1998), barley (8.7%, Vaz Patto 2001), *Thlaspi caerulescens* (7.6%, Assunção pers. comm.), tomato (17%, Saliba Colombani *et al.* 2000), melon (26.4%, Perin *et al.* 2002) and rice (10.6%, Maheswaran *et al.* 1997). The occurrence of allelic codominant markers in flax has previously been reported by (Spielmeyer *et al.* 1998). In total 19% of the segregating bands were classified by Spielmeyer *et al.* (1998) as putative codominant alleles. However, the observed proportion of codominant markers observed in the present study (35.9% in BxH and 37.9% in VxE) greatly exceeds the values detected by Spielmeyer *et al.* (1998) and in other species. We concluded that the high proportions found in flax are the result of the selection of primer combinations, because bi-allelic codominant markers are initially scored as two polymorphic bands in those primer combinations. A disadvantage of these bi-allelic codominant markers is the reduction of total number of marker loci. However, the utility of highly informative codominant markers for constructing genetic linkage maps and QTL analysis is of more value than the decrease of total polymorphic bands specifically in F4-

F5 RILs, as it can distinguish between homozygous and heterozygous regions in the genome of a specific genotype.

There are two causes of length polymorphisms in AFLPs. Although only two codominant markers were sequenced in this study, of which one was based on microsatellite size variation, we assume that a significant amount of microsatellite variation is present in all the allelic codominant markers obtained in flax. Next to microsatellite variation it is expected to find INDEL allelic codominant markers as well, as was observed in the other allelic codominant marker pair. In some cases, analysis of allelic codominant markers will result in two completely different sequences, suggesting that the two polymorphic bands are highly linked in coupling phase instead of being allelic. The latter case of the allelism is not expected to occur so often, as amplification of all allelic codominant markers with the extended primers resulted in the amplification of both alleles.

Remarkably, the ability to detect codominant markers is mainly restricted to self-fertilizing crops, which are in general low in percentage of polymorphic bands. Genetic bottlenecks will decrease the within population genetic variation of selfers more rapidly than in related outbreeding species (Savolainen *et al.* 2000). However, the decrease of genetic variation is highly dependent on the duration of the bottleneck (expressed in generations) and the number individuals passing the bottleneck (Eyre-Walker *et al.* 1998). Furthermore, the mutation rate of microsatellites ranges from 10^{-6} to 10^{-2} per generation and thus is significantly higher than estimated base substitution rates (Schlotterer 2000). A decrease of genetic diversity as a consequence of a bottleneck or breeding activities, which is the case in flax (Vromans *et al.*, Chapter 2), will therefore result in severe limitations to develop molecular markers, but proportionally the microsatellites are the first elements in the genome to regain DNA variation. Furthermore, the observed length differences between the alleles typically represent size variation expected for microsatellites. This means that the high number of codominant markers observed in this study mainly relies on microsatellite variation. Further investigation of these codominant markers would be necessary to validate this conclusion.

Enlarging the flax gene pool

This study demonstrates a low level of molecular variation within cultivated flax. This corroborates with the difficulty to breed new cultivars that meet the Distinctness, Uniformity and Stability (DUS) criterion (Everaert *et al.* 2001). Therefore the introduction of wild germplasm might be an important step forward to enlarge the molecular and phenotypic variation of the flax gene pool. Previous studies in tomato (Eshed and Zamir 1994), cabbage (Ramsay *et al.* 1996) and lettuce (Jeuken and Lindhout 2004) show that the introduction wild germplasm by means of a backcross introgression library can be a very powerful approach for the improvement of new varieties. There are not many wild flax

species which can be used for the introduction wild germplasm. A few species can be crossed with *L. usitatissimum* and produce fertile F1 interspecific hybrids, and *L. bienne* (synonymous: *L. angustifolium*) would be the good candidate. The construction of BILs would not suffer from a low level of molecular polymorphism, because the distance of *L. bienne* to the different cultivars is almost three times higher (Table 4) than between cultivars (Vromans *et al.*, Chapter 2).

Table 4 Distance matrix between four crossing parents and the wild flax species *L. bienne*. Distances are calculated using Jaccard similarity values. Values are based on five primer combinations. Distance values in parenthesis are based on all primer combinations of this study.

	Viking	Hermes	E1747	*L. bienne*
Hermes	0.08	x	x	x
E1747	0.13 (0.08)	0.14	x	x
L. bienne	0.27	0.30	0.26	x
Belinka	0.10	0.10 (0.05)	0.14	0.28

A manifold reduction of workload can be achieved for the construction of an interspecific linkage map as compared to the construction of an intraspecific linkage map without preliminary screening. If the most polymorphic AFLP primer combinations are selected for the interspecific cross, the average number of markers per primer combination of the selected primer combinations should exceed 20. Therefore an efficient construction of a BIL library in flax would be feasible. Although it is not known a priori what the added value of *L. bienne* is to flax breeding, positive alleles might be present in *L. bienne*. From for example tomato, we know that indeed unexpected positive alleles can lead to improved traits and cultivars (Tanksley and McCouch 1998; Zamir 2001).

General conclusion

This study shows that screening a large number of AFLP primer combinations can be beneficial to increase the efficiency in mapping studies in two ways. First, depending on the genetic distance of the chosen parents, a huge reduction of labor can be achieved. Secondly, a larger number of codominant markers are to be expected in the selected primer combinations, especially in self-fertilizers with low genetic diversity. We can conclude that screening a window of 1000 AFLP primer combinations is profitable for almost every crop and definitively for flax.

Acknowledgements

We are grateful to Steve Knapp for providing the VxE population. We gratefully acknowledge the Dutch breeding campanies Cebeco Seeds, ‘Van de Bilt Zaden en Vlas’ and Dobbelaar Breeding and the Dutch provinces Zeeland for their financial contribution.

References

Alonso Blanco, C., Peeters, A.J.M., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., Pot, J. and Kuiper, M.T.R. (1998) Development of an AFLP based linkage map of *Ler*, *Col* and *Cvi Arabidopsis thaliana* ecotypes and construction of a *Ler*/*Cvi* recombinant inbred line population. *Plant Journal* 14: 259-271

Bennett, M.D. and Leitch, I.J. (2004) Plant DNA C-values database (release 3.0) http://www.rbgkew.org.uk/cval/homepage.html

Bert, P.F., Charmet, G., Sourdille, P., Hayward, M.D. and Balfourier, F. (1999) A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers. *Theoretical and Applied Genetics* 99: 445-452

Brugmans, B., van der Hulst, R.G., Visser, R.G., Lindhout, P. and van Eck, H.J. (2003) A new and versatile method for the successful conversion of AFLP markers into simple single locus markers. *Nucleic Acids Research* 15; 31: e55

Eshed, Y. and Zamir, D. (1994) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: a tool for fine mapping of genes. *Euphytica* 79: 175-179

Everaert, I., de Riek, J., de Loose, M., van Waes J., and van Bockstaele, E. (2001) Most similar variety grouping for distinctness evaluation of flax and linseed (*Linum usitatissimum* L.) varieties by means of AFLP and morphological data. *Plant Varieties and Seeds* 14: 69-87

Eyre-Walker, A., Gaut, R.L., Hilton, H., Feldman, D.L. and Gaut, B.S. (1998) Investigation of the bottleneck leading to the domestication of maize. *Proceedings of the national academy of sciences of the United States of America* 95: 4441-4446

Fu, Y.B., Diederichsen, A., Richards, K.W. and Peterson, G. (2002) Genetic diversity within a range of cultivars and landraces of flax (*Linum usitatissimum* L.) as revealed by RAPDs. *Genetic Resources and Crop Evolution* 49: 167-174

Haanstra, J.P.W., Wye, C., Verbakel, H., Meijer-Dekens, F., van den Berg, P., Odinot, P., van Heusden, A.W., Tanksley, S., Lindhout, P. and Peleman, J. (1999) An integrated high density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* x *L. pennellii* F-2 populations. *Theoretical and Applied Genetics* 99: 254-271

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

Isidore, E., van Os, H., Andrzejewski, S., Bakker, J., Barrena, I., Bryan, G.J., Caromel, B., van Eck, H., Ghareeb, B., de Jong, W., van Koert, P., Lefebvre, V., Milbourne, D., Ritter, E., van der Voort, J.R., Rousselle Bourgeois, F., van Vliet, J. and Waugh, R. (2003) Toward a marker-dense meiotic map of the potato genome: lessons from linkage group I. *Genetics* 165: 2107-2116

Jeuken, M.J.W. and Lindhout, P. (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. *Theoretical and Applied Genetics* 2004; 109: 394-401

Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A. and Karp, A. (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* 3: 381-390

Karlin, S. and Mrazek, J. (1997) Compositional differences within and between eukaryotic genomes. *Proceedings of the national academy of sciences of the United States of America* 94: 10227-10232

Ma, H., Moore, P.H., Liu, Z.Y., Kim, M.S., Yu, Q.Y., Fitch, M.M.M., Sekioka, T., Paterson, A.H. and Ming, R. (2004) High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. *Genetics* 166: 419-436

Maheswaran, M., Subudhi, P.K., Nandi, S., Xu, J.C., Parco, A., Yang, D.C. and Huang, N. (1997) Polymorphism, distribution, and segregation of AFLP markers in a doubled haploid rice population. *Theoretical and Applied Genetics* 94: 39-45

Menendez, C.M., Hall, A.E. and Gepts, P. (1997) A genetic linkage map of cowpea (*Vigna unguiculata*) developed from a cross between 2 inbred, domesticated lines. *Theoretical and Applied Genetics* 95: 1210-1217

Negi, M.S., Sabharwal, V., Bhat, S.R. and Lakshmikumaran, M. (2004) Utility of AFLP markers for the assessment of genetic diversity within *Brassica nigra* germplasm. *Plant Breeding* 123: 13-16

Oh, T.J., Gorman, M. and Cullis, C.A. (2000) RFLP and RAPD mapping in flax (*Linum usitatissimum*). *Theoretical and Applied Genetics* 101: 590-593

Paran, I., van der Voort, J.R., Lefebvre, V., Jahn, M., Landry, L., van Schriek, M., Tanyolac, B., Caranta, C., Ben Chaim, A., Livingstone, K., Palloix, A. and Peleman, J. (2004) An integrated genetic linkage map of pepper (*Capsicum* spp.). *Molecular Breeding* 13: 251-261

Perin, C., Hagen, L.S., del Conto, V., Katzir, N., Danin Poleg, Y., Portnoy, V., Baudracco Arnas, S., Chadoeuf, J., Dogimont, C. and Pitrat, M. (2002) A reference map of *Cucumis melo* based on two recombinant inbred line populations. *Theoretical and Applied Genetics* 104: 1017-1034

Qi, X. and Lindhout, P. (1997) Development of AFLP markers in barley. *Molecular and General Genetics* 254: 330-336

Qi, X., Stam, P. and Lindhout, P. (1998) Use of locus-specific AFLP markers to construct a high-density molecular map in barley. *Theoretical and Applied Genetics* 96: 376-384

Ramsay, L.D., Jennings, D.E., Bohuon, E.J.R., Arthur, A.E., Lydiate, D.J., Kearsey, M.J., Marshall, D.F. (1996) The construction of a substitution library of recombinant backcross lines in *Brassica oleracea* for the precision mapping of quantitative trait loci. *Genome* 39: 558-567

Rowland, G.G. (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). *Canadian Journal of Plant Science* 71: 393-396

Saliba Colombani, V., Causse, M., Gervais, L. and Philouze, J. (2000) Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome* 43: 29-40

Savolainen, O., Langley, C.H., Lazzaro, B.P. and Freville, H. (2000) Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Molecular Biology and Evolution* 17: 645-655

Schlotterer, C. (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* 109: 365-371

Sebastian, R.L., Howell, E.C., King, G.J., Marshall, D.F. and Kearsey, M.J. (2000) An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations. *Theoretical and Applied Genetics* 100: 75-81

Sneath, P.H.A. and Sokal, R.R. (1973) Numerical taxonomy: the principles and practice of numerical classification. Freeman, San Francisco: 573 p

Spielmeyer, W., Green, A.G., Bittisnich, D., Mendhem, N. and Lagudah, E.S. (1998) Identification of quantitative trait loci contributing to Fusarium wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). *Theoretical and Applied Genetics* 97: 633-641

Tanksley, S.D. and McCouch, S.R. (1998) Seed banks and molecular maps: unlocking genetic potential from the wild. *Crop Genetic Resources*: 49-52

Van der Beek, J.G., Verkerk, R., Zabel, P. and Lindhout, P. (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. *Theoretical and Applied Genetics* 84: 1-2

Van Eck, H.J., van der Voort, J.R., Draaistra, J., van Zandvoort, P., van Enckevort, E., Segers, B., Peleman, J., Jacobsen, E., Helder, J. and Bakker, J. (1995) The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. *Molecular Breeding* 1: 397-410

Vaz Patto, M.C. (2001) the genetics and mechanism of avoidance of rust infection in *Hordeum chilense*. PhD thesis Wageningen University: 122p

Vivek, B.S. and Simon, P.W. (1999) Linkage relationships among molecular markers and storage root traits of carrot (*Daucus carota* L. ssp. *sativus*). *Theoretical and Applied Genetics* 99: 58-64

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414

Vuylsteke, M., Mank, R., Antonise, R., Bastiaans, E., Senior, M.L., Stuber, C.W., Melchinger, A.E., Lubberstedt, T., Xia, X.C., Stam, P., Zabeau, M. and Kuiper, M. (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theoretical and Applied Genetics* 99: 921-935

Zamir, D. (2001) Improving plant breeding with exotic genetic libraries. *Nature Reviews Genetics* 2: 983-989

Chapter 5

The construction of an integrated linkage map of flax (*Linum usitatissimum* L.) using conventional and novel mapping software

Jaap Vromans, Piet Stam and Herman J. van Eck

Abstract

For the construction of an intergrated map of flax, the value of the simultaneous use of JoinMap 3.0 and RECORD method is investigated, using data from two mapping populations. The VxE population comprised 19 linkage groups whereas the BxH population consisted of 21 linkage groups. The number of linkage groups in the integrated map is even higher (22). In the separate maps several gaps of 20 – 30 cM were present. These regions may represent interstitial segments that reached genetic fixation or reflect unequal distribution of marker loci and/or recombination. In total three and four linkage groups for the VxE and BxH population, respectively, showed a different marker order, as obtained with JoinMap or RECORD. In general the ambiguities could be solved easily, by performing a second round of RECORD analysis, comparing both marker orders with graphical genotyping and/or comparing the order of common markers across populations. It can be concluded that the simultaneous use of both RECORD and JoinMap resulted in better maps compared to linkage analysis with only one algorithm. The overall fit of the maps is very good and the order of the common markers showed nice co-linearity, which suggests that the marker order in this map is reliable and will be of great value for the mapping of QTL involved in resistance and quality traits.

Introduction

Genetic linkage maps have become an indispensable tool for locating genes or quantitative trait loci (QTL), marker assisted breeding (MAB) and map based cloning. In flax (*Linum usitatissimum*), a diploid, self-fertilizing, annual plant species, two genetic linkage maps have been reported. One genetic linkage map, based on 213 AFLPs resulted in 18 linkage groups covering approximately 1400 cM whereas flax has 15 chromosomes (Spielmeyer *et al.* 1998). This linkage map has been used to identify two quantitative trait loci involved in resistance to *Fusarium*. A second map was composed of 94 RFLP and RAPD markers comprising 15 linkage groups and covering 1000 cM (Oh *et al.* 2000). Among DNA fingerprinting techniques the AFLP method is known for its reproducibility and high multiplex ratio (Jones *et al.* 1997; Vos *et al.* 1995). For these reasons AFLP markers have been widely used in mapping studies. In spite of the high multiplex ratio of AFLP, only few polymorphic bands per primer combination can be obtained in flax. This exceptionally low level of polymorphism suggests a very narrow gene pool of cultivated flax. Especially the genetic distance between fiber flax cultivars is extremely low compared to the distance between two linseed cultivars and linseed cultivars and fiber flax cultivars (Vromans *et al.* Chapter 2).

To get around the low efficiency of marker development in flax, Vromans *et al.* (Chapter 4) showed that screening large numbers of AFLP primer combinations on parents of mapping populations is helpful to identify AFLP primer combinations that generate somewhat larger numbers of polymorphisms. This pre-screening resulted in two advantages. First, an almost 50% reduction of labor can be obtained when the most polymorphic primer combinations are used in a mapping study. Secondly, among the higher marker number obtained in this way, also a high number of length polymorphisms could be observed. These co-dominant AFLP markers are identified as allelic AFLP bands of different mobility in the selected primer combinations.

For the construction of genetic linkage maps various computer programs have been developed, such as JoinMap 3.0 (Van Ooijen and Voorrips 2001), MAPMAKER (Lincoln *et al.* 1992) and G-Mendel 3.0 (Holloway and Knapp 1994). The different computer programs make use of different cost functions such as likelihood (Jansen *et al.* 2001; Lander *et al.* 1987), the sum of adjacent recombination frequencies (SARF), the sum of adjacent LOD scores (SALOD) (Liu and Knapp 1990), number of crossovers (Thompson 1987) and ‘sum of squares’(Stam 1993).

Recently a new computer program, RECORD, has been described to calculate the order of genetic loci (Van Os *et al.* 2005). RECORD is not based on distances between marker pairs, but obtains the best marker order by minimization of the number of recombination events as counted in a data set of marker segregating data. Simulation studies have demonstrated that RECORD is specifically accurate in marker-dense regions

(Van Os *et al.* 2005). Furthermore, RECORD is hardly disturbed by singleton observations or missing values, in comparison to other mapping software packages using pairwise distance estimates (Van Os *et al.* 2005). RIL populations indeed have many apparent singletons (Nilsson *et al.* 1993), due to adjacent recombination events that took place in remaining regions of heterozygosity during the successive generations of inbreeding.

In this report we describe the value of the simultaneous use of JoinMap 3.0 and RECORD method for the improvement of an integrated genetic linkage map of flax. By comparing the marker order of JoinMap with the RECORD order the most plausible marker order could be determined. Additional evidence for the ability of RECORD to arrive at the best possible markers order could be obtained by comparing two linkage maps of different RIL mapping populations. Sufficient numbers of common markers allowed studying the colinearity of the linkage maps maps.

Materials and methods

Plant material

A population of 186 F9 recombinant inbred lines (RIL), indicated hereafter as VxE, was derived from a cross between the French fiber flax variety Viking and the Canadian linseed variety E1747 by means of single seed decent (SSD). E1747 is an EMS-induced McGregor mutant (Rowland 1991) and has a low level of linolenic acid content in the seed oil. A second RIL population, referred to as BxH, was also obtained by SSD, from a cross between two fiber flax varieties, Belinka and Hermes. The 205 lines descending from BxH were composed of two subpopulations of 131 F4 lines and 74 F7 lines, respectively. The combined BxH population could therefore be regarded as resembling F5 RILs during subsequent analysis.

AFLP analysis

Leaf material for DNA extraction was collected from 2-week old plants grown in the greenhouse. Genomic DNA was extracted from fresh leaves according to the procedure described by Van der Beek *et al.* (1992), adjusted for 96-well format using 1 ml tubes of Micronics (Micronic BV, Lelystad, The Netherlands). Leaf tissue was ground using a Retsch 300 mm shaker at maximum speed (Retsch BV, Ochten, The Netherlands). After hooking the DNA out of the isopropanol mixture, the DNA was washed overnight in 76% ethanol and 10 mM NH4Ac, dried in a vacuum drier and dissolved in 100 µl of mQ.

The AFLP procedure was performed according to the two step amplification as described by Vos *et al.* (1995) using the enzyme combinations *Eco*RI/*Mse*I and *Pst*I/*Mse*I. Instead of a radioactive label we used fluorescently (IRDyeTM 700/ IRDyeTM 800) labeled *Eco*RI+3 or *Pst*I+2 primers. The VxE population was analyzed with 36 AFLP primer combinations, and the BxH population with 45 AFLP primer combinations, of which 18

primer combinations are common. These selected primer combinations with an elevated number of polymorphic bands per lane were selected in an earlier study (Vromans *et al.*, Chapter 4; Table 1). The selection of primer combinations was based on a three step procedure. First the 45 most polymorphic primer combinations in the BxH populations were selected, because this population displayed the least amount of DNA polymorphisms. The next step was to determine the performance of these 45 primer combinations in the VxE population. The primer combinations with at least two common markers and in total more than four markers also applied on the VxE population (18 primer combinations). Finally, the total number of primer combinations in the VxE population was extended with the remaining 18 most polymorphic primer combinations to obtain enough markers.

Table 1 Selection of AFLP primer combinations for the genetic analysis of the RIL populations VxE and BxH.

*Eco*RI/*Mse*I			*Pst*I/*Mse*I		
Primer combinations common to both populations					
eAAC/mAAC	eAGT/mAAG		pCC/mACA	pGT/mCAA	pAT/mAGA
eAAT/mCCC	eAGT/mAGT		pGG/mCAA	pTT/mCCG	pAT/mCAA
eAGA/mATA			pGG/mCGT	pAG/mCTT	pAT/mCAG
eAGA/mATT			pGG/mCTC	pAT/mAAC	pCA/mACT
Primer combinations applied on the VxE population					
eAAA/mCCG	eAGA/mCAA	eAGT/mATG	pAA/mCCC	pCA/mCTG	
eAAC/mCAT	eAGA/mCAT	eAGT/mCAA	pCA/mAAG	pGA/mCCA	
eACT/mATT	eAGA/mCTA	eAGT/mCTG	pCA/mCAG	pGA/mCTG	
eAGA/mACT	eAGG/mCAT	eATC/mACC			
Primer combinations applied on the BxH population					
eAAA/mAAG	eAAT/mAGT	eAGC/mCCT	pAA/mCAT	pCA/mCTC	pGT/mCAT
eAAA/mCAG	eAAT/mATG	eATA/mCAG	pAC/mATC	pCC/mCCT	pGT/mCCA
eAAA/mCCA	eACG/mCAA	eATC/mCAT	pAT/mATT	pCT/mCTA	pTA/mAGG
eAAG/mAAC	eACG/mCCC	eATG/mAAA	pAT/mCCA	pGG/mCTG	pTA/mCAC
			pAT/mCTG	pGT/mAAC	pTT/mCCA

Gel electrophoresis and data capture from gel images

AFLP fragments were loaded on a polyacrylamide gel and analyzed on a NEN® Global Edition IR2 DNA Analyzer (LI-COR® Biosciences, Lincoln, NE). The marker data were collected in both populations by recording absence or presence, and intensity differences of the bands to distinguish between homozygote and heterozygote genotypes that could be recognized by QuantaPro software (Keygene product, NL). In some cases length polymorphisms were recognized in the same primer combination. We assign two polymorphic bands as two alleles from one codominant marker if the two polymorphic bands comply with the two criteria formulated by Alonso Blanco *et al.* (1998) (*i*) Two AFLP bands might be allelic when they are derived from different parents, within the same primer

combination. (*ii*) The two AFLP markers segregate complementary. Two allelic segregating bands of this type were manually scored as one co-dominant marker.

AFLP marker nomenclature

AFLP markers nomenclature is based on the enzyme combination (e=*Eco*RI, p=*Pst*I and m=*Mse*I), the selective nucleotides, followed by the mobility in the gel compared to a 10 base ladder (Sequamark; Research Genetics, Huntsville, AL). Decimal points in the mobility values are due to interpolation of band sizes between 10-bp markers by Quantar software. Furthermore the parental origin of the fragment is shown after the first underscore (V=Viking, E=E1747, B=Belinka and H=Hermes) and the linkage group number as calculated by RECORD (Van Os *et al.* 2005) after the second underscore. Common markers are displayed without the linkage group number as calculated by RECORD, as they occur in both populations and therefore have two RECORD group numbers.

Graphical genotyping

MS Excel was used to display graphical genotypes. Graphical genotypes are a representation of the raw data with the markers in rows, and plants in columns. The background color of the cells is formatted, conditional to the cell value, being the maternal or paternal allele. When the markers are sorted according to the marker order as calculated by mapping software, visual inspection of the map quality is easily performed, where the least interrupted vertical zebra-like pattern is perceived as the best marker order (see e.g. Figure 1). Vertical patterns that were highly interrupted were re-examined to seek for an improvement of the marker order.

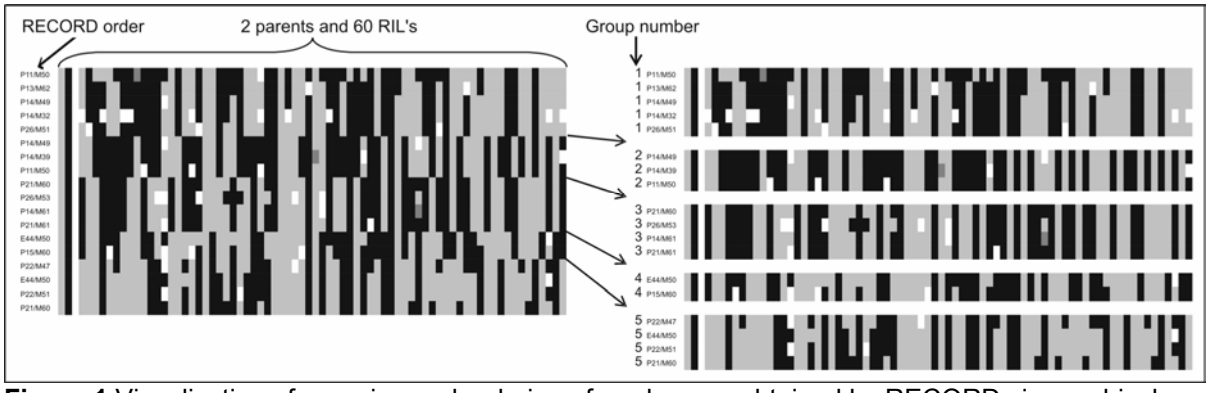

Figure 1 Visualization of grouping and ordering of markers as obtained by RECORD via graphical genotyping. Light grey or black indicate the observation of a marker allele originating from the female or male parent, Belinka or Hermes, respectively. Intermediate grey represent heterozygosity as inferred by the joint presence of co-dominant marker alleles. Heterozygosity for dominant markers is ignored. White represents missing observations.

Comparative grouping of markers with RECORD and JoinMap

Although RECORD has been developed to order markers that belong to a single linkage group, we have used RECORD to analyze the complete raw data set. This should place markers with similar segregation patterns at nearby positions to provide a first grouping of markers that belong to the same linkage group. The resulting ordered data set was imported in excel and inspected after graphical genotyping. Interruptions between patterns of vertical stripes, as observed in graphical genotyping images, were used to split the total data set into a tentative grouping. The group number was included in the marker name (Figure 1) to allow comparative analysis between this method of grouping and the grouping as obtained by the computer program JoinMap.

In JoinMap markers of the VxE and BxH populations were assigned to linkage groups using the 'grouping node' of the software. Upon scanning the number of groups by increasing the LOD score for grouping with steps of one LOD unit, starting with LOD = 1 up to LOD = 15. A stable grouping was in general obtained between LOD = 3 and 8 for VxE and LOD = 4 and 7 for BxH. However, the stability varied among groups. For both populations marker segregation distortion was tested against expected Mendelian ratios using a chi-square goodness-of-fit test with a threshold level for significance of 0.5%.

Map construction

Marker orders were calculated with RECORD and JoinMap as described by Van Os *et al.* (2005) and Van Ooijen and Voorrips (2001), respectively. Specific settings for the linkage mapping for VxE are as follows: rec < 0.45; LOD > 1.0; mapping function = Kosambi and popt = RIL9. For the BxH population the difference between F4 and F7 has to be taken into account when using JoinMap. Pair wise distances between marker loci (PWD output) were calculated within each subpopulation with corrections for the level of inbreeding. Subsequently, the F4 and F7 PWD output was used to generate a weighted PWD input file to calculate the map of the complete BxH mapping population. In case JoinMap and Record arrived at a different marker orders, inspection of graphical genotypes was used to choose the best solution. If necessary "fixed order files" were included in JoinMap to arrive at mapping results that was deemed proper. Specifically the integration of the VxE and BxH required fixed order files, to maintain the order of the markers as observed in the separate VxE and BxH maps. In rare cases the mapping parameters differed from the settings described before, which cases are indicated in the results.

Nomenclature of linkage groups was assigned as follows. Linkage groups of the separate VxE and BxH maps were numbered consecutively, according to decreasing map length (in cM) as criterion. The nomenclature of the 22 integrated linkage maps is indicated with letters A through V.

Results

General results

Separate maps of the VxE and BxH mapping population were constructed and subsequently integrated into a consensus map of flax. The VxE population was analyzed with 36 primer combinations resulting in 2352 amplification products of which 385 bands (16.5 %) segregated in the progeny. Of those 385 segregating bands 146 showed an alternating segregation of pairs of alleles, resulting in 73 bi-allelic codominant marker loci. As a result 312 markers were obtained and used for marker ordering by RECORD and linkage analysis with JoinMap (Table 2). In the BxH population more primer combinations were employed than in VxE, to compensate for the lower level of polymorphisms between both fiber flax parental genotypes. Of the 2785 AFLP fragments 323 bands were segregating in the population of which 116 are allelic bands (resulting in 58 bi-allelic codominant markers). Of the remaining 207 marker loci 75 loci showed clear intensity differences between homozygous and heterozygous genotypes. This allowed the collection of codominant marker data using Quantar pro software. This extra effort was taken because a subset of the mapping population contains a considerable number of heterozygous marker genotypes, which contributes to the reliability of the map and the ability to identify singletons or scoring errors. In addition, one morphological marker, the flower color locus *P*, was included in the BxH map. Finally, 265 markers were used for marker ordering and 266 markers for the construction of a linkage map (Table 2).

Table 2 Overview of the number of marker loci in the VxE and BxH mapping population

Population	Enzyme combination	# of primer combinations	Total # of bands	# of segregating bands	# of allelic codominant markers	Total # of markers
VxE	*Eco*RI/*Mse*I	18	1138	206	38	168
	PstI/*Mse*I	18	1214	179	35	144
	Total	36	2352	385	73	312
BxH	*Eco*RI/*Mse*I	18	1077	114	12	102
	PstI/*Mse*I	27	1708	209	46	163
	Total	45	2785	323	58	265

VxE linkage map

Analysis of the graphical genotyping images of the patterns of the segregating markers (example shown in Figure 1) resulted in the division of markers in 56 clear groups. To maintain this tentative grouping and to compare these grouping results with the results of JoinMap the markers names received an additional number reflecting the RECORD group. JoinMap could group 301 out of the 312 markers (96.5%) into 19 linkage groups, where each group contained at least four markers. Further reduction of the 19 groups to the expected 15 linkage groups by lowering the LOD threshold was not considered in this stage. This issue will be addressed during the construction of the integrated map of the

VxE and BxH population. The total map length comprised 1257 cM. Besides few dispersed markers with a skewed segregation three linkage groups had a region with a number of skewed markers: the bottom of VxE-1, the top of VxE-15 and the entire group VxE-20. In the F9 only 0.4% of the loci should be heterozygous, thus a more or less 1:1 ratio could be expected. However an average segregating ratio of 55:105 was observed for marker alleles of the VxE-1 group, 48:116 for VxE-15 and 33:130 for VxE-20. In all cases the segregation distortion was profoundly in favor of E1747 alleles.

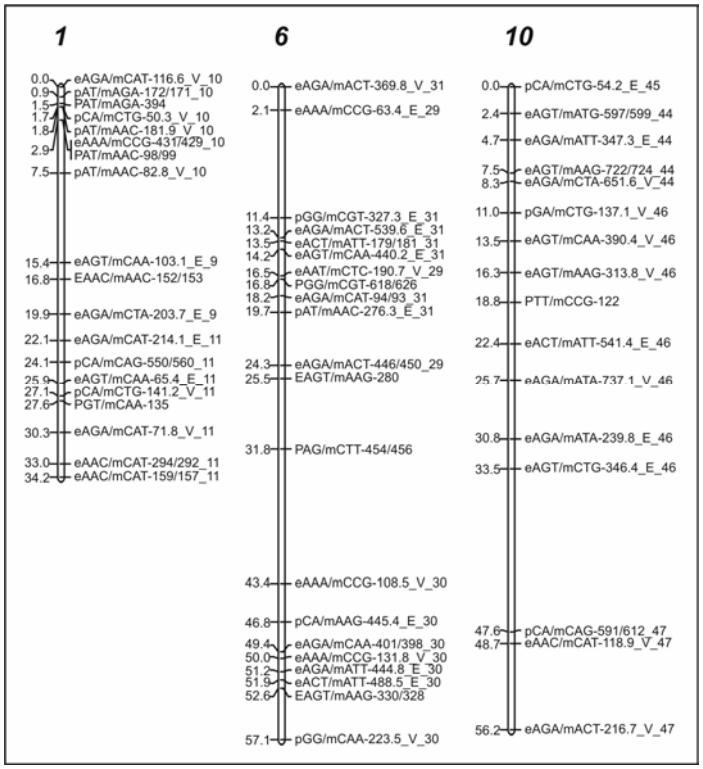

Figure 2 Three linkage groups of VxE with conflicting order of RECORD subgroups and/or jumbled order of markers when the results of JoinMap and RECORD are compared. The last two digits of the marker name reflect the RECORD group number.

In general a low mean chi-square square contribution (<1.0) was observed for all linkage groups, which implies good quality of the linkage groups. After grouping and the calculation of the linkage maps the grouping and marker order of RECORD and JoinMap could be compared. In all cases the RECORD groups fell entirely within JoinMap groups. In most cases the consecutive order of multiple RECORD groups within a single JoinMap linkage group was retained. However, three exceptions were observed where the order of RECORD groups and/or the order of markers differed between JoinMap and RECORD (Figure 2). When a new RECORD analysis was performed with only the markers of the groups VxE-1 and VxE-10, the order of the RECORD was identical to the marker order obtained with JoinMap. This initial difference is cause by the “attraction” between markers from different chromosomes that cause head-to-tail junctions in the RECORD analysis of

the entire marker dataset. However the new RECORD order of markers in linkage group VxE-6 remained stable and different to the JoinMap order.

Comparison of the graphical genotyping images of the marker order of linkage group VxE-6 as calculated by JoinMap and RECORD (Figure 3) showed that the marker order obtained by RECORD improved the appearance of the graphical genotyping images, because three markers (all belong to the initial RECORD group 29) were placed on the bottom of the linkage group. The final map was generated with JoinMap, while imposing the marker order of VxE-6 as obtained with RECORD by adding a fixed order file to JoinMap. The final map obtained with JoinMap even had a lower mean chi-square contribution and showed colinearity of the common markers.

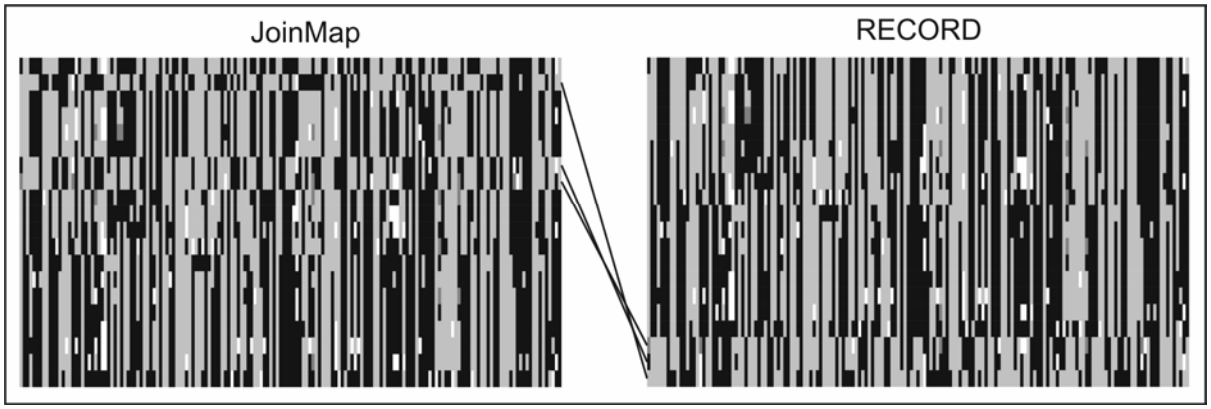

Figure 3 Graphical genotyping images of the marker order in linkage group VxE-6 according to JoinMap and RECORD

BxH linkage map

The raw AFLP data of the BxH subpopulations was processed with RECORD to order the marker loci. Again, analysis of graphical genotypes allowed the grouping of markers into 63 small groups. Subsequently, linkage maps of both BxH subpopulations were constructed with JoinMap, which resulted in 31 groups for the F7 population and 24 groups for the F4 population respectively, all containing at least 3 markers. The mean chi-square contribution contribution of almost all linkage groups in both subpopulations was lower than one. Only three groups of the F4 subpopulation had a mean chi-square contribution between one and 1.5. Skewed segregation was observed for all markers of RECORD group_9 in favor of Belinka alleles. To obtain one final BxH linkage map, the maps of both subpopulations were integrated by joining the F4 and F7 pair-wise distance estimates. This resulted in a map of 21 linkage groups, comprising 249 markers (93.6%) and leaving 16 markers unassigned. The total map length was 1178 cM. Again only three linkage groups with a mean chi-square contribution between one and 1.5, namely linkage group BxH-1, -13 and -9, were obtained.

Comparative analysis of the order of the RECORD and JoinMap results showed that the BxH map had four linkage groups with a contradicting marker order (Figure 4). A

second analysis with RECORD using markers that belong to a single linkage group (as inferred by JoinMap), did not resolve the contradictions in marker order as obtained with RECORD and JoinMap for linkage group BxH-1, BxH-3 and BxH-13. For linkage group BxH-11 it happened, that groups that were entirely separate in RECORD appeared to group in JoinMap. On the other hand, when RECORD analyzed the marker order using markers from the initial RECORD grouping: 33, 34, 50, 51 and 52 (linkage group BxH-11), the result was identical to the marker order calculated by JoinMap. The big gap between the _50 and _34/_33 markers is probably caused by either fixation between the parents and/or presence of a recombination hotspot. RECORD doesn't recognize _50 markers as being the markers with the least number of crossing overs with _34 markers, whereas the grouping function of JoinMap is able to put the markers in the same group before the markers are ordered.

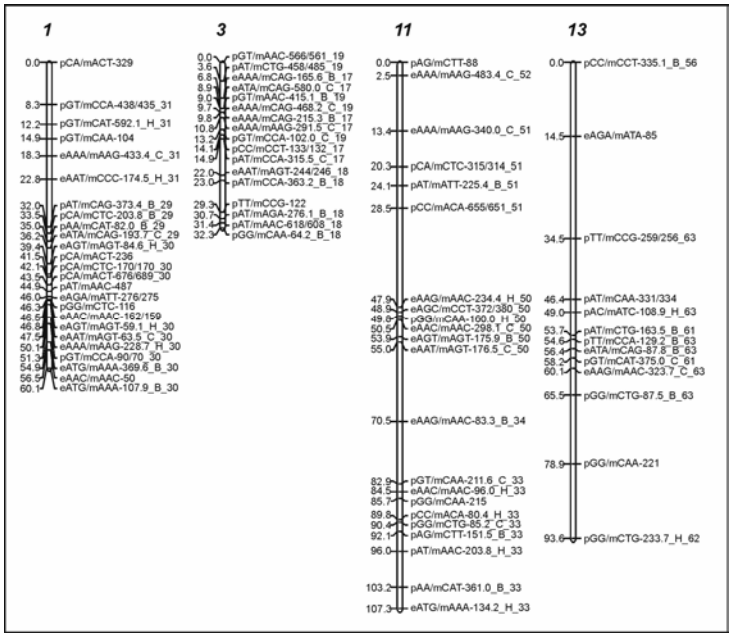

Figure 4 Four linkage groups of BxH with conflicting order of RECORD subgroups and/or jumbled order of markers when the results of JoinMap and RECORD are compared.

Comparison of the graphical genotyping images of the marker order of linkage groups BxH-3 and BxH-13 as calculated by JoinMap and RECORD (Figure 5) showed that the marker order obtained by RECORD improved the appearance of the graphical genotyping images. In more detail the raw data format feature showed that the _19 markers of linkage group BxH-3 should belong at the bottom of the linkage group, which was according to the RECORD order, instead of the group in between _17 markers (Figure 5). Similarly, the raw data format feature showed that the two _61 markers of linkage group BxH-13 should belong at the bottom of the linkage group, which was according to the RECORD order as well, instead of the middle of the linkage group (Figure 5). Both graphical genotyping images of the marker order of linkage groups BxH-1 as calculated by JoinMap

and RECORD showed a reliable pattern (data not shown). Therefore we could not decide which ordering algorithm resulted in the ‘best’order.

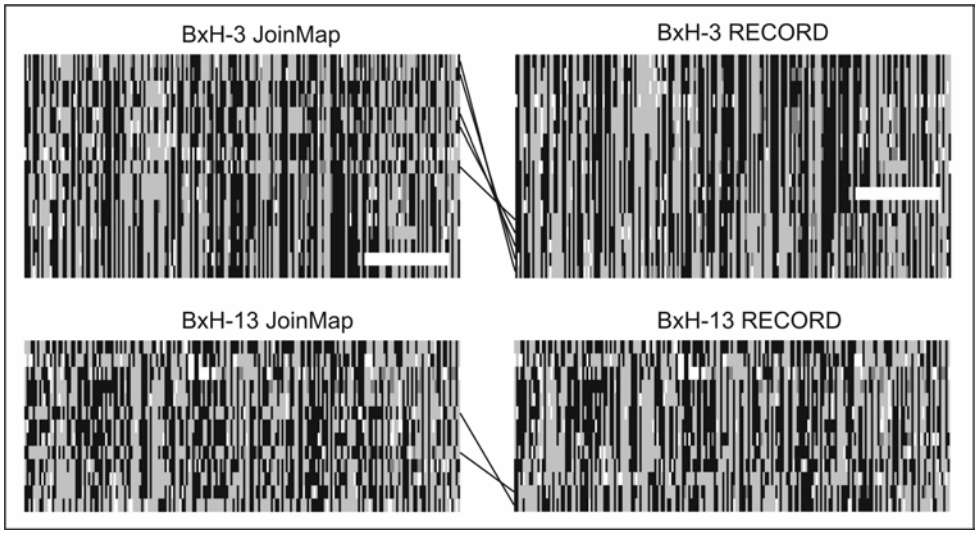

Figure 5 Graphical genotyping images of linkage group BxH-3 and BxH-13 according to JoinMap and RECORD marker order. The JoinMap derived marker order shows a vertical pattern often disrupted by markers at unlikely positions. Therefore the marker order of RECORD is regarded as more obvious.

Integrated map and co-linearity between the three maps

The integrated map of flax is comprised of 525 markers of which 502 (95.6%) are assigned to linkage groups and 23 markers remained unassigned. The integration of linkage groups is based on 51 common markers (10.2%) bridging the VxE and BxH linkage groups. The integrated map consists of 13 fully integrated linkage groups (374 markers). The next four groups share only one common marker (74 markers), which prevents orientation of the groups. Finally, two groups of the VxE linkage map (42 markers) and three groups of the BxH linkage map (12 markers) remained without common markers (Figure 6).

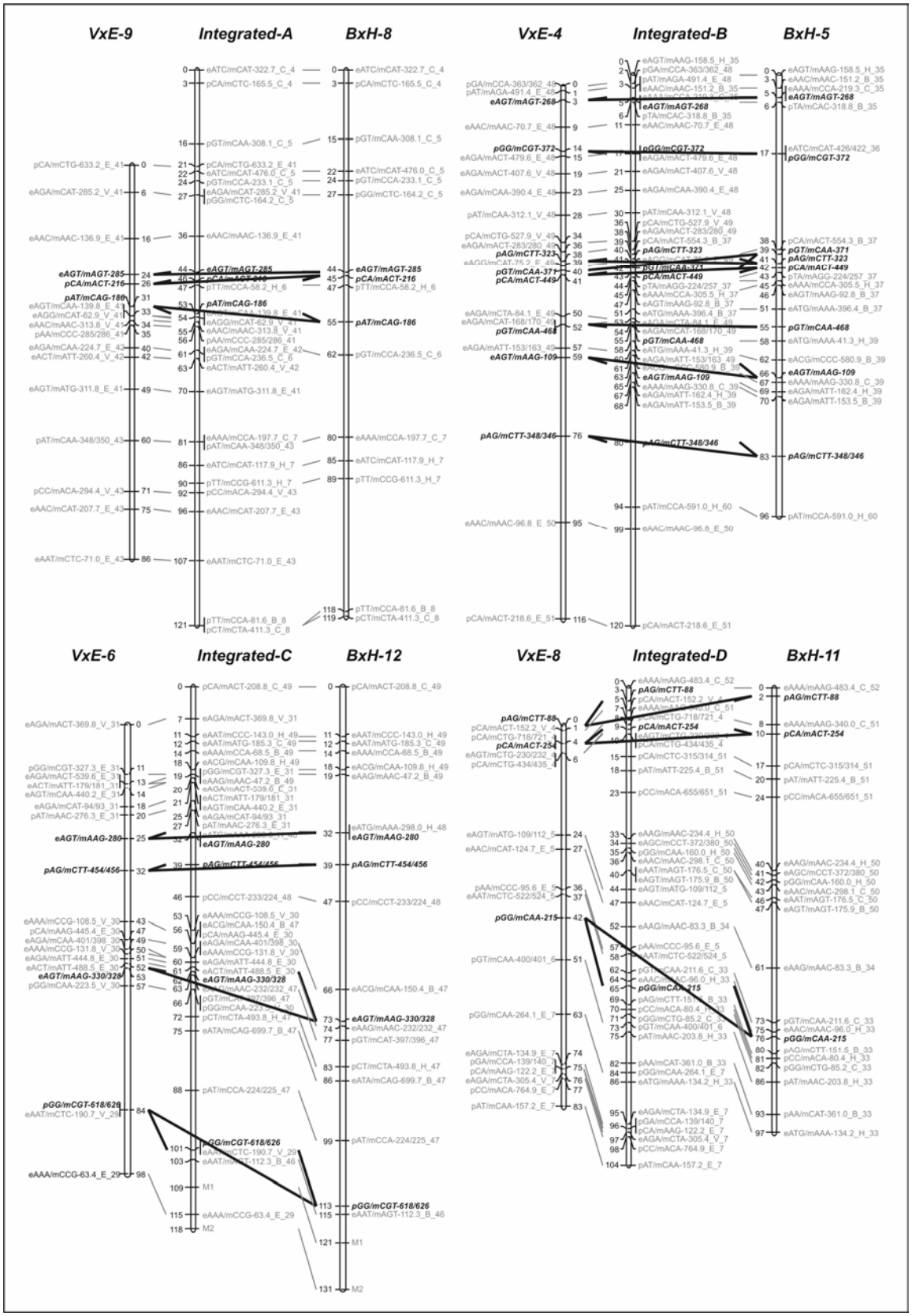

Figure 6 The integrated map of flax flanked by the maps of the BxH and VxE mapping population. Markers in grey were unique to one population. The markers that are common to VxE and BxH are shown in ***bold and italic,*** and were used to bridge linkage groups. The alignment is indicated with thick lines. Markers indicated in black belong to linkage groups that could not be integrated. Markers in **bold and underlined** are a potentially common marker, but the primer combination was not applied in the other mapping population, in spite of knowing that the marker is polymorphic (data from Chapter 4)

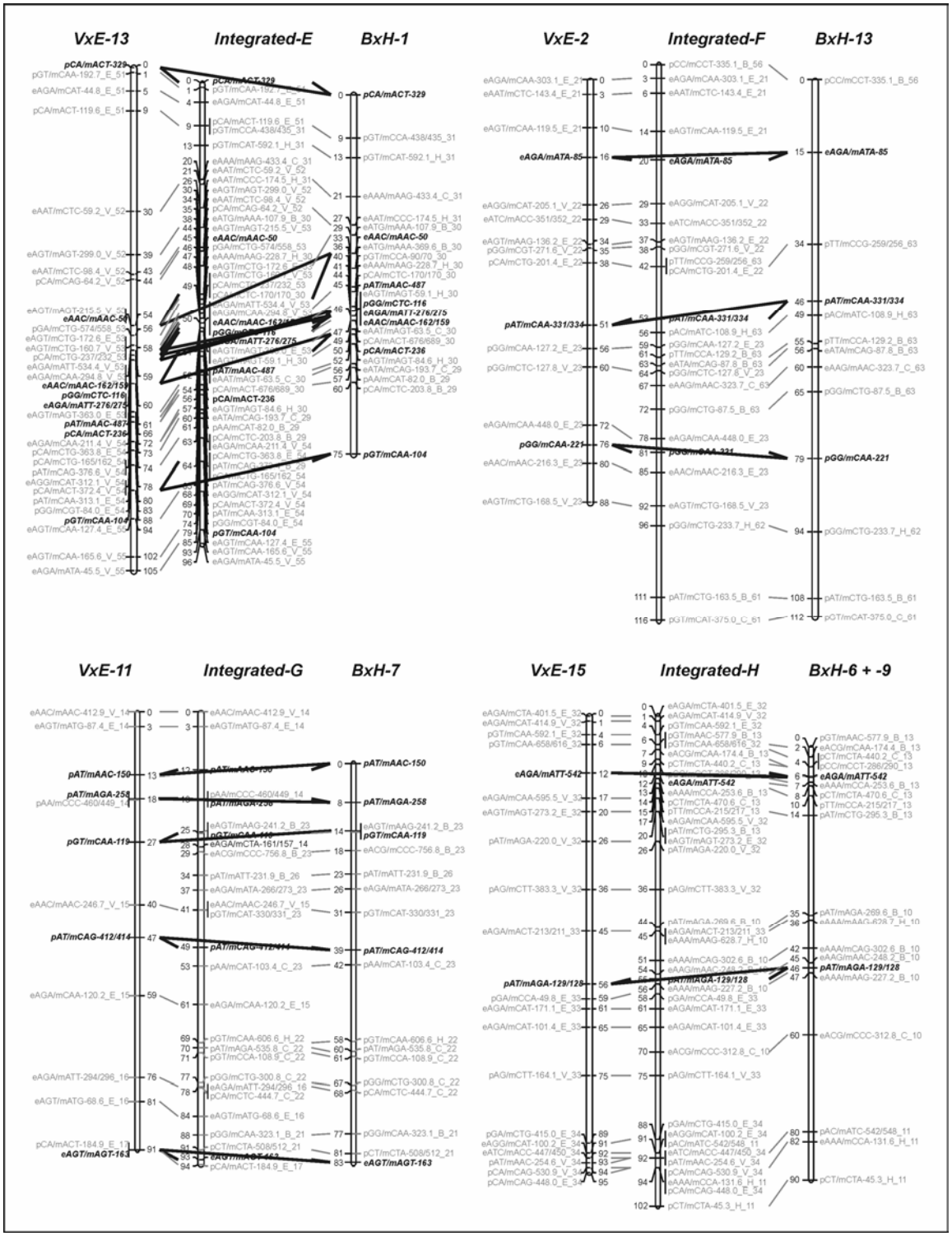

Figure 6 continued

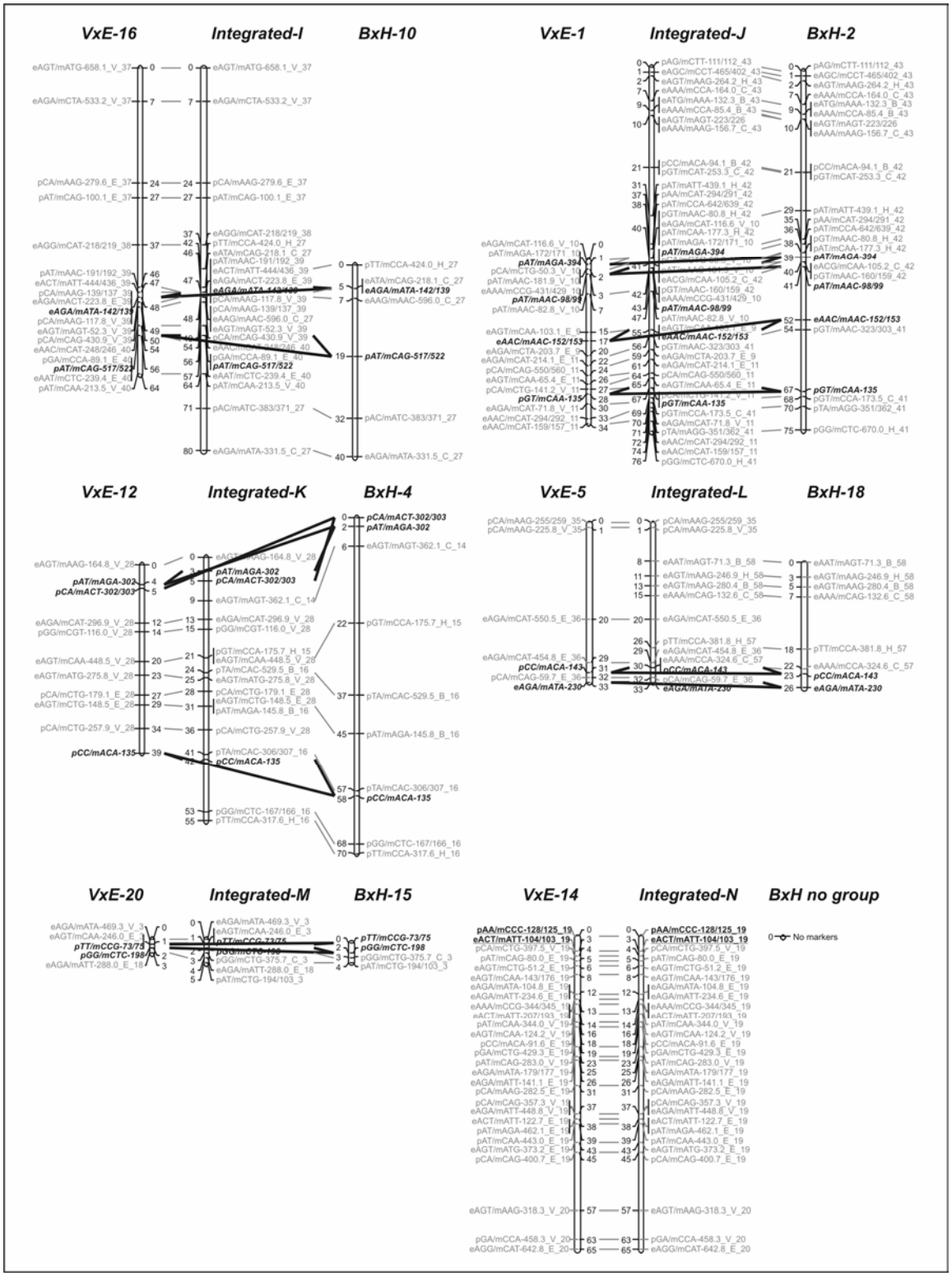

Figure 6 continued

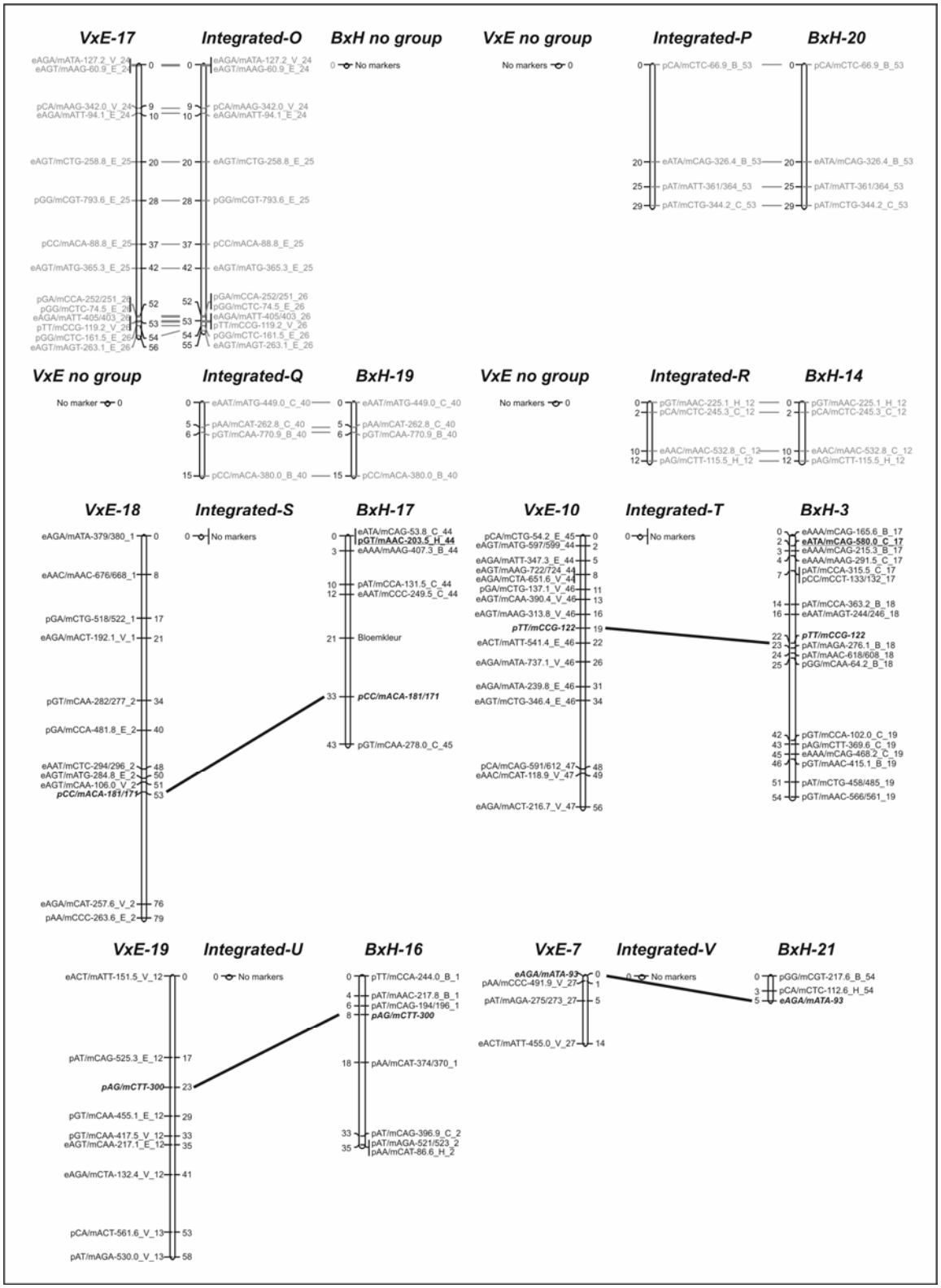

Figure 6 continued

Common markers on linkage group VxE-15 suggested the merging of two BxH linkage groups (BxH-6 and BxH-9). Ordering of markers from the initial RECORD groups 10, 11 and 13 resulted in an inversion of the markers of the groups 10 and 11 (Figure 7, middle). JoinMap could not merge these linkage groups without jumbling markers. Graphical genotyping of the marker order of JoinMap resulted in a graphical genotyping image with an unacceptable pattern (Figure 7, right). We assumed the second record order as the 'best' order.

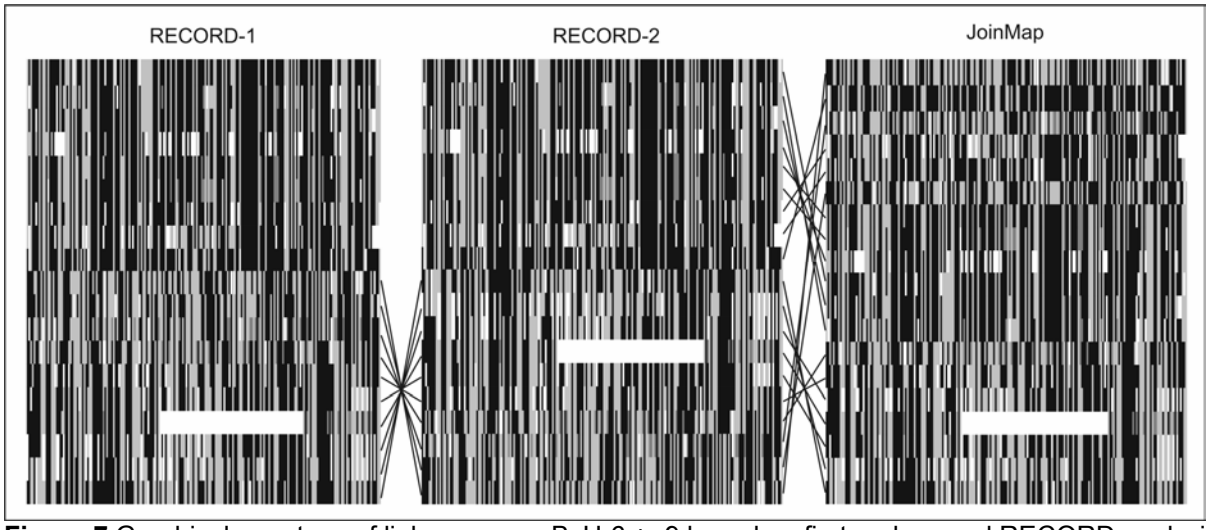

Figure 7 Graphical genotype of linkage group BxH-6 + -9 based on first and second RECORD analysis and JoinMap marker order

Among the 13 integrated linkage groups, five groups were bridged by only two common markers. Four groups were integrated with three common markers. The remaining 4 integrated groups had 4, 4, 7 and 8 common markers. The genetic distance between common markers in the separate linkage maps is quite similar, with exceptions on linkage groups integrated-D and K. This suggests that genetic distances are largely conserved across these two mapping populations. As expected the marker order was also conserved between the VxE and BxH maps. Only three out of the 51 common markers are flipped. These flips occurred only on integrated groups B, E and K between closely linked common markers.

After integration we can conclude that the ordering issue of linkage group VxE-6, discussed and solved before (Figure 2 and 3), has been solved correctly. The same marker order of common markers was observed on the homologous linkage map of the other population, that otherwise would have resulted in conflicts in the ordering of common markers, as illustrated by linkage group BxH-1 in Figure 8. Although, no final decision could be made after analyzing both graphical genotyping images of the marker order of linkage groups BxH-1 as calculated by JoinMap and RECORD, respectively, the order of the common markers in this group suggested RECORD to be considered as the best order

(Figure 8). It appeared that JoinMap is able to produce the same marker order in BxH-1 as RECORD if the Haldane's mapping function is used instead of the Kosambi's mapping function.

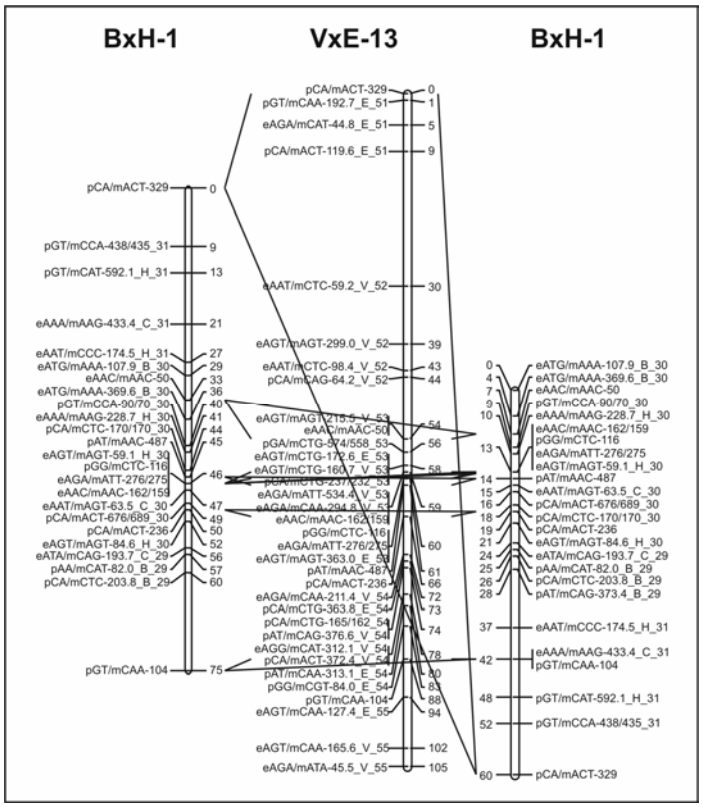

Figure 8 Alternative marker orders of BxH-1 as obtained by RECORD (left) or JoinMap (right) are compared via the order of common markers with the homologous linkage group VxE-13 (middle). The most likely order, as obtained by RECORD is indeed without conflict with the marker order of VxE13.

Discussion

The use of RECORD and graphical genotyping images to assign markers to groups

Although RECORD has not been developed with the intention to group markers into linkage groups, we have demonstrated that RECORD and graphical genotyping images can generate useful information. At this moment marker grouping is not an automated process in JoinMap and requires human choices. In this study we decided to excise many small groups of markers from the total ordered list of marker loci obtained by RECORD. Such small groups are advantageous, because this grouping (documented in the marker name) often guided our choices during the assignment of markers into groups in JoinMap and it also prompted us towards inspection of the marker order calculated by JoinMap.

Grouping has not resulted in separate maps or an integrated map with a number of linkage groups that equals the haploid chromosome number of flax (n=15). The VxE and BxH map is composed of 19 and 20 linkage groups, respectively. The number of linkage groups in the integrated map is even higher (22) due to the lack of common marker loci

bridging the homologous VxE and BxH groups. The low level of polymorphism may also result in large chromosomal segments with the same marker haplotypes. Due to these genetic fixations between the parental genotypes these map regions remain invisible to the geneticist. For example, BxH-10 seems to be represented by only a half map. On the other hand, the missing top part of BxH-10 might be disguised as BxH-14, BxH-19 or BxH-20. Several gaps of 20 – 30 cM still remain. These regions may represent interstitial segments that reached genetic fixation, but it may also reflect unequal distribution of marker loci and/or recombination. Genetic maps may split in case of interstitial regions without markers, but this is not coinciding in both VxE and BxH maps. Genetic fixation of an interstitial region, across all four parental genotypes, is not expected because one parent was not derived from the fiber but from the linseed gene pool. Unfortunately, we still we cannot offer a full explanation for the elevated number of linkage groups. Neither could Spielmeyer who arrived at 18 linkage groups and a similar map length of approximately 1400 cM (Spielmeyer *et al.* 1998). Both studies were conducted with a sufficiently large number of markers to reach genome saturation, but remarkably Oh *et al.* (2000) obtained 15 linkage groups with only one third of the marker loci. It is not unlikely that that Oh *et al.* (2000) have missed some linkage groups and split some others to compensate for a number below 15 groups.

Comparison of the marker order calculated by JoinMap or RECORD

When the order of markers on a linkage map is calculated on the basis of perfect data (no missing values and/or errors), the results of JoinMap and RECORD will not differ (Van Os *et al.* 2005). Realistic data (including missing values and scoring errors) easily lead to erroneous marker orders, but RECORD is hardly sensitive to scoring errors or missing values (Van Os *et al.* 2005). In this paper we have shown marker orders and linkage maps that were calculated with two different computer programs. Our experience suggests that the marker order of RECORD was often more plausible. This observation is in agreement with the results obtained by simulation studies (Van Os *et al.* 2005). RIL mapping populations are notoriously difficult, because during the generations of inbreeding singletons will accumulate. Singletons are equally harmful in linkage mapping as data error. We conclude that the simultaneous use of the two different computer algorithms in combination with graphical genotyping is rewarding, because ambiguities are easily revealed and the most obvious marker order can be accepted. In our view this allowed us to publish better maps than we could have obtained with only one algorithm.

Seven linkage groups showed a different marker order, as obtained with JoinMap or RECORD. In general the ambiguities could easily be solved. In two cases a second RECORD analysis solved the problem. The initial difference is cause by the “attraction” between markers from different chromosomes that cause head-to-tail junctions in the

RECORD analysis of the entire marker dataset. Comparing both marker orders with graphical genotyping appeared to be the most successful approach, as in three cases the best order could be determined. In one case the order of common markers provided useful information for the determination of the 'best' order. Only linkage group BxH-11 showed to be more problematic. In case of conflicting groups, both JoinMap and RECORD provided the final order. This supports our philosophy that better maps could be obtained if two algorithms are used simultaneously. With all the analysis done so far it is still not clear what the 'best' order of markers is for this linkage group. However the current marker order is considered for the time being as the 'best' order.

Apart from only one obvious conflicting linkage group (BxH-11) it can be concluded that the performance of both RECORD and JoinMap are equal for this data set. RECORD was mainly written to order markers in ultra-high dense linkage groups. The RECORD algorithm is much faster than the ordering algorithm of JoinMap and the cost function of RECORD is not much influenced by missing observations or scoring errors, because their effects depend on the context of such a data point, as provided by the flanking markers (Van Os *et al.* 2005). In view of the stable results that we obtained, we conclude that our data set is of high quality, containing only few scoring errors.

Reliability of the three maps

The mean chi-square contribution of all the linkage groups is very low. This suggests that the marker order in this map is robust. Furthermore, the co-linearity of the order of the common markers also supports our notions on the reliability of the three maps. The co-linearity of the three maps allows comparative analysis of QTLs that are detected on homologous positions in both mapping population. Such cases may also offer insight in the number of different QTL alleles at that specific locus.

Conclusion

In this report we described the simultaneous use of two ordering algorithms to improve the construction of genetic linkage maps. We showed that more plausible results could be obtained if the two marker orders were carefully compared. Although RECORD was initially developed to order extremely large numbers of loci (>500) (Isidore *et al.* 2003), it is recommended to use RECORD (with a cost function that minimizes recombination events) as a tool complementary to JoinMap (with a cost function that is based on pair wise distance estimates). This map will be of great value for the mapping of QTL loci involved in resistance and quality traits. The existence of regions without markers, is not disturbing the analysis of QTL, because putative QTL in those regions will not segregate either.

Acknowledgements

We are very grateful to dr.Steve Knapp (Oregon State University) for providing the VxE population. We gratefully acknowledge the Dutch breeding campanies Cebeco Seeds, ‘Van de Bilt Zaden en Vlas’ and Dobbelaar Breeding and the Dutch provinces Zeeland for their financial contribution.

References

Alonso Blanco, C., Peeters, A.J.M., Koornneef, M., Lister, C., Dean, C., van den Bosch, N., Pot, J. and Kuiper, M.T.R. (1998) Development of an AFLP based linkage map of *Ler*, *Col* and *Cvi Arabidopsis thaliana* ecotypes and construction of a *Ler*/*Cvi* recombinant inbred line population. *Plant Journal* 14: 259-271

Holloway, J.L. and Knapp, S.J. (1994). G-MENDEL 3.0: A tool for constructing genetic linkage maps and analyzing Medelian phenotypes. Department of Crop and Soil Science, Oregon State University.

Isidore, E., van Os, H., Andrzejewski, S., Bakker, J., Barrena, I., Bryan, G.J., Caromel, B., van Eck, H., Ghareeb, B., de Jong, W., van Koert, P., Lefebvre, V., Milbourne, D., Ritter, E., van der Voort, J.R., Rousselle Bourgeois, F., van Vliet, J. and Waugh, R. (2003) Toward a marker-dense meiotic map of the potato genome: lessons from linkage group I. *Genetics* 165: 2107-2116

Jansen, J., de Jong, A.G. and van Ooijen, J.W. (2001) Constructing dense genetic linkage maps. *Theoretical and Applied Genetics* 102: 1113-1122

Jones, C.J., Edwards, K.J., Castaglione, S., Winfield, M.O., Sala, F., van de Wiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevschi, A., Marmiroli, N., Aert, R., Volckaert, G., Rueda, J., Linacero, R., Vazquez, A. and Karp, A. (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding* 3: 381-390

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

Lincoln, S., Daly, M. and Lander, E. (1992). Constructing Genetic Maps with MAPMAKER/EXP 3.0. Whitehead Institute Technical Report. 3rd edition.

Liu, B.H. and Knapp, S.J. (1990) GMENDEL: a program for Mendelian segregation and linkage analysis of individual or multiple progeny populations using log-likelihood ratios. *Journal of Heredity* 81: 407

Nilsson, N.O., Sall, T. and Bengtsson, B.O. (1993) Chiasma and recombination data in plants - are they compatible. *Trends in Genetics* 9: 344-348

Oh, T.J., Gorman, M. and Cullis, C.A. (2000) RFLP and RAPD mapping in flax (*Linum usitatissimum*). *Theoretical and Applied Genetics* 101: 590-593

Rowland, G.G. (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). *Canadian Journal of Plant Science* 71: 393-396

Spielmeyer, W., Green, A.G., Bittisnich, D., Mendhem, N. and Lagudah, E.S. (1998) Identification of quantitative trait loci contributing to *Fusarium* wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). *Theoretical and Applied Genetics* 97: 633-641

Stam, P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant Journal* 3: 739-744

Thompson, E.A. (1987) Crossover counts and likelihood in multipoint linkage analysis. *Journal of Mathematics Applied in Medicine and Biology* 4: 93-108

Van der Beek, J.G., Verkerk, R., Zabel, P. and Lindhout, P. (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: *Cf9* (resistance to *Cladosporium fulvum*) on chromosome 1. *Theoretical and Applied Genetics* 84: 1-2

Van Ooijen, J.W. and Voorrips, R.E. (2001). JOINMAP® Version 3.0: Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands

Van Os, H., Stam, P., Visser, R.G. F. and van Eck, H.J. (2005) RECORD: a novel method for ordering loci on a genetic linkage map. *Theoretical and Applied Genetics* 112: 30-40

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414

Chapter 6

QTL mapping of resistance and quality traits in Flax (*Linum usitatissimum* L.).

Jaap Vromans, Hein de Jong, Peter Keijzer, Raimon Laan, Piet Stam and Herman J. van Eck

Abstract

To unravel the inheritance of 18 quantitative and qualitative traits in flax, QTL analysis was performed on previously developed linkage maps. Trait values were recorded in two mapping populations, in two years at several locations. Disease resistance was scored at two points in time (first and second observation). The inheritance of four traits is discussed in more detail, as they are important in fiber flax and linseed breeding. A total of 63 QTL positions were identified with varying LOD values and explained variances. For flax scorch three QTLs were detected, of which the major was consistent across locations and years, but the minor QTLs were year and location specific. Both minor QTLs have an additive effect to the major QTL. The genetic model of *Fusarium* resistance is more complex. Three QTLs in the BxH population were consistent across the first and the second observation, but two QTLs were specific to the first or the second observation, respectively. At the second observation, two minor QTLs have an additive effect and the remaining major and minor QTL showed epistatic interaction. At four loci the favorable QTL allele originated from the resistant Hermes parent, but one favorable QTL allele originated from the susceptible parent cultivar Belinka. For linoleic and linolenic acid content of the seed oil two QTLs were identified; this was expected since the E1747 parent was known to be a double mutant for two desaturase genes that are involved in the proportion of linolenic and linoleic acid. Furthermore, this study showed that fiber content is a complex trait but accumulation of favorable alleles with additive and duplicate gene effect will result in high fiber content values. For the remaining traits co-localization of QTLs is discussed in terms of pleiotropic effects of single loci.

Introduction

Fiber flax and linseed are two distinct cultivar groups, grown for the production of fiber and linseed oil respectively. Some breeding goals are important for both cultivar groups, for example pathogen resistance and resistance to lodging. Fiber content, and resistance to *Fusarium* and scorch are the most important breeding goals for fiber flax (Fouilloux 1988), whereas seed yield, fatty acid composition and resistance to rust (*Melampsora lini*) and *Fusarium* are important for linseed breeding (Rowland 1998).

Flax scorch is a complex disease of which the identity of the causal pathogen(s) has been controversial since the beginning of the 20th century (Cariou *et al.* 2003; Delon and Kiffer 1978; Maddens 1976; Wiersema 1955). Resistant cultivars do exist, but the inheritance of the resistance is still unknown. Chaboche and Fouilloux (1996) proposed a di-genic model of complementary genes, both required for *Fusarium* resistance. According to Chaboche and Fouilloux (1996) cultivar Culbert is fully resistant and has both resistance genes (A and B), cultivar Tapa Parana carried two recessive genes (d and e) conferring full susceptibility, and a fifth suppressive recessive gene (called c) was detected in cultivar Ocean. These results were confirmed by Spielmeyer et al (1999) who mapped two highly significant QTLs, contributing 38% and 26% of the phenotypic variance, as well as a strong interaction. The mode of interaction and the amount of variance explained by the interaction was not published. Furthermore, the nomenclature of their linkage groups and markers loci does not easily allow the transfer of this information to other studies.

Little is known about the inheritance fiber content, but the high observed heritabilities suggest that selection for fiber content should be effective (Fouilloux 1988). Furthermore it is known that both additive and dominant effects of genes are involved in the heredity of fiber content and both effects are influenced by environmental conditions (Popescu *et al.* 1998). However it is still uncertain how many genes are involved in the heredity of fiber content.

The inheritance of the fatty acid composition is understood much better. Independently, both Green (1986) and Rowland (1991) developed EMS-induced low-linolenic-acid mutants of flax. They could identify two independent fatty acid desaturase (FAD) genes with additive gene action. Later, both FAD genes were cloned by Fofana *et al.* (2004), and they observed that both genes had partial overlapping but non-identical sequences. Earlier, a stearoyl-acyl carrier protein desaturase (SAD) had been cloned. The function of the gene is to introduce a double bond between carbons 9 and 10 of the stearoyl-acyl carrier protein. This results in the mono-unsaturated oleic acid (Surinder *et al.* 1994).

Since the development of molecular genetic markers, marker assisted selection (MAS) has become an important supportive tool in plant breeding. Trait loci with a simple Mendelian inheritance can be localized on a genetic linkage map easily, where the flanking

DNA markers can be exploited for MAS. For quantitative traits however, without a direct relation between trait value and genotype, the DNA markers for MAS are identified by QTL analysis.

So far two genetic linkage maps in flax have been published (Oh *et al.* 2000; Spielmeyer *et al.* 1998). Oh *et al.* (2000) used 94 RFLP and RAPD loci to construct 15 linkage groups, without trait loci. Spielmeyer *et al.* constructed 18 linkage groups with 213 AFLP markers to identify loci involved in *Fusarium* resistance (see above). Recently Vromans *et al.* (Chapter 5) constructed an integrated map of flax based on two mapping populations and comprising 502 AFLP markers. In Chapter 5 the methods and resulting maps are described, that were obtained by the combined use of two different software packages to improve the reliability of the integrated linkage map of flax. These maps are used in this study to unravel the inheritance of quantitative and qualitative traits. Furthermore, we describe the hereditary basis of four traits in more detail, as they are important in fiber flax and linseed breeding.

Materials and methods

Plant material

Two mapping populations, VxE and BxH, were used to detect loci involved in quantitative and qualitative traits in flax. The first mapping population, VxE, comprised 186 F9 recombinant inbred lines (RILs), derived from a cross between the French fiber flax variety Viking and the Canadian linseed variety E1747 by means of single seed decent (SSD). E1747 is an EMS-induced mutant from cultivar McGregor (Rowland 1991) and has a low level of linolenic acid compared to Viking. Furthermore the two genotypes differed in several other quantitative and qualitative traits. A second RIL population, BxH, was also obtained by SSD from a cross between two fiber flax varieties, Belinka and Hermes. The BxH mapping population is composed of 131 F4 lines and 74 F7 lines and is therefore regarded as an F5. These two parents were mainly chosen because of their difference in fiber content, *Fusarium* resistance and scorch resistance. The construction of both linkage maps is described in Chapter 5.

Trait data collection

In total, 18 different traits were observed, of which 12 traits were studied in both mapping populations. The traits, their scale, parental values, and the seed company that contributed to the observations and years are shown in Table 1.

Table 1 List of traits analyzed

Trait name	Trait code	Trait scale	Company[1]	Year of trial	Parental values			
					Viking	E1747	Belinka	Hermes
Fiber content	FC	% of fibers in straw	Ceb/VdB	2003/ 2004	40.6	27.6	32.0	37.3
Fiber yield	FY	Total fiber yield in kg/ha	Ceb	2003	n.a.	n.a.	2310.1	2894.1
Protein content seed	PCS	% of protein in seed	Ceb	2002	22.6	18.1	n.a.[2]	n.a.
Oil content seed	OCS	% of oil in seed	Ceb	2002	40.1	44.8	n.a.	n.a.
16-0 content oil	OP	% of palmitic acid in oil	Ceb	2002	4.6	6.4	n.a.	n.a.
18-0 content oil	OS	% of stearic acid in oil	Ceb	2002	2.8	3.2	n.a.	n.a.
18-1 content oil	OO	% of oleic acid in oil	Ceb	2002	15.4	12.0	n.a.	n.a.
18-2 content oil	OLE	% of linoleic acid in oil	Ceb	2002	19.6	74.9	n.a.	n.a.
18-3 content oil	OLN	% of linolenic acid in oil	Ceb	2002	57.6	3.5	n.a.	n.a.
Seed yield	SY	Total seed yield in kg/ha	Ceb	2002/2004	n.a.	n.a.	1650.0	1349.0
Thousand grain weight	TGW	Thousand grain weight in g	Ceb	2002	6.3	4.8	5.5	5.9
Scorch resistance	SR	1=resistance, 9=susceptible	Ceb/VdB	2003/2004	n.a.	n.a.	1.0	8.2
Fusarium resistance 1st observation	FR1	1=resistance, 9=susceptible	Ceb	2003	1.9	5.0	4.3	1.5
Fusarium resistance 2nd observation	FR2	1=resistance, 9=susceptible	Ceb	2003	n.a.	n.a.	6.6	2.7
Plant length 2002	PL1	Plant length in cm	Ceb	2002	n.a.	n.a.	81.5	88.5
Plant length 2003	PL2	Plant length in cm	Ceb	2003	n.a.	n.a.	89.9	92.3
Flowering time	FT	Number of days that plants start to flower aftersowing	Ceb	2003	n.a.	n.a.	5.4	7.1
Earliness ripening stem	ERS	1=early, 9=late	Ceb	2003	n.a.	n.a.	4.7	4.8
Lodging 1st observation	L1	1=resistance, 9=susceptible	Ceb/VdB	2003	n.a.	n.a.	5.3	4.8
Lodging 2nd observation	L2	1=resistance, 9=susceptible	Ceb	2003	3.4	1.8	3.8	4.7
Lodging 3rd observation	L3	1=resistance, 9=susceptible	Ceb	2003	n.a.	n.a.	3.8	5.6
General impression	GI	1=ideal, 9=bad impression	Ceb	2003	n.a.	n.a.	5.3	6.5

[1] Ceb = Cebeco and VdB = Van De Bilt Zaden en Vlas

[2] n.a. = not available

In a field trial for, *Fusarium* in both populations, scorch in the BxH population and both fiber content and lodging in the VxE population, seeds of the RILs were sown in a randomized block design with two replications. This allowed the estimation of broad sense heritabilities (h^2) for these traits, which was calculated by:

$$h^2=V_g/(V_g +V_e).$$

Vg represents the genetic variation and Ve the environmental variation; both values could be obtained from an ANOVA on the field data. The remaining traits were observed on the basis of a trial field without replication. Some of the traits were observed in different years, by the breeders of “Cebeco” and “Van De Bilt Zaden en Vlas” in the Dutch agricultural areas Flevo Polders and Zeeuws Flanders, respectively. The trait values of the two replications were averaged and used for further QTL analysis. Only scorch resistance and plant height of the BxH population was determined in two consecutive years. Fiber content and seed yield of the BxH population was measured in 2004 whereas fiber content and seed yield of the VxE population was measured different years, 2002 and 2003 respectively.

QTL analysis

The software program MapQTL 5.0 (Van Ooijen 2003) was used to detect QTLs with both the Interval Mapping method (Lander and Botstein 1989) and the MQM Mapping method (Jansen 1994) for all traits. Putative heterozygosity in the BxH population will hardly affect QTL estimates, because 59 percent of the AFLP markers could be scored as co-dominant markers, either because alternating allelic bands could be observed in the same fingerprint (length polymorphisms) or because of clear band intensity polymorphisms (See Chapter 5).

First interval mapping was performed to detect QTLs. During the next step the automatic cofactor selection option of MapQTL was used to identify co-factors. This was followed by MQM mapping of the same trait with the selected cofactor(s) to identify other QTLs. Then if necessary, new cofactors were determined and followed by another round of MQM mapping. This was done till no new QTLs could be found. This approach has two consequences. First, the estimated percentage of explained phenotypic variance will not be overestimated, because these co-factors will absorb variance components. Second, the smaller but specifically the QTLs with epistatic interactions, which may not have a significant effect on their own, can be detected as significant when the co-factor is placed at a marker locus near the other member in the interaction. The LOD-profiles and the percentage of explained phenotypic variance as shown in the results section were obtained with the MQM approach with cofactors on other linkage groups nearby the other QTL(s) if more then one QTL was detected. In case only one QTL is present the LOD-profile of the Interval Mapping procedure is shown. For declaring a significant QTL a LOD threshold of 3.0 was applied (Van Ooijen 1999). LOD profiles are graphically displayed, including the LOD minus 1 / LOD minus 2 support interval to approximate a 95 percent

confidence interval (Van Ooijen 1992).

For both, *Fusarium* resistance in the BxH population and fiber content in the VxE population the genetic model was described by general linear regression analysis for the observed QTLs in MapQTL. Markers with the highest LOD-value in the MQM analysis were used to describe the QTLs in the model. With ANOVA the main effects and interactions of the genetic model could be screened for their significance. Both general linear model analysis and ANOVA were performed in GenStat 8.11. Correlations between traits were used to describe the interaction between co-localizing QTLs of different traits. QTL nomenclature is based on the trait name (for example; **S**corch **R**esistance=SR), followed by the population name (VxE or BxH), linkage group number (for example; **12**.1) and consecutive QTL number for the trait on that specific linkage group (12.**1**).

Calculation of the expected gain obtained by marker assisted selection on the trait value

Of *Fusarium* resistance, scorch resistance, fiber content and fatty acid composition, the expected gain obtained by marker-assisted selection was estimated by averaging the trait values for the different genotypic classes. For example, in the case of only one QTL, the average trait value was calculated for those offspring genotypes with either the maternal or paternal genotype at the marker locus nearest to the QTL peak. In the case of two QTLs, four possible genotypic classes were regarded. As the possible number of classes grows exponentially with the number of QTLs, we only calculated the average additive effect of allele substitutions in the case of three and four QTLs. In this way the effect of pyramiding QTLs could be approached. In all cases (except for the fiber content of the BxH population) putative heterozygosity at QTL and marker loci of the RILs was ignored, which is a necessity in case of a dominant marker system. The estimated genotypic value might be slightly biased, due to a small percentage of heterozygosity in the RILs of the BxH population.

Results

General overview of QTL analysis

For 18 different traits a total of 60 QTL positions were identified with a LOD value exceeding a threshold of 3.0 (Table 2, Figure 1). Another three QTLs with a LOD value of 2.9 have been included, because of circumstantial evidence for their significance. This evidence is based on the presence of a QTL at a homologous position in the other mapping population (*Fusarium* resistance FR1-VxE_13.1 and FR1-BxH_1.1), confirmation across locations (scorch resistance ScR-BxH_4.1) and QTLs for related traits in fatty acid metabolism (PCO-VxE_6.1, LeCO-VxE_6.1 and LnCO-VxE_6.1).

The LOD values were variable, suggesting minor and major QTLs. Some of the minor QTLs were specific to only one year or location. Some of the major QTLs have such high LOD values and percentages of explained phenotypic variance that a Mendelian classification is almost feasible. However, we consider philosophical discussions on the

boundary between quantitative and Mendelian genetics beyond the scope of this article. In the following paragraphs we wish to describe results on the inheritance of resistance to scorch and *Fusarium*, as well as fiber content and fatty acid composition of linseed oil.

Table 2 observed QTLs

Trait	Name QTL	Integrated linkage group	Favorable genotype*	Lod value	% explained variance
Fiber content	FC-VxE_9.1	Integrated-A	Viking	14.6	26.6
	FC-VxE_6.1	Integrated-C	Viking	4.6	6.8
	FC-VxE_1.1	Integrated-J	Viking	4.3	5.5
	FC-VxE_17.1	Integrated-O	Viking	7.4	12.4
	FC-BxH_8.1	Integrated-A	Hermes	22.9	57
Fiber yield	FY-BxH_8.1	Integrated-A	Hermes	6.6	18
	FY-BxH_11.1	Integrated-D	Belinka	3.4	7.7
Protein content seed	PCS-VxE_9.1	Integrated-A	Viking	8.4	19.1
	PCS-VxE_6.1	Integrated-C	E1747	4.3	8.2
	PCS-VxE_13.1	Integrated-E	E1747	3.5	7.2
	PCS-VxE_16.1	Integrated-I	Viking	3.3	7.2
Oil content seed	OCS-VxE_9.1	Integrated-A	E1747	9.9	20.3
	OCS-VxE_6.1	Integrated-C	Viking	3.1	5.7
	OCS-VxE_16.1	Integrated-I	E1747	4.9	10.6
palmitic acid content seed	OP-VxE_6.1	Integrated-C	E1747	2.9	4.4
	OP-VxE_13.1	Integrated-E	E1747	17.8	34.6
	OP-VxE_15.1	Integrated-H	E1747	4.3	6.9
	OP-VxE_19.1	Integrated-U	Viking	3.0	5.9
stearic acid content seed	OS-VxE_2.1	Integrated-F	E1747	3.8	6.5
	OS-VxE_14.1	Integrated-N	E1747	19.4	46.2
oleic acid content seed	OO-VxE_19.1	Integrated-U	Viking	4.9	15.3
linoleic acid content seed	OLE-VxE_6.1	Integrated-C	E1747	33.1	47
	OLE-VxE_13.1	Integrated-E	E1747	30.6	47.5
linolenic acid content seed	OLN-VxE_6.1	Integrated-C	Viking	36.1	45.7
	OLN-VxE_13.1	Integrated-E	Viking	31.5	48
Seed yield	SY-VxE_11.1	Integrated-G	Viking	4.7	14.6
	SY-BxH_7.1	Integrated-G	Hermes	3.3	8.1
	SY-BxH_6.1	Integrated-H	Belinka	6.0	14.5
Thousand grain weight	TGW-VxE_11.1	Integrated-G	Viking	3.0	7.8
	TGW-VxE_19.1	Integrated-U	Viking	4.0	9.1
	TGW-BxH_12.1	Integrated-C	Belinka	3.1	7.0
	TGW-BxH_1.1	Integrated-E	Hermes	3.2	7.5
	TGW-BxH_7.1	Integrated-G	Belinka	3.4	5.5
Scorch resistance	SR-BxH_12.1	Integrated-C	Belinka	32	62
	SR-BxH_2.1	Integrated-J	Belinka	5.7	6.3
	SR-BxH_4.1	Integrated-K	Belinka	2.9	4.6
Fusarium resistance	FR1-VxE_13.1 = FR2-VxE_13.1	Integrated-E	E1747	2.9	7.8
	FR1-BxH_1.1 = FR2-BxH_1.1	Integrated-E	Hermes	4.4	6.5
	FR1-BxH_10.1 = FR2-BxH_10.1	Integrated-I	Hermes	13.5	24
	FR2-BxH_17.1	Integrated-S	Hermes	3.2	4.2
	FR1-BxH_3.1 = FR2-BxH_3.1	Integrated-T	Belinka	4.5	7.0
	FR1-BxH_16.1	Integrated-U	Hermes	3.6	6.9
Plant length	PL1-BxH_8.1	Integrated-A	Belinka	5.3	14.3
	PL1-BxH_5.1	Integrated-B	Hermes	5.1	10.9
	PL1-BxH_12.1	Integrated-C	Hermes	4.1	9.5
	PL2-BxH_8.1	Integrated-A	Belinka	8.0	20.3
	PL2-BxH_5.1	Integrated-B	Hermes	4.1	8.3
	PL2-BxH_6.1	Integrated-H	Hermes	3.1	6.4
	PL2-BxH_18.1	Integrated-L	Hermes	3.4	7.6
	PL2-BxH_19.1	Integrated-Q	Hermes	4.2	9.6
Flowering time	FT-BxH_5.1	Integrated-B	Hermes	10.5	33.5
Earliness (ripening stem)	ERS-BxH_12.1	Integrated-C	Belinka	3.7	9.5
	ERS-BxH_6.1	Integrated-H	Belinka	3.5	8.1
Lodging	L1-VxE_11.1	Integrated-G	Viking	6.0	13.9
	L1-VxE_12.1	Integrated-K	Viking	3.2	7.6
	L2-BxH_12.1	Integrated-C	Hermes	4.9	10.7
	L1-BxH_12.1	Integrated-C	Hermes	3.5	8.9
	L2-BxH_13.1	Integrated-F	Hermes	3.5	7.3
	L2-BxH_7.1	Integrated-G	Belinka	3.1	5.6
	L2-BxH_2.1	Integrated-J	Hermes	6.8	14.2
	L1-BxH_2.1	Integrated-J	Hermes	4.7	11.9
	L3-BxH_2.1	Integrated-J	Hermes	7.3	19.2
General impression	GI-BxH_2.1	Integrated-J	Hermes	6.6	17

* Parent contributing the favorable allele

Scorch resistance

High heritabilities were observed for scorch resistance in the BxH population for both years and location. Small differences occurred between the heritabilities calculated from scorch evaluation done by the breeders of “Cebeco” and “Van De Bilt zaden en Vlas”, 0.92 (2003) and 0.92 (2004) versus 0.68 (2003) and 0.83 (2004), respectively. On the basis of field experiments in 2003 and 2004 in Flevo Polders, and 2003 in Zeeuws Flanders one major and two minor QTLs were detected (Table 2, Figure 1). The major QTL ScR-BxH_12.1, with a LOD value exceeding 30 and an explained phenotypic variance of 62%, was consistent across locations and years, but the minor QTLs were year and location specific (ScR-BxH_2.1 in 2003 only at Cebeco, Flevo Polders; ScR-BxH_4.1 in 2004 in both locations). For all three QTLs the cultivar Belinka derived allele contributed to scorch resistance (Table 2). Epistatic interaction between the major and minor QTLs was not significant, confirming that both minor QTLs have an effect additive to the major QTL (Figure 2). Although it might be feasible to visually classify the offspring on the presence or absence of the major QTL allele, breeding for a superior level of scorch resistance could require the presence of both minor QTL, whose presence is almost improbable without the use of DNA markers.

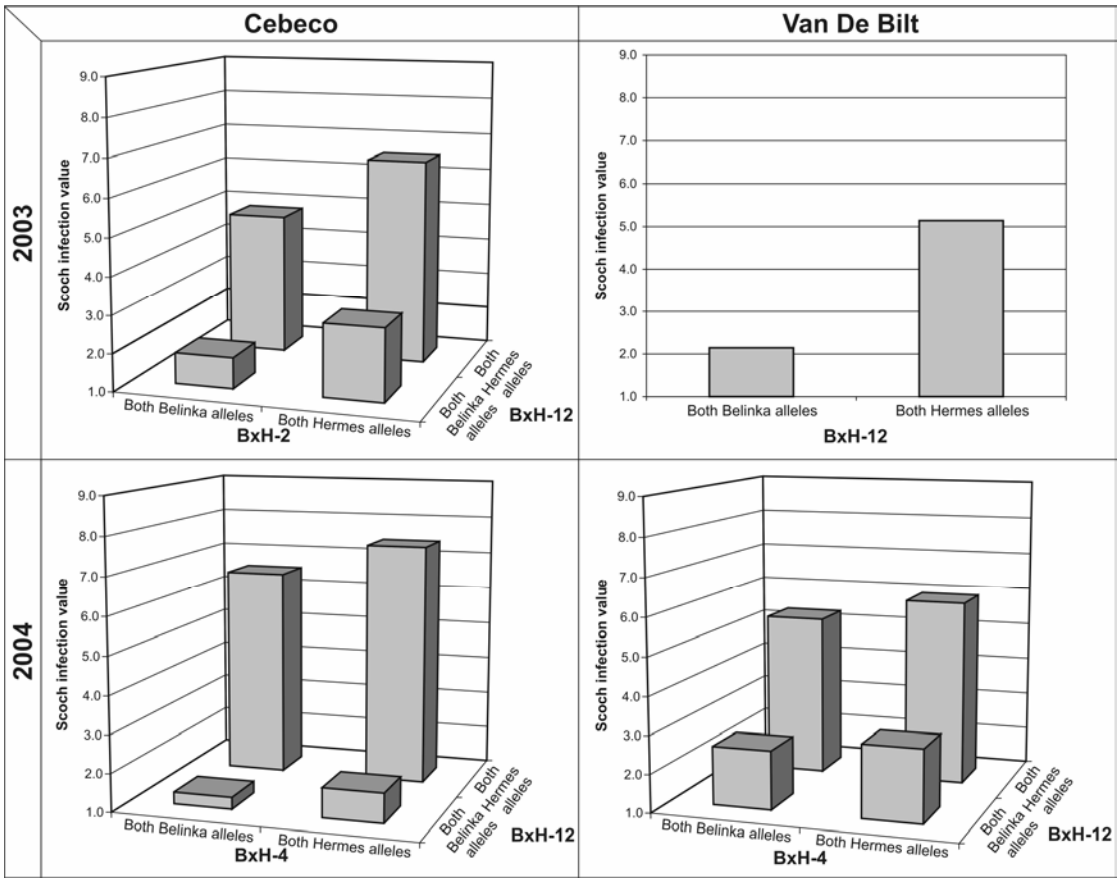

Figure 2 Effect of selection of markers linked to scorch resistance QTLs

Fusarium resistance

The heritabilities for both populations at both time points were moderate to high, ranging from 0.55-0.69. In the BxH population the heritability of *Fusarium* calculated at time one was higher than at time point two (0.69 over 0.56), whereas for the VxE population it was the other way around (0.66 over 0.55). Five QTLs involved in *Fusarium* resistance have been identified in the BxH population and one QTL in the VxE population. The latter QTL FR1-VxE_13.1 and QTL FR1-BxH_1.1 have been localized on linkage group 13 and 1 in VxE and BxH respectively. FR1-VxE_13.1 and FR1-BxH_1.1 could be alleles of the same gene, because the positions of these QTLs coincide on the linkage group E of the integrated map (Figure 1).

The five QTLs of the BxH population differed in significance and explained phenotypic variance. In contrast to four minor QTLs, the locus FR1-BxH_10.1 is a highly significant major QTL with a LOD value of 13.5 and an explained variance of 24% (Table 2). In the BxH mapping population the female parent cultivar Hermes is considered resistant. Indeed, at four loci the favorable QTL allele originated from Hermes, but one QTL allele that contributed to *Fusarium* resistance (FR1-BxH_3.1) originated from the susceptible parent cultivar Belinka (Table 2). This fits well with the observation of transgressive segregation in the BxH mapping population. Three QTLs were consistent across the earlier and the later observation, but two QTL FR1-BxH_16.1 and QTL FR2-BxH_17.1 were specific to the first or the second observation, respectively.

Unfortunately, none of the BxH RILs combined all five QTL alleles that contribute to resistance. Hence, the added value of the Belinka allele in a Hermes background can be inferred only indirectly. With five segregating QTL loci a total of 32 different genotypic classes are expected in the offspring. For further analysis we used phenotypic data of the second observation in 2003. The best linear model for *Fusarium* resistance in the BxH population reads:

$$FR2_{ijkl} = \mu + BxH\_1_i + BxH\_3_j + BxH\_10_k + BxH\_17_l + BxH\_3\times BxH\_10_{(jk)},$$

since it contains significant main and interaction effects, only. ANOVA showed that all four loci made a significant contribution to the level of *Fusarium* resistance (Table 3). Prediction of the means from the regression model for the offspring genotypic classes at the minor QTLs FR2-BxH_1.1 and FR2-BxH_17, showed that the Hermes derived alleles 1 each contributed 0.38 and 0.35 scale units to *Fusarium* resistance, respectively. The minor QTL FR2-BxH_3.1 and the major QTL FR2-BxH_10.1 showed significant epistatic interaction (Table 3).

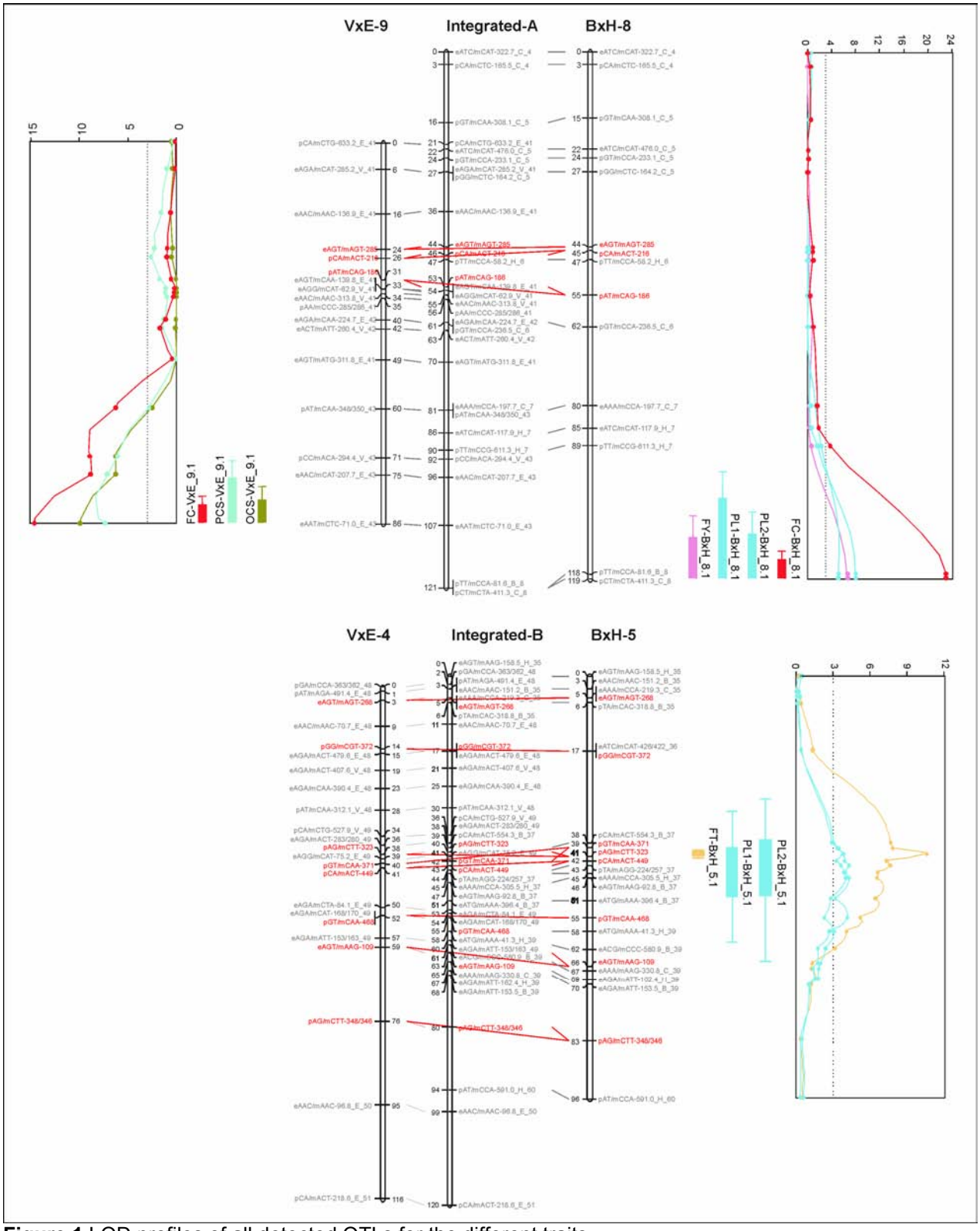

Figure 1 LOD profiles of all detected QTLs for the different traits

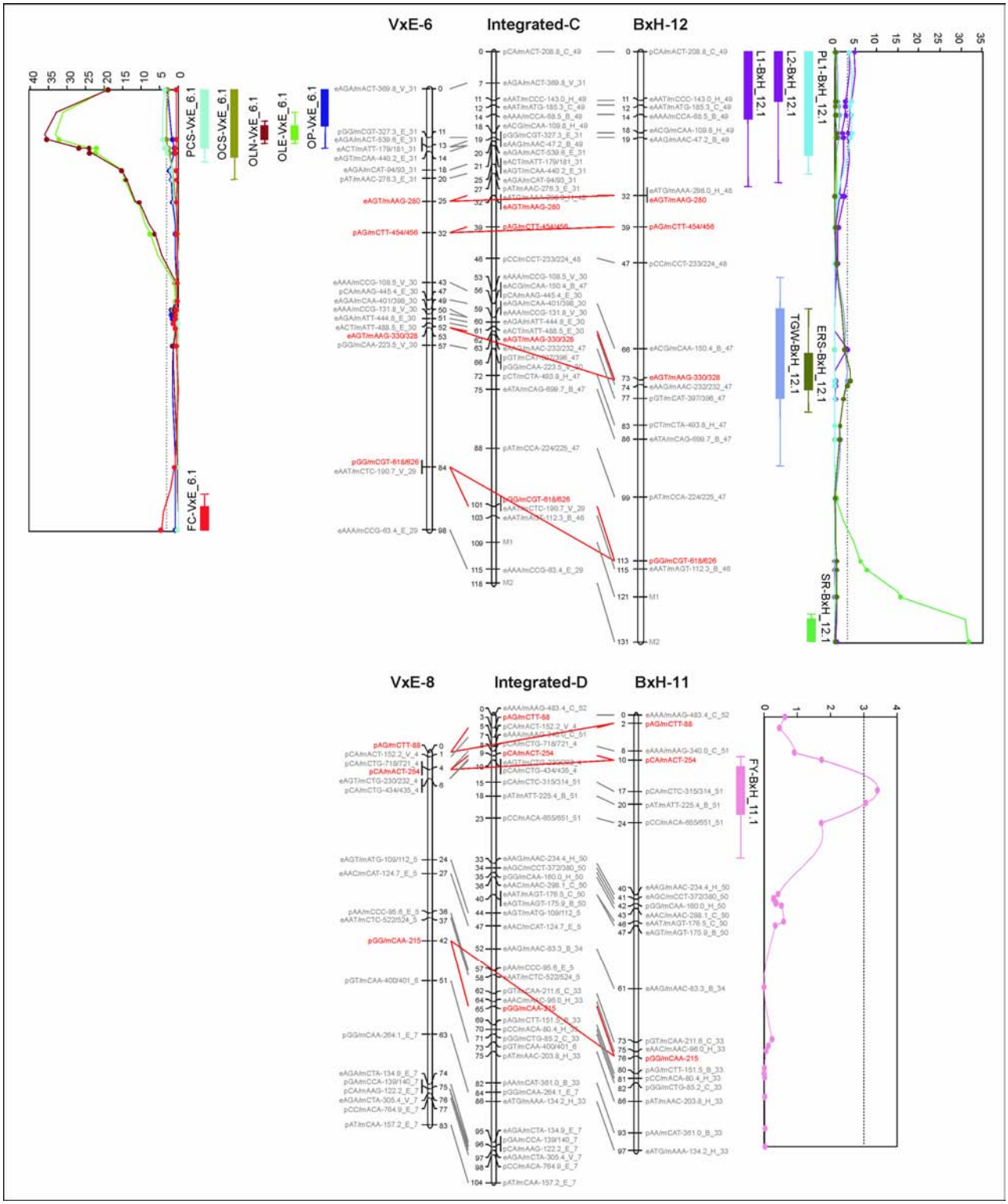

Figure 1 continued

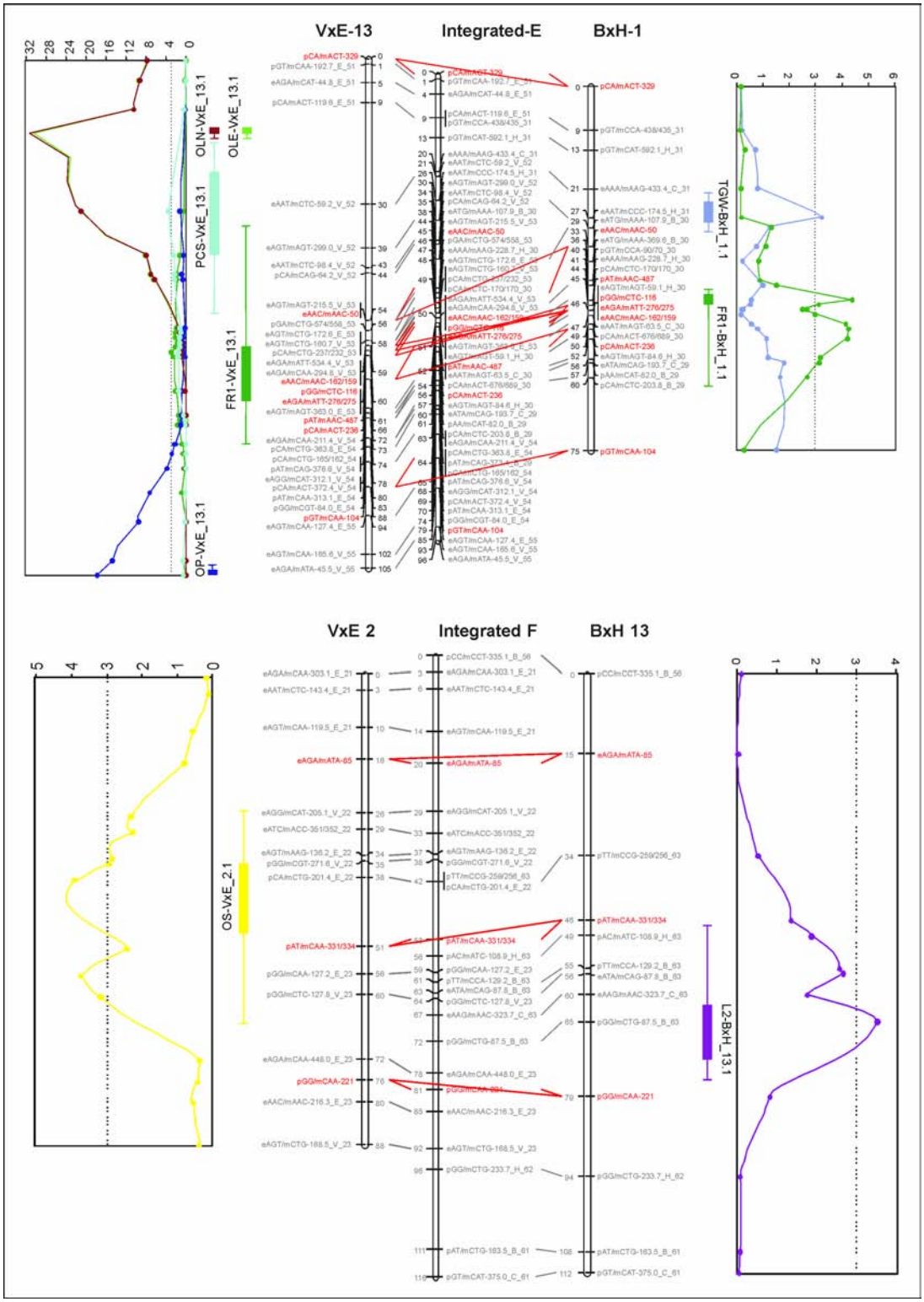

Figure 1 continued

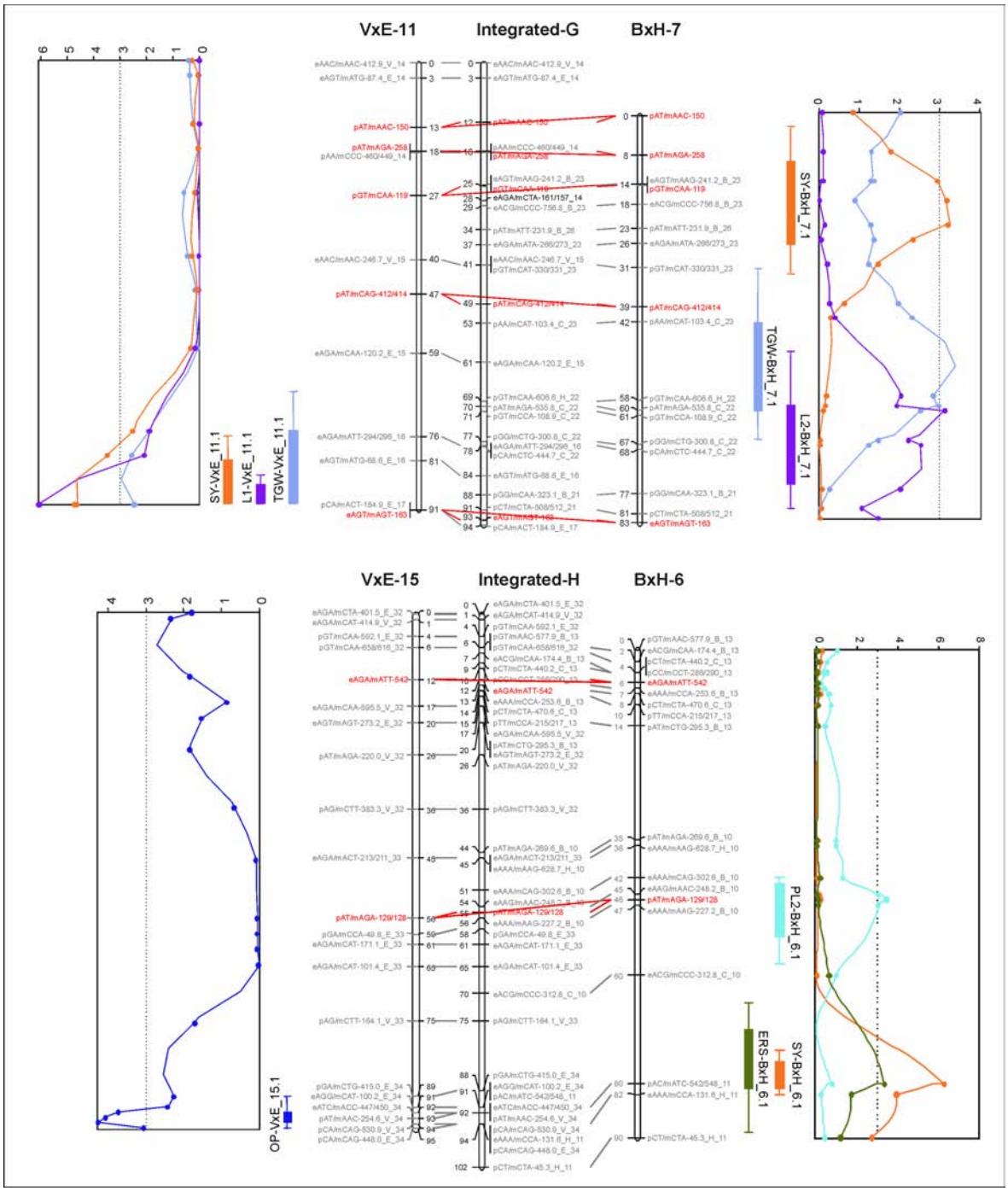

Figure 1 continued

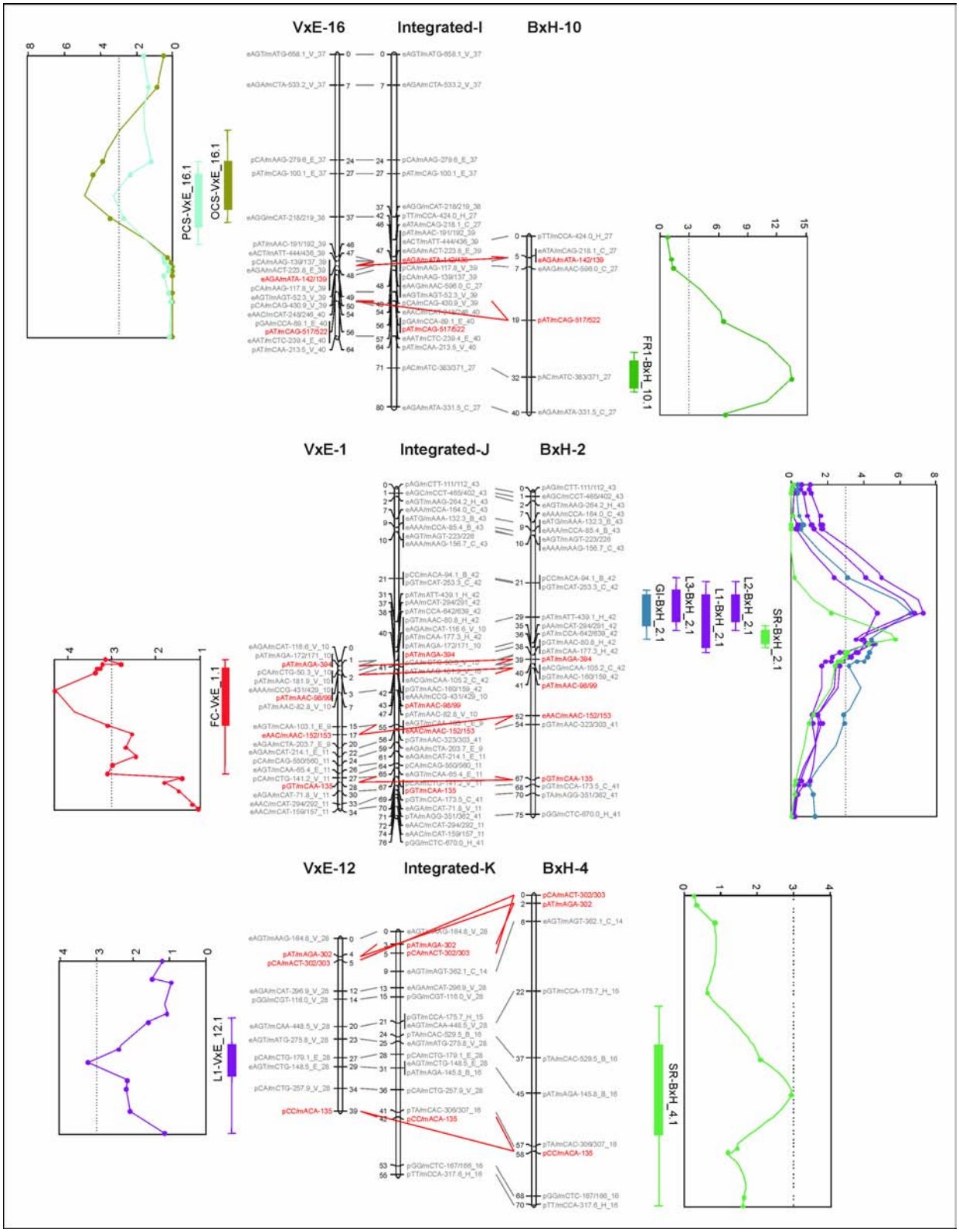

Figure 1 continued

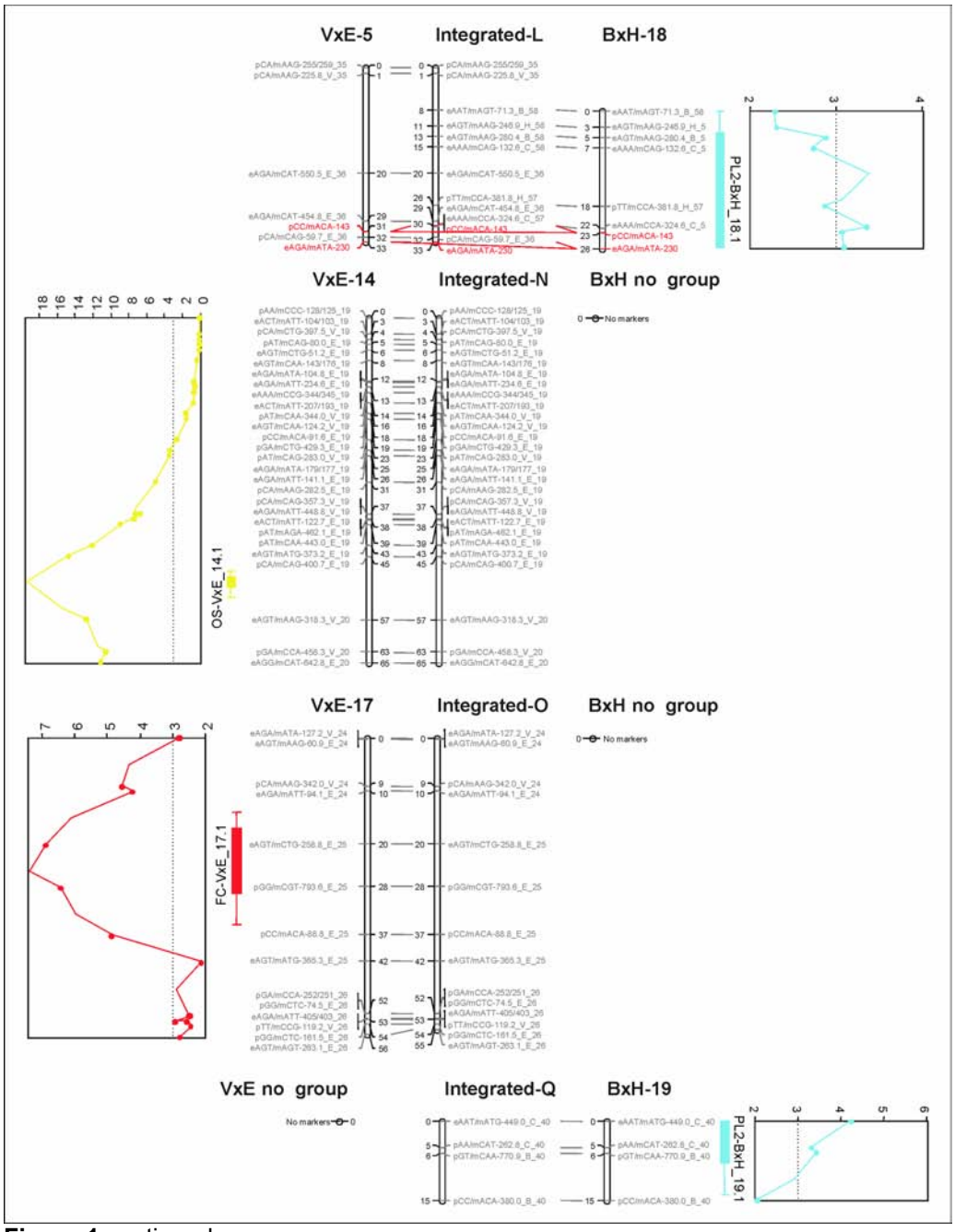

Figure 1 continued

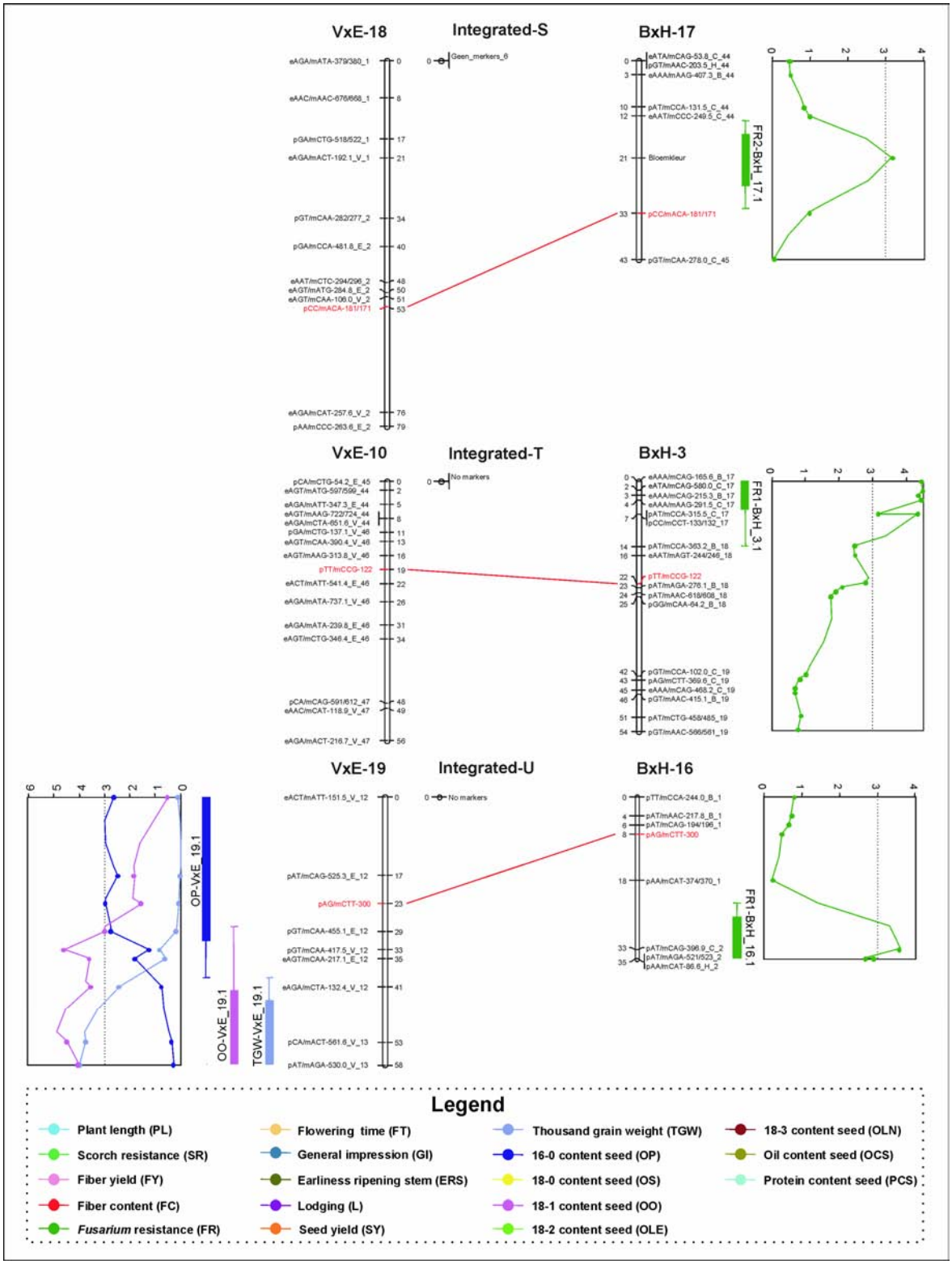

Figure 1 continued

Table 3 ANOVA for *Fusarium* QTLs

Change	d.f.	s.s.	m.s.	v.r.	F pr.
FR2-BxH_1.1	2	28.01	14.00	10.4	<0.001
FR2-BxH_3.1	2	20.86	10.43	7.74	<0.001
FR2-BxH_10.1	2	73.84	36.92	27.41	<0.001
FR2-BxH_17.1	2	18.98	9.49	7.04	0.001
FR2-BxH_3.1* FR2-BxH_10.1	4	14.76	3.69	2.74	0.031
Residual	140	188.57	1.35		
Total	152	345.00	2.27		

The heterozygous genotypic classes could not be significantly distinguished from the homozygous classes, because of large least significant differences of predictions at a 5% level. The large least significant differences of predictions are the result of the low number of heterozygous genotypes in the RIL population. Therefore, the heterozygous classes were ignored for the interpretation of the epistatic interaction. Figure 3 shows the interaction type where the effect of Hermes-derived alleles is boosted by the presence of the Belinka alleles at the other locus. The joint absence or presence of the Hermes alleles at QTL FR2-BxH_10.1 and the Belinka alleles at QTL FR2-BxH_3.1 caused a difference of 2.5 scale units to *Fusarium* resistance.

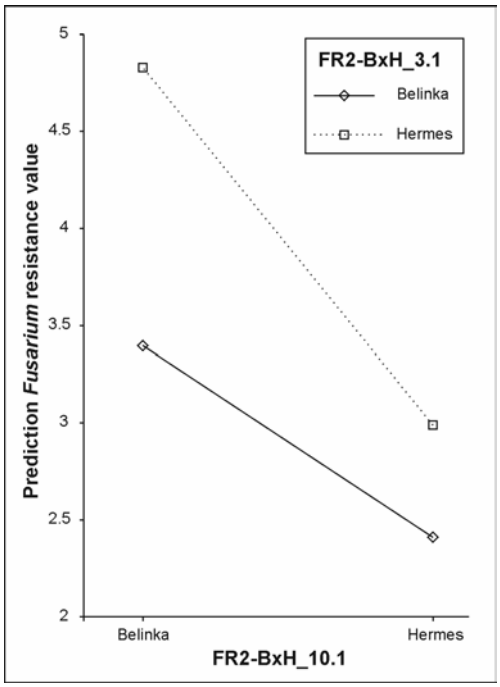

Figure 3 Interaction effect of QTLs FR2-BxH_3.1 and FR2-BxH_10.1 on the level of *Fusarium* resistance

Figure 4a shows transgressive segregation for *Fusarium* resistance and in Figure 4b the level of resistance is shown in relation to the number of QTL alleles that contribute to resistance at time point two. This figure supports the more general conclusion that a higher level of resistance is achieved by pyramiding QTL alleles involved in resistance. Such is only attainable with MAS.

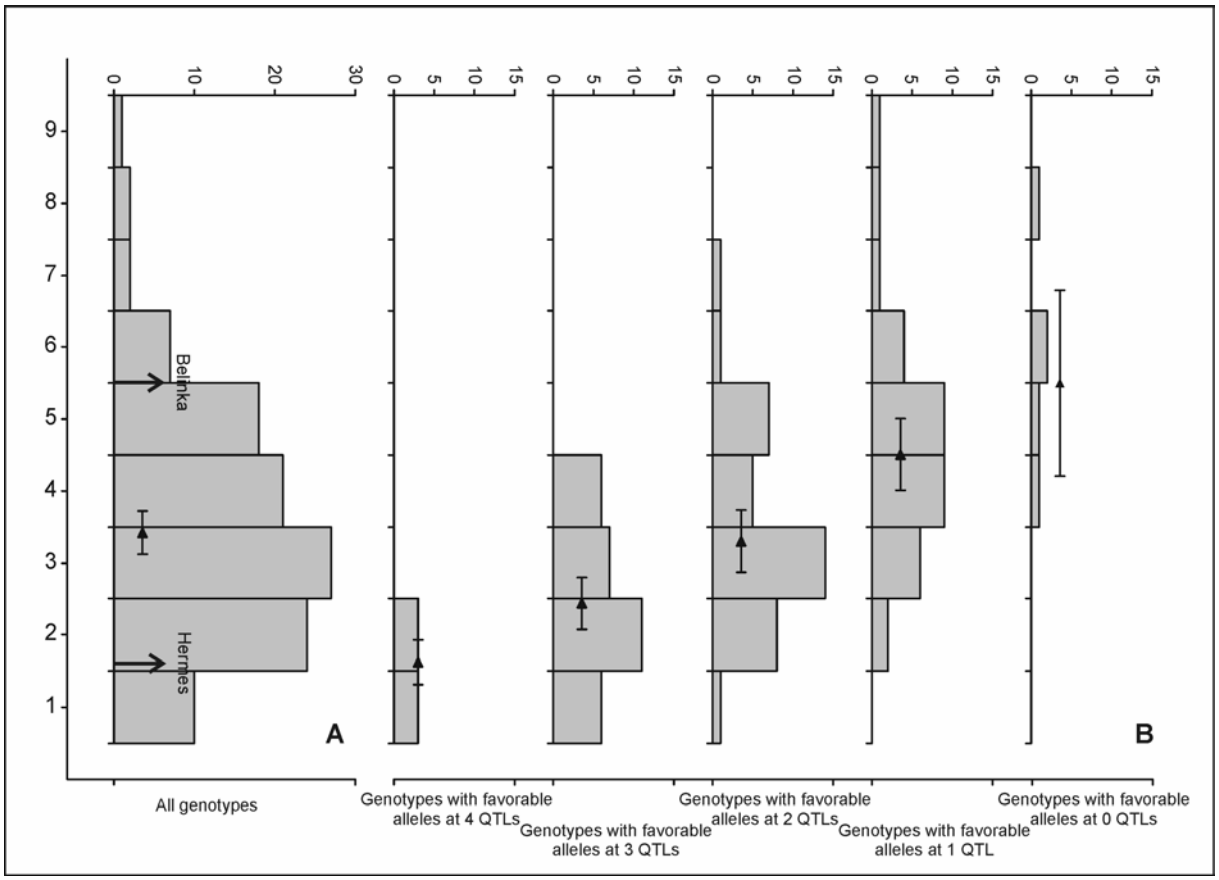

Figure 4 The distribution of *Fusarium* resistance trait values in the BxH population (N=112). The average and the 95% confidence interval of the RIL population are shown. The average parental values over the two time points are plotted as well. **B** The expected gain obtained by MAS is approximated by calculating the average value of resistance caused by increasing number of favorable QTL alleles at time point two. The average trait value and the 95% confidence interval of the RILs of the different genotypic groups is shown

Fatty acid composition of seed oil - linoleic / linolenic acid content

As expected the trait values for linoleic and linolenic acid show a strong negative correlation (r = -0.99) in the RILs of VxE, because both components depend on the same substrate (Figure 5). Although linoleic and linolenic acid QTLs have been shown as separate traits on the VxE map, it is obvious that the minus allele for one trait should be allelic to the favorable allele for the other. The two QTL for both linoleic as linolenic acid (OLE-VxE_6.1/OLN-VxE_6.1 and OLE-VxE_13.1/OLN-VxE_13.1) are highly significant with LOD-values exceeding 30 and the sum of the explained phenotypic variance at both loci reached ~ 90%. This suggests that discrete classification of genotypes is nearly possible. In Figure 5 the expected four genotypic classes are indicated and only the V1V1v2v2 and v1v1V2V2 classes showed overlap, where the symbol V1 and V2 represent the Viking derived dominant desaturase alleles at locus OLE-VxE_6.1/OLN-VxE_6.1 and OLE-VxE_13.1/OLN-VxE_13.1, respectively. Four groups are expected and indeed observed, because heterozygosity is highly improbable in F9 RILs.

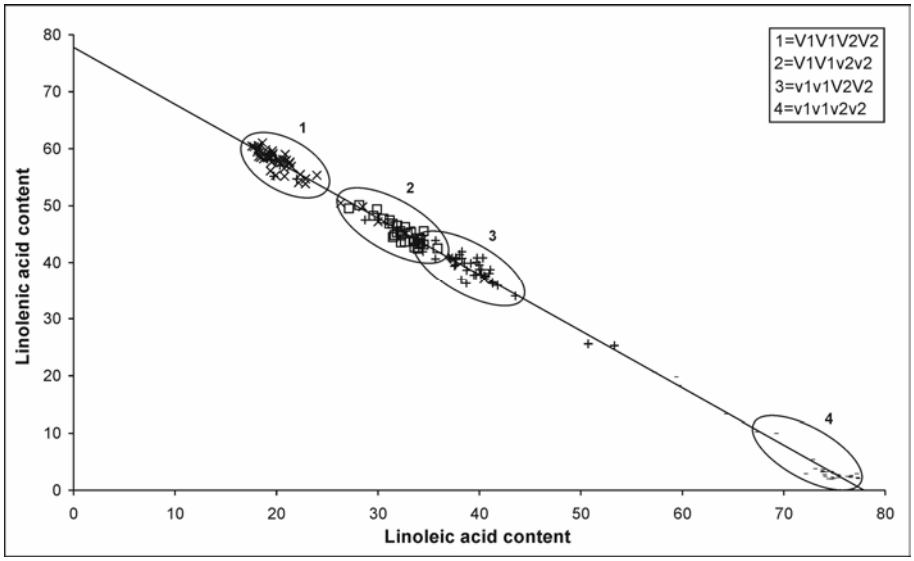

Figure 5 The Linolenic / linoleic acid content of linseed oil of the VxE RILs and their genotypic constitution

The RIL population is highly homozygous and therefore it's expected to observe four genotypic groups, if differences are present in the activity of both desaturase loci. Group 1 contained both intact desaturase genes (resemblance of Viking parent) whereas group 2 had a mutation at locus OLE-VxE_13.1/OLN-VxE_13.1 and group 3 at locus OLE-VxE_6.1/OLN-VxE_6.1. Genotypes in group 4 had the same linoleic/linolenic content as parent E1747 with both desaturase mutations. A few genotypes could be determined in Figure 5 with an intermediate level of linoleic/linolenic content compared to the linoleic/linolenic content of the genotypic groups. Table 4 shows that the presence of one mutated desaturase gene resulted in a higher proportion of linolenic acid content and a lower linoleic acid content than the midparent values (respectively (57.6+3.5)/2=30.6 and (74.9+19.6)/2=47.3). Furthermore it is obvious, from the wild type perspective (which is high linolenic acid content), that the mutation at locus OLE-VxE_6.1/OLN-VxE_6.1 has a larger effect than the mutation at the other locus, because the average values of both genotypic groups are significant different from each other.

Table 4 Allelic composition and trait values of the different group

Genotype	Linolenic acid content	Linoleic acid content	Number of genotypes (N)
Viking	57.6	19.6	2
E1747	3.5	74.9	2
V1V1V2V2	58.0 se 0.27	19.9 se 0.23	45
V1V1v2v2	45.4 se 0.39	32.3 se 0.35	30
v1v1V2V2	39.3 se 0.36	38.9 se 0.35	36
v1v1v2v2	2.8 se 0.14	75.0 se 0.26	27

Fatty acid composition seed oil – remaining components

Besides two major QTLs for linoleic and linolenic acid content, seven other QTLs for fatty acid composition were detected. Of the four palmitic acid content QTLs one co-localized

and with the major QTL OLE-VxE_6.1/OLN-VxE_6.1 (Figure 1). A second palmitic acid content QTL was located at the other end of the linkage group, which contained the other major QTL for linoleic/linolenic acid content (VxE-13, Figure 1). Anther palmitic acid content QTL was located on the same linkage group (VxE-19, Figure 1) as the oleic acid content QTL. The last palmitic acid content QTL and two stearic acid content QTLs were mapped on different linkage groups (Figure 1).

Fiber content

A high heritability (h^2=0.92) was observed for fiber content in the VxE population. Fiber content in the VxE population was at least controlled by four QTLs of which the major QTL FC-VxE_9.1 had the strongest effect with the highest LOD value and explained variance (Table 2). In the BxH population only one a major QTL was observed (FC-BxH_8.1). The position of both the major QTLs coincides on the linkage group A of the integrated map, and therefore FC-VxE_9.1 and FC-BxH_8.1 could be alleles of the same gene (Figure 1).

In total the four VxE QTLs took account for 51.3% of the total phenotypic variation. No transgressive segregation was observed (Figure 6a) which was supported by the Viking origin of the alleles with a positive effect on fiber content (Table 2). ANOVA showed that besides significant gene effects of individual QTL also a significant interaction could be determined (Table 5).

Table 5 ANOVA for QTLs involved in fiber content

Change	d.f.	s.s.	m.s.	v.r.	F pr.
FC-VxE_1.1	1	76.84	76.84	25.78	<0.001
FC-VxE_6.1	1	79.87	79.87	26.79	<0.001
FC-VxE_9.1	1	217.07	217.07	72.82	<0.001
FC-VxE_17.1	1	90.92	90.92	30.50	<0.001
FC-VxE_6.1* FC-VxE_17.1	1	13.29	13.29	4.46	0.036
Residual	149	444.16	2.98		
Total	154	922.15	5.99		

The best linear model for fiber content in the VxE population reads:

$$FC_{ijkl} = \mu + VxE\_1_i + VxE\_6_j + VxE\_9_k + VxE\_17_l + VxE\_6*VxE\_17_{(kl)}.$$

Prediction of the means from the regression model for the offspring genotypic classes at the minor QTL FC-VxE_1.1 and major QTL FC-VxE_9.1 showed that the Viking derived alleles at both QTLs contributed 0.58 and 1.29 to fiber content, respectively. Prediction of the means from the regression model for the offspring genotypic classes at QTLs FC-VxE_6.1 and FC-VxE_17.1 showed an interaction known as duplicate gene action; only if at both QTLs the E1747 alleles are present a reduced fiber content (approximately 2 scale units less) can be observed (data not shown).

In Figure 6b the level of fiber content is shown in relation to the number of QTL alleles that contribute to fiber content. This figure supports the more general conclusion that a higher level of fiber content is achieved by pyramiding QTL alleles involved with fiber content.

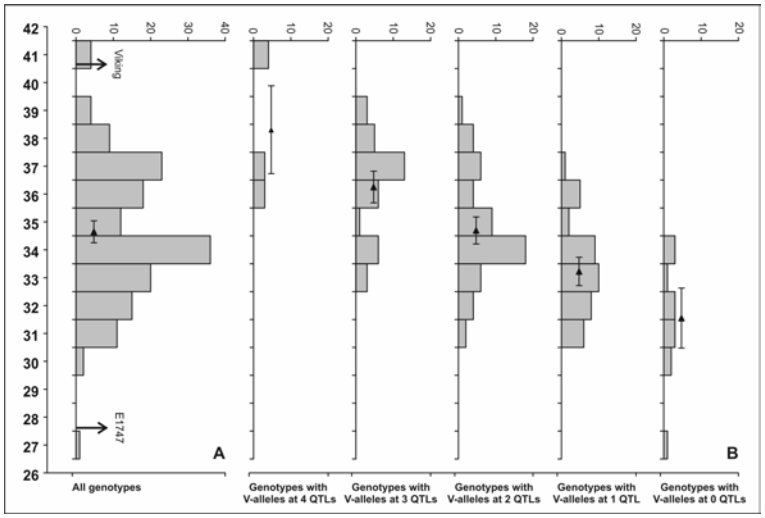

Figure 6 A The population distribution of fiber content of the VxE population (N=156). The average and the accompanying 95% confidence interval of fiber content of the RILs are shown. The average parental fiber content is plotted as well. **B** Effect of selection of markers linked to fiber content QTLs for different genotypic groups with a descending number of favorable fiber content QTLs (all Viking). The average and the accompanying 95% confidence interval of fiber content of the different genotypic groups are shown

The clear difference between fiber content of both parents of the BxH population was caused by only one QTL (Table 2). The BxH fiber content QTL FC-BxH_8.1 had a very high LOD value and accounted for 57% of the phenotypic variation.

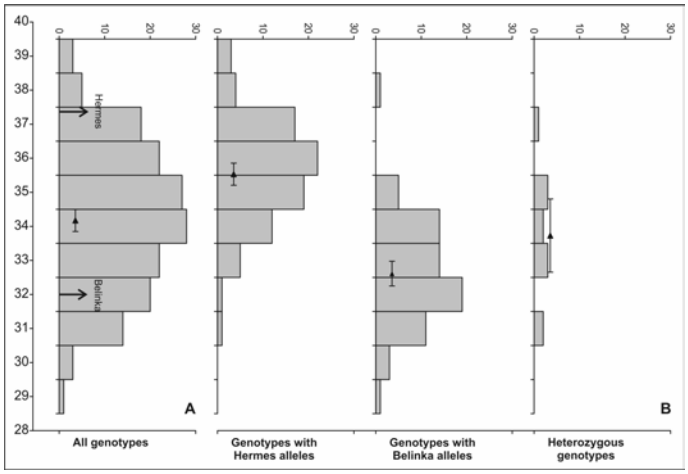

Figure 7 A The population distribution of fiber content of the BxH population (N=163). The average and the accompanying 95% confidence interval of fiber content of the RILs are shown. The average parental fiber content is plotted as well. **B** Effect of selection of the marker linked to fiber content QTL for different genotypic groups. The average and the accompanying 95% confidence interval of fiber content of the different genotypic groups are shown

And as expected the favorable QTL allele originated from Hermes, which had the highest fiber content of the two parents. As a codominant marker was most closely linked to the QTL a heterozygous offspring class could be identified as well. But as expected, just few offspring with heterozygosity are present in the BxH RIL population, which resulted in a large 95% confidence interval (Figure 7). Although the difference between the Belinka alleles and the heterozygous class was not significant, this locus seems to contribute only additive genetic variance.

Miscellaneous traits

This paragraph describes the remaining QTLs, which are listed in Table 2 and mapped in Figure 1. Besides fiber content, fiber yield is important for flax breeding as well. Two QTLs, explaining 25% of the phenotypic variation, were detected of which the major QTL co-localized with both the major QTL of fiber content and plant length (at time point two) on linkage group BxH-8. The minor QTL for fiber yield was mapped on linkage group BxH-11. Five other plant length QTLs were detected across the BxH linkage map, of which one co-localized with flowering time (BxH-5) and one with lodging (BxH-12). Five thousand grain weight QTLs were observed in the two populations across four different integrated linkage groups. Only TGW-VxE_11.1 and TGW-BxH_7.1 were mapped on the same integrated group (Integrated-G) and both co-localized with a lodging and seed yield QTL. Furthermore, TGW-BxH_12.1 co-localized with a QTL for earliness (ripening stem, ERS-BxH_12.1). In the BxH population a second QTL for seed yield (SY-BxH_6.1) and earliness (ripening stem, ERS-BxH_6.1) co-localized at linkage group BxH-6. Three oil and protein content QTLs co-localized at linkage group VxE6, VxE-9 and VxE-16, whereas a forth protein content QTL mapped at VxE-13. Of the total nine observed lodging QTLs, some QTLs were consistent across mapping populations (Integrated-G) or time points (BxH-2 and BxH-12). The remainder localized at different positions. Finally, one general impression QTL was detected at a similar position to the time point consistent lodging QTLs at linkage group BxH-2.

Discussion

General overview

This study confirmed the good quality of the linkage maps (see Chapter 5) as many smooth running LOD profiles could be determined in both mapping populations. Secondly, the fact that for some traits more than two QTLs could be detected, indicated the high accuracy of the observed trait values. Furthermore the QTLs detected in different populations, but located at similar regions, give the opportunity to postulate about the inheritance of these traits. Hence, this study provides good knowledge about the genetics of some traits in flax.

The effects of some QTLs were specific for a given year, location or point in time, which points towards genotype by environment interaction. This phenomenon has previously been reported in many other crop species and has been of great interest of research in plant breeding for the past 15 years.

Inheritance of scorch resistance

The observation of three additive QTLs for scorch resistance is in support with the breeder's perception, but seemingly in disagreement with commonly accepted elicitor-receptor model of plant - pathogen interaction. If the major QTL is an *R*-gene, we still need to account for its incomplete action and the additive value of the two minor QTL. Possibly the phenotypic observations of disease symptoms are imprecise due to the effect of more than one causal pathogen, since *Phytium buismaniae* and *Chalara elegans* have both been proposed as the causal pathogen (Cariou *et al.* 2003; Delon and Kiffer 1978; Wiersema 1955). Transgressive segregation for scorch resistance was not observed, which agrees with the Belinka origin of the more resistant QTL alleles. Possibly all resistance factors of the flax gene pool have been pyramided in Belinka, whereas Wiersema (1955) obtained resistant offspring from crosses between two less resistant parents. Wiersema therefore expected scorch resistance to be controlled by at least two genes. Alternatively, our two mapping populations do not imply a comprehensive search of all genetic factors involved in scorch resistance in the flax genepool.

Inheritance of Fusarium resistance

The inheritance of resistance to *Fusarium* wilt has been investigated by Chaboche and Fouilloux *et al.* (1996) and Spielmeyer et al (1998). Both papers suggest two complementary genes, both required for full resistance. This was not observed in our study, because only one major QTL involved in *Fusarium* resistance has been observed, although this QTL (FR1-BxH_10.1) could be homologous to one of the QTLs described before. The major QTL explained a large part of the phenotypic variation but also showed interaction with the minor QTL FR1-BxH_3.1. This interacting minor QTL might be the complementary gene implied by Chaboche and Fouilloux (1996), but the QTL effect is much smaller than the percentage of explained phenotypic variance published by Spielmeyer *et al.* (1998). The minor QTL FR1-BxH_3.1 is most curious, because the resistance was derived from the susceptible parent cultivar Belinka. This locus might correspond to the gene called c from cultivar Ocean (Chaboche and Fouilloux, 1996). From the transgressive segregation in the BxH population it is obvious, that only the pyramiding of favorable QTLs via marker assisted selection will result in the highest level of resistance.

Inheritance of polyunsaturated fatty acid composition

As expected two QTLs were identified that are involved in the proportion of linolenic and

linoleic acid, because the E1747 parent was known to be a double mutant genotype for the two desaturase genes (Rowland 1991). The strong negative correlation between linoleic and linolenic acid is in close correspondence with the r-values reported by Green (1986) and Rowland (1991). This strong negative correlation suggests that the two QTLs detected for linoleic acid and linolenic acid are allelic. Therefore, we assume that the two loci (OLE-/OLN-VxE_13.1 and OLE-/OLN-VxE_6.1) detected are the two mutated desaturase genes, which were present in the E1747 parent.

We reported a significant difference between the activities of both loci by comparing the two single mutant genotypic classes. This is in agreement with differences observed in sensitivity and activity of the desaturase genes in flax to high temperature during seed maturation (Green 1986). He showed that both mutant genotypes containing mutations in different desaturase genes, but similar genetic background (both mutants of cultivar Glenelg), reacted differently to high temperature (higher than 27 day and 22 night temperature) during seed maturation and had different levels of linoleic and linolenic acid content. This suggest that different genes are involve which was confirmed by Fofana *et al.* (2004) who detected two overlapping but non-identical fatty acid desaturases.

Inheritance of fiber content QTLs

Little is known about the inheritance of fiber content in flax, but high heritabilities have been obtained before (Fouilloux 1988), which is in correspondence with the high heritability obtained in this study. The VxE population revealed four quantitative trait loci involved in fiber content differences observed between the fiber flax cultivar Viking and linseed genotype E1747. For all four QTLs the Viking alleles are associated with high fiber content. It is therefore questionable whether or not linseed cultivars can contribute to fiber flax breeding concerning fiber content. The fiber content detected in the BxH population was located at the same position as the major fiber content QTL of the VxE population. Therefore, it is most likely that the same locus is detected. Although both Viking and Hermes harbor the same major QTL, Hermes has a slightly higher fiber content than Viking, which means that either more loci are involved in fiber content or multiple alleles occur at the same locus. Nevertheless this study showed that fiber content is a complex trait but accumulation of favorable QTLs showed to have an additive effect and will result in high fiber content values.

QTL FC-BxH_8.1 had a larger LOD-value and explained phenotypic variation than co-localizing QTL FC-VxE_9.1. However, the relative contribution of QTL FC-BxH_8.1 to the variance in the BxH population is high compared to QTL FC-VxE_9.1 in the VxE population, because fiber content has a high heritability (this study; Fouilloux 1988). Hence, we assume that both QTLs are allelic.

Related traits

For 18 traits we have identified 63 QTLs, but this is a somewhat redundant list. There are two ways to remove redundancy in our list of QTLs. First, there are observations across

years, locations and mapping populations. Proof of homology of QTLs from different mapping populations cannot be obtained easily with the test for allelism, because the QTLs are localized at similar positions. The second way to remove redundancy in our list of QTLs is based on an interpretation of the trait correlations, because a single QTL could have a pleiotropic effect on two different traits. Several co-localizing QTLs have no obvious relation, and in such cases it is safe to assume that a few centimorgan interval of the flax genome may harbor a large number of linked genes. Below we wish to point out a number of co-localizing QTLs where the trait correlation has an obvious explanation.

The QTLs PL1-BxH_12.1 and L1/L2-BxH_12.1 for plant length and lodging, respectively, are probably a single locus with a pleiotropic effect, as longer plants are more susceptible to lodging. The QTLs FT-BxH_5.1 and PL1/PL2-BxH_5.1 for flowering time and plant length, respectively, are probably a single locus with a pleiotropic effect as well, as late flowering results in taller plants, because of extended vegetative phase. Because lodging susceptible plants tended to have a poor general impression score, both major QTLs for lodging (L1/L2/L3-BxH_2.1) and general impression (GI-BxH_2.1) co-localize. The QTLs SY-VxE_11.1 and TGW-VxE_11.1 for seed yield and thousand grain weight, respectively, are probably a single locus with a pleiotropic effect, as plants with higher thousand grain weight have a higher seed yield.

Both mutated desaturases had pleiotropic effects as well. The mutations decreased linolenic acid content and increased both linoleic acid content and protein content. Furthermore, the mutation at locus V1 on linkage group VxE-6 also influenced palmitic acid content and oil content. These mutants contained lower levels of oil content and higher levels for palmitic acid content. Since oil and protein content show strong negative correlation, it is not surprising that QTLs for these traits co-localized (at linkage groups VxE-9 and VxE-16).

Conclusion

The results presented in this paper offer a comprehensive overview of the genetic basis of many agronomically important traits in flax. In several cases it has been confirmed that implementation of MAS is essential to be able to identify and pyramid all genes that contribute to the trait value. Further research is necessary to clarify the real cause of year and location differences in for example scorch resistance and to be able to understand the interactions between QTLs within a trait and between QTLs across traits.

Acknowledgements

We are very grateful to Marcos Malosetti and Dr. Fred van Eeuwijk for their statistical assistance. We gratefully acknowledge the Dutch breeding companies Cebeco Seeds, ‘Van de Bilt Zaden en Vlas’ and Dobbelaar Breeding and the Dutch provinces Zeeland for their financial contribution.

References

Cariou, E., Brochard, M. and Raynal, G. (2003) La brulure du lin, due a *Chalara elegans* et plusieurs especes de *Pythium*. *Phytoma* 566: 6-9

Chaboche, L. and Fouilloux, G. (1996) Resistance of flax to *Fusarium*. 4th Workshop of the FAO Network on Flax, Rouen, France: 243-246

Delon, R. and Kiffer, E. (1978) *Chalara elegans* (*Thielaviopsis basicola*) and allied species, 1: Generalities and pathology. *Annales du Tabac* 2: 159-186

Fofana, B., Duguid, S. and Cloutier, S. (2004) Cloning of fatty acid biosynthetic genes beta-ketoacyl CoA synthase, fatty acid elongase, stearoyl-ACP desaturase, and fatty acid desaturase and analysis of expression in the early developmental stages of flax (*Linum usitatissimum* L.) seeds. *Plant Science* 166: 1487-1496

Fouilloux, G. (1988) Breeding flax methods. In Flax: Breeding and Utilisation: 14-25

Green, A.G. (1986) Effect of temperature during seed maturation on the oil composition of low-linolenic genotypes of flax. *Crop Science* 26: 961-965

Jansen, R.C. (1994) Controlling the type-I and type-II errors in mapping quantitative trait loci. *Genetics* 138: 871-881

Lander, E.S. and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199

Maddens, K. (1976) Centre for flax research: Results of agronomic investigations 1960-1975. *Fibra* 21: 4-32

Marchal, E. (1900) Recherches biologiques sur une *Chytridinée* parasite du lin. *Bulletin de l'Agriculture* 16: 511-554

Oh, T.J., Gorman, M. and Cullis, C.A. (2000) RFLP and RAPD mapping in flax (*Linum usitatissimum*). *Theoretical and Applied Genetics* 101: 590-593

Rowland, G.G. (1991) An EMS-induced low-linolenic-acid mutant in McGregor flax (*Linum usitatissimum* L.). *Canadian Journal of Plant Science* 71: 393-396

Rowland, G.G. (1998) Growing flax: Production, management and diagnostic guide. Flax Council of Canada and Saskatchewan Flax Development Commission: 56p

Spielmeyer, W., Green, A.G., Bittisnich, D., Mendhem, N. and Lagudah, E.S. (1998) Identification of quantitative trait loci contributing to *Fusarium* wilt resistance on an AFLP linkage map of flax (*Linum usitatissimum*). *Theoretical and Applied Genetics* 97: 633-641

Surinder, S., McKinney, S. and Green, A. (1994) Sequence of a cDNA from *Linum usitatissimum* encoding the stearoyl-acyl carrier protein desaturase. *Plant Physiology* 104: 1075

Van Ooijen, J.W. (1992) Accuracy of mapping quantitative trait loci in autogamous species. *Theoretical and Applied Genetics* 84: 803-811

Van Ooijen, J.W. (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* 83: 613-624

Van Ooijen, J.W. (2003). MapQTL 5, Software for the calculation of QTL positions on genetic linkage maps. Kyazma B.V., Wageningen, Netherlands.

Wiersema, H.T. (1955) Flax scorch. *Euphytica* 4: 197-20

Chapter 7

Summarizing discussion and concluding remarks

Molecular markers as a tool in genetic studies in flax.
This thesis describes the application of molecular marker techniques and genetic analysis of several studies in flax. By means of different diversity indices and cluster analysis the variation within and between species can be analyzed. Besides the investigation of the variation of flax we describe the application of molecular markers in genetic studies as well. Optimizing the AFLP fingerprinting technique is necessary for optimal genotyping of mapping populations. Furthermore, good quality QTL mapping relies on high quality linkage maps. The use of different methods and careful genotyping might be the key to success. At last, QTL analysis can be used to understand the genetic basis of traits by combining molecular marker data with phenotypic trait values. One of the objectives of these studies is to obtain insight in the structure and size of the flax genepool. Furthermore, we want to provide recommendations to reveal new alleles via several new strategies. And the last objective is to make MAB ready to use in flax breeding. The latter will be mainly focused on several disease resistances and quality traits.

In **Chapter 2**, the molecular genetic diversity within and between the two morphological groups of flax (fiber flax and linseed) were investigated by means of AFLPs. From literature it is suggested that fiber flax has been selected from linseed and that the genetic diversity in linseed is larger than in fiber flax. Our results are in agreement with this concept: cluster analysis showed that fiber flax seemed to be a subset of the population of linseed cultivars. In addition, PCA showed that Dutch fiber flax cultivars diverge more from linseed cultivars, which is probably the effect of modern breeding activities with a narrow focus on fiber flax characteristics. Not only the PCA but also the AMOVA analysis showed that the genetic diversity of fiber flax is much smaller than in linseed. Fiber flax cultivars and linseed cultivars can be differentiated by several markers which are significantly associated with the morphotypes. The fact that morphotype associated markers were not necessarily genetically linked indicates that different regions of the genome are involved in the morphological differences between fiber flax and linseed. For the development of new cultivars with improved characteristics the more variable linseed cultivars should be considered as valuable genetic resources.

In **Chapter 3**, AFLPs were used to group species and to differentiate between groups within the genus *Linum*. Neighbor Joining dendrograms showed that not all species of the section *Linum* belong to a monophyletic group. This was also the case for species belonging to section *Linastrum*. However, within the groups the accessions from one species could be nicely differentiated. Although the AFLP analysis provides clear results on the differentiation or grouping of accessions within species and species within subgenera we do not draw conclusions on the phylogeny of *Linum* spp., because AFLP fingerprints are too dissimilar between the main groups to allow meaningful distance estimation. Furthermore, AMOVA and cluster analysis showed that accessions of the

same species within a group could not be differentiated from other accessions in the same group. Therefore, a revision of the genus *Linum* taking the molecular marker data into account is recommended.

In **Chapter 4** the effects of a preliminary screening of AFLP primer combinations of four parental genotypes of two mapping populations are presented. This screen aimed to identify primer combinations with the highest number of polymorphic bands to minimize the number of fingerprinting assays for the construction of a linkage map. The smaller genetic distance between two parents of a mapping population the larger the reduction of workload for genotyping the mapping population. An unexpected side effect of this selection of primer combinations is the observation of an elevated level of fragment length polymorphisms. In contrast to SNP based absence/presence AFLPs that behave as dominant markers, the AFLP markers based on length polymorphisms behave as codominant markers. The elevated number of allelic codominant length polymorphisms is causing a reduction of the total number of marker loci, but the utility of highly informative codominant markers for constructing genetic linkage maps and QTL analysis is of more value, since part of the BxH mapping population contained a substantial number of heterozygous loci. The molecular origin of these length polymorphisms is most likely based on microsatellite variation. The large number of fingerprints that were generated with two enzymes and various selective nucleotides were described in terms of number of bands per lane and number of polymorphisms. This dataset allows data mining to explain fingerprint complexity and level of polymorphism. Fingerprint complexity could be explained by compositional bias of C/G versus A/T residues in the restriction enzyme recognition site and similarly in the selective nucleotides. The level of polymorphism could not be explained by these factors, and thus cannot be predicted either.

In **Chapter 5** two linkage maps are constructed of different mapping populations, as well as an integrated linkage map. During the construction of the map we made use of two different computer algorithms to calculate marker order: JoinMap and RECORD. By comparing graphical genotype images of marker orders obtained by JoinMap or RECORD the most plausible order could be determined. This was done in several cases where the order of common markers on homologous linkage groups did not show co-linearity. In all conflicting cases the best marker order was presented by RECORD. Our results demonstrate that the RECORD algorithm is not only useful in high-density linkage mapping studies but can be used at lower marker densities as well. Furthermore, RECORD and JoinMap could be used as complementary tools in linkage map construction. Furthermore, the presence of more than two common markers in linkage groups enabled us to validate the best order of markers when contradictions in colinearity were observed.

In **Chapter 6** QTLs were identified involved in resistance and quantitative traits by combining field observation on the offspring of the VxE and BxH mapping populations with the VxE and BxH genetic linkage maps described in Chapter 5. Four traits important to flax breeding were investigated in more detail: scorch and *Fusarium* resistance, fatty acid composition of the seed oil and fiber quality traits. The inheritance of flax scorch resistance is based on a major QTL that explained the largest part of the phenotypic variation. However year and location specific minor QTLs were observed as well, which contributed significantly to a more desirable resistance level. It would be interesting to investigate the effect of multiple environments on scorch resistance. In this way a better understanding of the genetic model of scorch resistance might be obtained. The heredity of linoleic versus linolenic acid content was in accordance with literature reports: two unlinked additive desaturase loci (mutated in the E1747 parent) are involved in the biosynthesis of linolenic acid out of linoleic acid. *Fusarium* resistance in the BxH population and fiber content in the VxE population appeared to be more complex traits. Fusarium is controlled by at least five QTLs of which two occurred at different time points. The genetic model for *Fusarium* resistance at time point two was investigated more carefully and consisted of two loci with additive gene effect and an epistatic interaction between the two other loci. The differences observed at the two points raise intriguing questions about the mechanism of *Fusarium* resistance. More detailed studies on plant pathogen interaction at the different time points might give some answers. Fiber content in the VxE population appeared to be controlled by at least four QTLs, of which two QTLs had additive effect and the two other loci showed duplicate gene epistatic interaction. Furthermore, the position of fiber QTLs on the VxE and BxH map coincided. In this case we assume that both QTLs represent the same locus. Many co-localizing QTLs involved in different traits could be explained by the presence of a single locus with a pleiotropic effect on those different traits. However, more detailed correlation studies at these loci might show the true relation between traits.

Is it possible to make progress in flax breeding with the current gene pool?
Apparently, it is difficult to develop new cultivars with improved fiber content, fiber yield and disease resistance, as last ten years only few new cultivars appeared on the National List of varieties of field crops in the Netherlands. The results described in this thesis show that for all traits one or more loci are present in the flax genepool which are polymorphic between cultivars. Combining all positive alleles at these QTLs might result in cultivars that outperform the current elite material. For fiber content we could not identify QTLs that were not present already in cultivar Hermes, and thus little perspective is offered for this trait due to the limited amount of genetic diversity in the fiber flax gene pool. This is confirmed by the modest achievements of breeders. The fiber content of the current new cultivars does not exceed the fiber content of Hermes, which was introduced approximately ten

years ago. Only the straw yield could be improved slightly in the new cultivars. Combining all positive QTL alleles for loci involved in fiber content and straw yield might result in a high fiber yielding cultivar.

With markers efficient pyramiding of positive alleles of the QTLs becomes feasible. A major drawback of QTL mapping, as a strategy to identify favorable alleles at loci involved in important traits, is the limited amount of germplasm involved. In our study only four parental genotypes have been studied for the presence of favorable alleles, which is only a fraction of the total number of cultivars. The parental genotypes were carefully chosen that they should carry favorable alleles with high probability. To guarantee the segregation of those favorable QTL the other parent should not carry those alleles. For example favorable alleles for scorch resistance, seed yield and lodging should be expected to segregate from Belinka; fiber content, fiber yield, plant length and *Fusarium* resistance should be expected to segregate from the Hermes; fiber content and *Fusarium* resistance should be expected to segregate from the Viking; oil content seed, linoleic acid content seed oil and lodging should be expected to segregate from E1747. This experimental design is such that crosses between a favorable and inferior parent are used, but these favorable alleles may already be fixed in germplasm that is superior for the traits involved. An experimental design were a superior parents is crossed with another superior parent would either render unsuccessful, because QTLs are not segregating or transgressive segregation will expose favorable allele combinations that have not yet been achieved. Only in the case of oil content seed, seed yield, *Fusarium* resistance, plant length and lodging transgressive segregation has been observed.

An experimental design that circumvents the disadvantages of QTL analysis is association mapping. A large collection of unrelated genotypes is screened with a large number of markers to identify associations between the trait value and marker alleles. Although Chapter 2 was designed only with the purpose to study the genetic diversity in the flax genepool and were the numbers of markers too low to approach a genome wide scan linkage disequilibrium, yet a number of marker trait associations have been observed. The next logic molecular genetic experiment in flax would be an elaboration and integration of the results of Chapter 6 and Chapter 2. Markers that were positively associated with traits in Chapter 2 should be tested in the QTL approach by analyzing that specific AFLP primer combination on the mapping populations. And in reverse, the markers near QTLs that have been identified in Chapter 6 should be tested in the association mapping approach to verify if these marker QTL associations can be extrapolated to a wider range of germplasm.

Will MAS become a cost effective tool in fiber flax breeding?
In general MAS can be implemented in breeding programs in case of (*i*) disease resistance where reliable inoculation and scoring methods are not available (e.g. scorch resistance), (*ii*) complex traits with low heritabilities (e.g. straw yield) or (*iii*) traits where a reduction in costs per data point can be obtained by replacing more expensive phenotyping methods with less expensive genotyping methods. (e.g. fiber content). Furthermore, for the effective implementation of MAS, the association between the marker allele and the QTL allele should hold true in a larger set of genotypes. In such linkage desequilibrium (LD) studies, population structure should be taken into account. In view of these criteria, it is obvious that some favorable alleles that have been detected in the BxH population and VxE population should be very suitable for MAS in fiber flax or linseed breeding programs.

Recommendations for flax breeders.
Throughout the chapters in this thesis the level of genetic diversity has been an issue. The low level of genetic diversity is an obstacle for molecular genetic research projects; this is a concern for the practical breeder as well. Neither should genetic diversity be seen as a goal in itself. It is breeders themselves who are facing the limitations caused by a narrow genetic diversity. In the first place, their breeding efforts could be more rewarding if there is a substantial amount of selectable genetic variation. Secondly, according to the UPOV criteria to grant Plant Breeders Rights, the newly selected cultivars have to be morphologically distinct from existing cultivars. If the need to broaden the genetic basis of fiber flax is indeed accepted by commercial breeders, then this thesis provides some background information to achieve this goal. The gene pool of linseed cultivars is more variable, and crosses between fiber flax and linseed cultivars should be encouraged. However, earlier attempts to identify positive QTLs from otherwise inferior germplasm have often been unsuccessful.

The results in Chapter 2 demonstrate that the level of molecular diversity between linseed and fiber flax is sufficient to allow the development of a backcross introgression library (BILs; Eshed and Zamir 1995) In such a library it is likely that even minor QTLs can be identified, because the subtle QTL effects are not overshadowed by the huge negative effects of major QTL without breeding value from the donor parent (Tanksley and Nelson, 1996).

The development of BILs of linseed segments in the recurrent parent of fiber flax should be regarded as a conservative strategy to broaden the genepool of fiber flax. In several other crop species backcross introgression libraries have been developed using wild relatives. Remarkably, all these efforts have resulted in the identification of favorable QTL alleles from an otherwise unfavorable background. In view of the long cultivation

history of flax as a crop species and the mode of reproduction (selfing) it to be expected that genetic bottlenecks have occurred and consequently the loss of positive alleles. The more recent split between linseed and fiber flax is a good reason not to expect favorable alleles in the linseed genepool, because the loss of genetic diversity has occurred before. Hence it should be expected that exploitation of *L. bienne* or *L. angustifolium* germplasm is more promising to identify favorable QTL alleles that have not been retained during the breeding history of fiber flax. Moreover, this wider cross will result in a higher level of polymorphism, and therefore a more efficient development of molecular markers.

Recommendations for future molecular genetic research

The results of this thesis can be applied in flax breeding programs. At the same time the results are helpful in understanding the genetic basis of resistance and quality traits in greater detail. To understand the process underlying fiber quality and fiber production in flax, the molecular role of the genes underlying the QTLs shown in Chapter 6 should be characterized. Genes involved in cell wall metabolism are suitable candidate genes to test their role on phenotypic variation in fiber content, fiber quality and fiber yield. If those candidate genes, as well as expressed sequence tags (ESTs) for cell wall metabolism, obtained during the fiber formation process, can be localized on our linkage maps, their position can be compared with the fiber related QTLs detected in Chapter 6. If a QTL involved in fiber yield or quality traits and a specific EST co-localize, then this is a strong indication for the EST to be the gene affecting the trait.

Likewise, the molecular genetics of disease resistance can exploit the results shown in Chapter 6. Scorch resistance appeared to be controlled by a major QTL (Chapter 6). However, we cannot conclude whether or not this major QTL is an *R*-gene, of the elicitor-receptor model of plant - pathogen interaction. The BxH mapping population could be very well used for more elaborate plant - pathogen studies to investigate the molecular role of scorch resistance locus, and also to determine the causal pathogen of scorch. For the latter issue, the BxH mapping population should be exposed to the different pure isolates of putative pathogens and/or combination of pathogens in controlled field trials or greenhouse experiments.

References

Eshed, Y. and Zamir, D. (1995) A genomic library of *Lycopersicon pennellii* in *L. esculentum*: A tool for fine mapping of genes. *Euphytica* 79: 175-179

Tanksley, S. D., and Nelson, J.C., (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics* 92: 191-203

Summary

In this thesis five molecular genetic studies on flax (*Linum usitatissimum* L.) are described, of which two chapters aim to characterize the genetic structure and the amount of genetic diversity in the primary and secondary gene pool of the crop species. Three chapters describe the development of AFLP markers, linkage map construction and QTL analysis of resistance and quality traits.

Genetic diversity in the primary gene pool was studied by AFLP fingerprinting 110 varieties representing linseed and fiber flax germplasm, the two cultivar groups in flax. Fiber flax germplasm is a narrow subset within linseed flax, and modern breeding tends to further narrow the genetic diversity. Forty two AFLP bands showed strong association with cultivar group. In a second diversity study 52 accessions of 27 *Linum* species could be differentiated in clearly distinct groups. These groups showed some resemblance to the current taxonomic grouping of species in sections of the genus *Linum*, but many aberrations were observed. In our view the sexually compatible wild species *L. bienne* and *L. angustifolium* should be excellent sources to broaden the genetic basis of *L. usitatissimum*.

In view of the extremely low number of polymorphic markers in AFLP fingerprints, an entire study was dedicated to select primer combinations with the highest number of markers. Unselected primer combinations show on average 2.3 and 3.8 polymorphisms for two different parental combinations. Our selection of favourable primer combinations has on average 7.2 and 10.7 markers, allowing a 50% reduction of the workload involved in linkage map construction. An unexpected side effect of this selection was the frequent occurrence of allelic codominant length polymorphisms among the resulting AFLP markers.

Two linkage maps were constructed in mapping populations descending from a fiber x fiber and a fiber x linseed cross. These maps were integrated to obtain a linkage map of flax. Two different computer algorithms were used to calculate marker order, JoinMap and RECORD, which are complementary, and their joint use allowed reaching the most plausible marker order. For 18 different traits involved in disease resistance (*Fusarium* wilt and scorch) and quality traits related to fiber yield and quality, as well as fatty acid composition we identified a total of 63 QTL positions. In practical breeding some of these QTL loci could be of high practical value. Phenotypic selection of these loci would require field trials of the highest quality, but genotypic selection of these loci is easily pursued with DNA diagnostics. In view of the results obtained with these five molecular genetic studies recommendations are made to implement Marker Assisted Breeding (MAB), to proceed with linkage disequilibrium studies and to broaden the genetic basis of fiber flax breeding, for example with the aid of a library of backcross introgression lines.

Samenvatting

Dit proefschrift beschrijft vijf genetische studies in vlas (*Linum usitatissimum* L.), waarvan twee als doel hebben het karakteriseren van de genetische structuur en de hoeveelheid genetische variatie in de primaire en secundaire genenpool van het gewas. De drie andere hoofdstukken beschrijven de ontwikkeling van AFLP merkers, de constructie van een koppelingskaart en localistaie van resistentie en kwaliteitseigenschappen in vlas via QTL analyse.

De genetische diversiteit in de primaire genenpool is bestudeerd door middel van AFLP *fingerprinting* van 110 rassen die het olie- en vezelvlas germplasm, de twee cultivar groepen van vlas, representeren. Het vezelvlas materiaal is een deelverzameling van olievlas. Vezelvlas blijkt een smalle genetische basis te hebben die door moderne plantenveredeling verder dreigt te af te nemen. Tweeenveertig AFLP banden bleken sterk te zijn geassocieerd met het verschil in cultivar groepen. In een tweede genetische diversiteitsstudie konden 52 accessies van 27 *Linum* soorten worden onderscheiden als duidelijk verschillende groepen. Op enkele afwijkingen na, vertoonden deze groepen overeenkomsten met de huidige taxonomische groepering van soorten in de vijf secties van het genus *Linum*. De kruisbare wilde soorten *L. bienne* en *L. angustifolium* zouden uitstekende bronnen kunnen zijn om de genetische basis van *L. usitatissimum* te verbreden.

Gezien het extreem lage aantal polymorfe merkers in de AFLP fingerprints tussen vlas cultivars, is een volledige studie besteed aan het selecteren van primercombinaties met het hoogste aantal merkers. Ongeselecteerde primercombinaties bevatten gemiddeld 2,3 en 3,8 merkers voor respectievelijk de vezel x vezel en de vezel x olie oudercombinatie. De door ons gemaakte selectie van primercombinaties met het hoogste aantal merkers resulteerde in gemiddelden van 7,2 en 10,7 merkers, waardoor een arbeidsreductie van 50% kon worden bereikt voor de constructie van koppelingskaarten. Een onverwacht bijeffect van deze selectie was het veelvuldige voorkomen van allelische codominante lengte polymorfismen tussen de verkregen AFLP merkers.

Twee koppelingskaarten zijn geconstrueerd in mappingpopulaties afkomstig van een kruising tussen twee vezelvlasrassen en een vezelvlas x olievlas kruising. Deze kaarten zijn gecombineerd tot één geïntegreerde kaart van vlas. Twee verschillende computer algoritmes, Joinmap en RECORD, zijn gebruikt om de merkers op volgorde te plaatsen. Het gelijktijdige gebruik van deze complementaire algoritmes maakt het mogelijk de meest waarschijnlijke merkervolgorde te verkrijgen. Voor 18 verschillende eigenschappen die onder andere betrokken zijn bij ziekteresistentie, vezelopbrengst, vezelgehalte en vetzuursamenstelling, konden 63 QTL posities worden geïdentificeerd. Voor praktische veredelingsdoeleinden kunnen enkele van deze QTL loci waardevol zijn. Fenotypische selectie van deze loci vergt veldproeven van de hoogste kwaliteit, terwijl genotypische selectie eenvoudig kan worden gedaan met DNA diagnostiek. Gezien de verkregen resultaten van de vijf genetische studies, zijn er aanbevelingen gemaakt voor de inplementatie van merkergestuurde veredeling, het uitvoeren van ‘linkage disequilibrium’ analyses en het verbreden van de genetische basis voor vezelvlasveredeling door middel van bijvoorbeeld teruggekruiste introgressielijnen.

Nawoord

Het is zover: voor je ligt het proefschrift waarmee ik een leuke en leerzame periode van ruim vijf jaar bij het Laboratorium voor Plantenveredeling afsluit. Voordat ik een aantal mensen persoonlijk ga toespreken wil ik iedereen bedanken die direct of indirect een bijdrage aan dit proefschrift heeft geleverd.

Beste Herman, als dagelijks begeleider en copromotor heb jij verreweg de belangrijkste rol gespeeld tijdens mijn promotiewerk. Ondanks jouw drukke agenda kon ik altijd bij je binnenvallen met vragen of nieuwe resultaten. Jouw wetenschappelijke kennis en ervaring hebben ervoor gezorgd dat er uiteindelijk meer diepgang in mijn proefschrift is gekomen. Veel tijd heb ik niet van je gevraagd de afgelopen 5 jaar, maar de laatste paar maanden is daar wel verandering in gekomen. Dat je op vrijdag tot 4.00 uur ’s nachts door bent gegaan met het corrigeren van een manuscript waardeer ik zeer. Ook de sportieve avondjes op de squashbaan waren altijd erg leuk. Bedankt voor de goede samenwerking.

Piet, je hebt mijn onderzoek op een afstandje gevolgd, maar als ik vragen had stond je altijd paraat. De laatste twee maanden heb je echter een belangrijke rol vervuld door je frisse kijk op mijn experimentele hoofdstukken. Dit heeft tot een mooie ‘finishing touch’ geleid en daar wil ik je hartelijk voor bedanken.

Mijn paranimfen, Hans en Björn, ik vind er erg leuk dat jullie me willen bijstaan tijdens de verdediging van mijn proefschrift. Hans, we hebben samen een gezellige tijd gehad tijdens het pingpongen, klaverjassen en andere activiteiten. Door onze verschillende hobbies waren de gesprekken altijd erg leuk en divers. En jij bent tenminste niet zo’n watje die niet durft te katapulting op de kermis in Nijmegen! Daarnaast heb je me geholpen met de constructie van de genetische kaart. Ik ben blij dat je een leuke baan hebt gevonden en een collega van Inge bent geworden. Björn, hoeveel hobbies kun je samen delen. Lekker slap ouwehoeren over voetbal, feestje hier biertje daar, Hans Teeuwen, wielrennen en veldrijden, maar ook vooral het klaverjassen niet willen missen. Ik hoop dat we in de toekomst nog eens een paar leuke toertochten (de hel van Wageningen?) samen kunnen fietsen. Verder wil ik je nog veel succes toewensen met jouw promotieonderzoek.

Ja, als je twee keer verhuisd moet je behoorlijk wat kamergenoten bedanken. In het echte oude gebouw op de Lawickse Allee heb ik, als beginnde onderzoeker, met veel plezier de kamer gedeeld met Jean-Paul, Ronald en Ji. We hadden altijd leuke gesprekken over van alles en nog wat. Ji, thanks for the lovely tea you brought from China. Op de Binnenhaven kreeg ik twee nieuwe kamergenoten. Dear Manga, I really loved our conversations about life, politics, sports, work and of course ‘coach van het jaar’ during coffee breaks and working time. I was very happy that you could fill in the empty place in our new office at ‘De Haaff’. Marieke, ook met jou heb ik een gezellige tijd gehad. Lekker elke maandagochtend bijkletsen wat we allemaal gedaan hadden in het weekend,

maar ook de inhoudelijke discussies waren altijd leuk en leerzaam. Tot slot wil ik ook Adillah, René en John bedanken voor de gezellige sfeer op onze kamer in ‘De Haaff’.

De collega’s van de statistiek groep hebben er de laatste maanden voor gezord dat het schrijven nog een beetje draaglijk was. Zinvolle en zinloze gesprekken tijdens het ‘bakkie’ hebben voor een prettige afleiding gezorgd. Special thanks go to Marcos for helping me out with some statistical problems.

Anton Vels en Michel Hagendoorn van Unifarm wil ik bedanken voor het verzorgen van mijn vlasplanten en het helpen bij het dorsen van het zaad. Door jullie kundigheid zijn alle proeven goed verlopen.

Waar zouden we zijn zonder Annie en Letty. Of ik nou kwam declaren, klagen of een gezellig praatje maken, altijd waren jullie bereid te helpen of te luisteren. Nogmaals bedankt voor jullie goede zorgen.

Naast de veredelingsbedrijven Cebeco Seeds (Innoseeds), Dobbelaar Breeding en Van de Bilt Zaden en Vlas die het onderzoek financieel steunden, wil ik hun veredelaars, Hein de Jong, Peter Keijzer en Raimon Laan bedanken voor hun bijdrage tijdens de halfjaarlijkse vergaderingen en het nauwkeurig fenotyperen van de mappingpopulaties. Ook de provincie Zeeland wil ik bedanken voor de financiele steun van mijn project. David, jouw inzet voor de subsidie aanvraag bij de provincie Zeeland en de administratieve taken tijdens mijn onderzoek en halfjaarlijkse vergaderingen heb ik erg gewaardeerd.

Mijn studenten Gerben Mennink, Bjorn Pieper, Eugenie van de Bilt en Menno ter Maat wil ik graag bedanken voor hun uitstekende inzet tijdens hun stage en afstudeervak. Veel succes in jullie verdere loopbaan.

Verder wil ik mijn familie en vrienden bedanken voor de interesse die jullie hadden in mijn onderzoek. Lieve Pa en Ma, ik ben jullie erg dankbaar dat ik de mogelijkheid heb gekregen te gaan studeren en dat jullie mij altijd hebben gesteund met de keuzes die ik heb gemaakt. Lieve Inge, jij hebt een heel belangrijke bijdrage geleverd aan de totstandkoming van dit proefschrift. Ik wil je vooral bedanken voor je onvoorwaardelijke steun de afgelopen 5 jaar. Vooral het laatste jaar was het niet altijd even gemakkelijk, maar jij hebt ervoor weten te zorgen dat de vaart erin is gekomen.

Curriculum Vitae

Jaap Vromans werd geboren op 19 december 1975 in Bergeijk. In juni 1995 behaalde hij aan het Rythovius College te Eersel het VWO diploma. In datzelfde jaar begon hij zijn studie plantenveredeling en gewasbescherming aan de Wageningen Universiteit. Bij het laboratorium voor plantenveredeling heeft hij als student onderzoek verricht naar de inactivatie van het korrelgebonden zetmeelsyntasegen in aardappel door middel van transposons. Tijdens zijn stage bij Seminis Vegetable Seeds in Enkhuizen heeft hij onder andere gewerkt aan hybride zaadopbrengst bij radijs en moleculaire karakterisatie van witte kool met behulp van ‘Inter Simple Sequence Repeat’ (ISSR). In september 2000 behaalde hij zijn doctoraal diploma. In diezelfde maand trad hij in dienst als AIO bij het laboratorium voor plantenveredeling aan de Wageningen Universiteit in het project over de toepasbaarheid van moleculaire merkers in de vlasveredeling. De resultaten van dat project staan beschreven in dit proefschrift.

Education Statement of the Graduate School Experimental Plant Sciences

Issued to: Jaap Vromans
Date: March 15, 2006
Group: Laboratory of Plant Breeding, Wageningen University

1) Start-up phase	date
► **First presentation of your project**	
The use of molecular markers in flax (*Linum usitatissimum* L.) breeding	November 30, 2000
► **Writing a review or book chapter**	
General introduction: 'Flax breeding and marker assisted selection'	Oktober, 2005
Subtotal Start-up Phase	*5 credits**

2) Scientific Exposure	date
► **EPS PhD student days**	
PhD students day 2001, Wageningen	January 25, 2001
PhD students day 2002, Wageningen	January 24, 2002
PhD students day 2003, Utrecht	March 27, 2003
► **EPS theme symposia**	
Theme IV symposium, Nijmegen	December 13, 2001
Theme IV symposium, Wageningen	December 20, 2002
► **NWO Lunteren days and other National Platforms**	
ALW meeting Lunteren 2001, Plant Science	26-27 April, 2001
ALW meeting Lunteren 2002, Plant Science	15-16 April, 2002
ALW meeting Lunteren 2003, Plant Science	7-8 April, 2003
ALW meeting Lunteren 2004, Plant Science	5-6 April, 2004
► **Seminars (series), workshops and symposia**	
Plant breeding seminars (15x)	2000-2004
Frontiers in plant development seminar series (5x)	2000-2002
PE&RC annual meeting: Ethics in Science	2002
► **Research discussion**	
Research discussion plant breeding	2000-2004
Discussion Group at PE&RC	2002
► **International symposia and congresses**	
Plant & Animal Genome XII	January 10-14, 2004
International conference of FAO European research network on flax and allied fiber plants for human welfare	December 8-11, 2003
► **Presentations**	
Oral presentation at meeting Plant Breeding, Genetics & Breeding and PPO	May 17, 2002
Poster	December 8, 2003
Oral presentation at Lunteren	April 6, 2004
► **IAB interview**	March 28, 2003
► **Excursions**	
Flax breeders: Application of molecular markers in plant breeding (Seminis Vegetable Seeds)	2003
Subtotal Scientific Exposure	*11.5 credits**

3) In-Depth Studies	date
► **EPS courses or other PhD courses**	
AFLP markers in systematics and breeding	May 16-18, 2001
The analysis of natural variation within crop and model plants	April 22-25, 2003
Subtotal In-Depth Studies	*1.4 credits**

4) Personal development	date
► **Skill training courses**	
Scientific writing level 5	May-July 2003
Digital guide to scientific artwork	March 4-5, 2002
► **Organisation of PhD students day, course or conference**	
Organisation of PhD students day 2002	January 24, 2002
► **Membership of Board, Committee or PhD council**	
Membership of PhD council	2002-2004
Subtotal Personal Development	*3.6 credits**

TOTAL NUMBER OF CREDIT POINTS*	21.5

** A credit represents a normative study load of 40 hours of study*

The work described in this thesis was carried out in the Graduate Schools Experimental Plant sciences and Production Ecology and Resource Conservation at the Laboratory of Plant Breeding, Wageningen University. Main financial support was provided by the Dutch provinces Zeeland and the Dutch breeding campanies, Cebeco Seeds, ‘Van de Bilt Zaden en Vlas’ and Dobbelaar Breeding.

Cover: ‘Vlas, Rilland’ painted by Dick Bakker

Printed by PrintPartners Ipskamp