GENETIC ANALYSIS OF OIL CONTENT AND COMPOSITION IN OAT (*Avena sativa* L.)

BY

CHRIS ANDERSON

A Thesis

Submitted to the Faculty of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

Department of Botany

University of Manitoba

Winnipeg, Manitoba

August, 2000

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION PAGE

Genetic Analysis of Oil Content and Composition in Oat (*Avena Sativa L.*)

BY

Chris Anderson

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirements of the degree of Master of Science

ACKNOWLEDGEMENTS

The author would like to acknowledge the contributions of many people who were instrumental in bringing this project to completion. The advice and supervision of Dr. Penner, Dr. Brown, Dr. Ford, Dr. Punter and Dr. Scarth has been appreciated.

My thanks are also extended to the entire Penner lab at the Cereal Research Centre. Everyone in the lab added to the project in some way but special thanks are due to Marcy Zirino, Leslie Bezte, Elsa Riemer and particularly Allison Zacharias.

Dr. Duguid and his oat crew also deserve special thanks for their advice and assistance in the field trials. Thanks also to Dr. Noll, Dr. White and Colin Demaniuk for their assistance with the NIR analysis and oil extraction. Dr. Daun and his technical staff at the Grain Research Lab also provided valuable advice and assistance with the lipid analysis.

I would also like to express my sincere appreciation and thanks to my family and friends, especially Bob, Joan and Kelly, who offered encouragement, support and occasional prodding along the way. Your good-natured barbs finally got through, thank you.

This project was funded by in part by an Agriculture and Agri-Food Canada/NSERC Research Partnership Support Program Grant. Resources were also provided by the Quaker Oat Company and General Mills.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Oat oil is an important quality characteristic for the end users of oat (*Avena sativa* L.) and oat products. The objective of this study was to investigate the inheritance of oil content and composition and identify markers linked to genes affecting oat oil using a linkage map. Two hundred and twenty three lines from an F_7 population of a cross between the high oil cultivar AC Marie (approximately 5 percent oil) and the low oil cultivar Cascade (approximately 3.5 percent oil) were analyzed for fatty acid composition. Correlations (P=0.01) between total oil content and palmitic, oleic and linoleic fatty acids were found to be 0.92, 0.94, and 0.97 respectively. Results indicated that a single gene was responsible for control of each different fatty acid and for total oil content. Near infrared reflectance (NIR) spectroscopy was used to analyze an 80 line subpopulation of the AC Marie/Cascade lines. Broad sense heritability for total oil was 0.19 and for groat oil (total oil content corrected for percent hull content), 0.12. A linkage map of the cross was created using amplified fragment length polymorphism* (AFLP) analysis. A total of 591 markers were identified. These were organized into 35 linkage groups covering a map distance of 1077 cM. Twenty two markers were unlinked and the overall genome size was estimated to be approximately 2900 cM. A single QTL affecting total oil content was identified linked to the marker CRO217. This QTL was responsible for a 0.58 percent increase in oil content when comparing the class means for those lines carrying CRO217 with the class mean for lines not carrying CRO217. The results of this study will increase knowledge of oat oil inheritance, assist in the development of improved oil cultivars, facilitate QTL identification for other traits, and increase knowledge of oat genome structure.

* AFLP is a registered trademark of Keygene Inc.

1.0 INTRODUCTION

Oat (*Avena sativa* L.) is the third largest cereal crop grown in Canada, in terms of both harvested area and production. In 1996, Canada was the second largest producer of oats in the world with exports exceeding one million metric tonnes (Canada Grains Council, 1996). Oat is the fourth most important cereal for human consumption behind wheat, rice and corn. Worldwide consumption of food oat is increasing (Burnette et al., 1992). Oat continues to be an important crop both internationally and within Canada.

Currently, the major uses for oat are either as human food or as a component of animal feed. Oat provides a high level of energy, compared to other cereals, due to its high lipid content. Lipids have a higher energy value than carbohydrates. Oat lipids have a favourable composition in that they are high in unsaturated fatty acids, particularly linoleic acid which is an essential fatty acid in human diets (Peterson, 1992). Raising the lipid content of oat has been suggested as a way to add value to the crop (Frey and Hammond, 1975; Kalbasi-Ashtari and Hammond, 1977). Oat oil compares favourably with soybean oil but oat contains more active lipases, which are extracted with the oil and need to be deactivated to allow for stable storage.

A great deal of variation in oil content exists in oat germplasm. Brown and Craddock (1972) found oil contents ranging from three to eleven percent within the world oat collection. The majority of individual cultivars range from between five and nine percent oil content. Fatty acid composition has also been examined and significant variation exists among cultivars (Youngs, 1986).

Inheritance of oil content and composition is complex. Oat oil content is polygenically controlled with both additive (Brown et al., 1974) and dominant (Thro and Frey, 1985) gene action. Individual fatty acids are also controlled by both additive and dominant gene effects, with multiple genes/alleles contributing to variation (Karow and Forsberg, 1984). Correlations between individual fatty acids and total oil content exist, as well as relationships between specific fatty acids (Youngs and Puskulcu, 1976). Modes of inheritance for both total oil and oil composition seem to be specific to the populations used in the study.

Manipulating oil content in oat poses several challenges for breeders. Evaluation of early generation or elite material for quality traits such as oil content can be limited by lack of seed availability, time required for analysis and the cost of biochemical analysis (Forsberg and Reeves, 1992). Frey and Hammond (1975) indicated that selection for high oil results in changes to the fatty acid profile. High oil cultivars would tend to have high levels of oleic acid and reduced levels of linoleic and linolenic acids.

The complex nature of oil inheritance and the inability to visually score materials for phenotype make this group of traits a good candidate for marker development. Tanksley et al. (1989) have suggested that DNA based markers and genomic maps will allow complex traits to be broken down into simple genetic components, increase understanding of trait expression and interaction in given backgrounds, and allow faster introgression of genetic material from wild type or related genomes.

The objectives of this study were: 1) to determine the variation of oil content and composition in a population of a high oil by medium oil cross, 2) to create a molecular marker map of the genome of this oat cross, and 3) to identify molecular markers linked to the inheritance of oil content and composition. The availability of molecular markers linked to oil content and quality and an increased understanding of the oat genome will be valuable for the development of improved oil cultivars.

2.0 LITERATURE REVIEW

2.1 Oat Oil

Oat contains the highest lipid concentration of the cereal grains (Price and Parsons, 1975) which makes it an excellent energy source for use in foods or feed. Youngs et al. (1977) compared two oat cultivars and found that the majority of oat lipids are found in the endosperm. Although the concentration of lipid in the endosperm is not high, the endosperm comprises the majority of the kernel and thus contains the highest total amount of oil. Much of the oil found in the endosperm is found in the aleurone and subaleurone cells which contain large lipid globules (Youngs, 1986). The germ contains the highest concentration of lipid, up to 20 percent in the scutellum, but the germ represents only six to eight percent of the total kernel. The hull makes up about 25 percent of the kernel but contributes only about 2 percent oil and therefore adds very little to total oil content.

2.2 Lipid Biosynthesis

Lipid biosynthesis in higher plants is a complex process. The primary storage product for lipids in seeds is in the form of triacylglycerides. Saturated fatty acids are synthesized *de novo* in plastids or proplastids. Most plants create chains that are either 16 or 18 carbons long using a common system of seven enzymes known as fatty acid synthetase (FAS) (Stumpf, 1987). Monounsaturated fatty acids are also produced in the plastids by substrate specific desaturase enzymes (Jaworski, 1987). Further desaturation and elongation also appears to be governed by specific enzymes. Most of this occurs in the cytoplasm and endoplasmic reticulum. Final formation of triacylglycerides also occurs in the endoplasmic reticulum. Fatty acids released from the plastids are moved into the endoplasmic reticulum where they are attached to a glycerol backbone in a series of enzyme mediated steps (Stymne and Stobart, 1987). The newly formed triacylglycerides are collected and released from the endoplasmic reticulum into the cytoplasm as oil bodies.

2.3 Oil Content Variation and Inheritance

The total amount of oil shows considerable variation among cultivars. Brown and Craddock (1972) tested over 4,000 entries of the world oat collection for total oil content. They

found lines with oil content as low as 3.1 percent and as high as 11.6 percent. Over 90 percent of the lines tested fell between five and nine percent oil content. These results are consistent with those reported in similar studies, although lines with oil contents either higher or lower have also been found (Frey and Hammond, 1975; Baker and McKenzie, 1972; Sahasrabudhe, 1979). Since these studies were conducted, lines with substantially higher groat oil content have been created. Branson and Frey (1989) created a 14.5 percent groat oil line and in 1991, Schipper and Frey reported developing a line with 16.2 percent groat oil content.

Several studies into the inheritance of oil content in oat have been performed and it appears that the mode of inheritance is dependent upon the populations under study. Baker and McKenzie (1972) found that heritabilities ranged from 0.68 to 0.93 in crosses between unrelated lines. A cross between sister lines resulted in a heritability of only 0.18. They suggested that oil inheritance is controlled through multiple factors. Similar results were presented by Brown et al. (1974). Heritabilities in this study ranged from 0.60 to 0.79 and additive genetic action appeared to be the major mode of inheritance. Frey et al. (1975) also postulated that oil content is polygenically inherited but the lines they used showed partial dominance for high oil inheritance. In 1985, Thro and Frey found a large number of transgressive segregants in their populations, also indicating that inheritance of groat oil content was largely additive. Applying knowledge of the modes of inheritance postulated above, several groups have used recurrent selection strategies to produce high oil oat lines (Branson and Frey, 1989; Thro and Frey, 1985; Schipper and Frey, 1991).

Correlations between oat oil and other traits are important if successful selection for total oil content is to occur. Brown et al. (1966) reported a significant negative correlation between oil content and protein content and a slightly negative correlation between oil content and kernel weight. Schipper and Frey (1992) also reported negative correlations between oil content and protein, and oil content and test weight when selecting for high oil but interestingly these correlations did not hold up when selecting for high protein content. No correlation between oil content and kernel weight, kernel density, or percent hull was detected by Brown and McKenzie

(1972) or between oil percentage and groat weight, heading date or plant height by Frey et al. (1975).

2.4 Oil Composition Variation and Inheritance

The composition of oil in oat is also highly variable between cultivars. Triglycerides compose the majority of the stored lipid in oat but the relative amounts of the fatty acids found are variable. Palmitic acid (16:0), oleic acid (18:1) and linoleic acid (18:2) make up over 90 percent of the fatty acids found. Stearic acid (18:0) and linolenic acid (18:3) each comprise between one and four percent of the total. The remainder is made up of trace fatty acids (Youngs et al., 1986). Frey and Hammond (1975) tested 64 lines and found that palmitic acid ranged from 14 to 23 percent, oleic acid from 29 to 52 percent, and linoleic acid from 26 to 48 percent of total fatty acids. Similar results for fatty acid composition were reported by Youngs and Puskulcu (1976), de la Roche et al. (1977), Sahasrabudhe (1979), and Karow (1980). In each of these studies, linoleic acid was, based on the mean, the most abundant fatty acid, followed by oleic acid, palmitic acid, linolenic acid and stearic acid. While linoleic is, on average, the most abundant fatty acid, many lines exist where oleic acid represents a greater portion of the total fatty acid component. Oleic acid also exhibits a much greater range of variation than the other fatty acids.

Heritability estimates for the major fatty acids (palmitic, oleic and linoleic) have been found to be very high. Youngs and Puskulcu (1976) reported heritabilites of 91, 99, and 96 percent respectively for palmitic, oleic, and linoleic acids. Thro et al. (1985) determined that additive gene action was the most important mode of inheritance and their estimates of heritabilities were also quite high at 68 percent for palmitic acid, 72 percent for oleic acid and 64 percent for linoleic acid. Similar heritabilities were calculated by Karow and Forsberg (1984), who tested a high oil by medium oil population (Dal/Sauk) and a high by low population (Dal/Exeter). Heritabilities for oleic and linoleic acids were very consistent between the two populations at 66 and 79 percent respectively. Their work found that additive genetic effects control linolenic inheritance but partial dominance was observed in the inheritance of oleic and linoleic acids.

Interesting relationships between total oil content and various fatty acids have been described as well as correlations between individual fatty acids (Karow and Forsberg, 1984; Thro

et al., 1985; Youngs and Puskulcu, 1976; Frey and Hammond, 1975). Oleic acid content positively correlated with total oil content. Sahasrabudhe (1979) postulated that this was due to a higher incidence of oleic acid in triglycerides in high oil lines, since the additional lipid tends to be stored as triglycerides. Linoleic and palmitic acid content tends to be neutral or negatively correlated with total oil content. Oleic acid is negatively correlated with each of the other fatty acids and a similar relationship exists between palmitic and linoleic. Karow and Forsberg (1985) have suggested that a single gene with multiple alleles may be responsible for the oleic – linoleic relationship. Similar genetic control mechanisms have been reported in corn (de la Roche et al., 1971) and safflower (Knowles, 1965). These relationships also indicate that, to a certain extent, each of the fatty acids is produced at the expense of the others.

2.5 Environmental Effects

Kibite and Edney (1992) studied the effects of location and growing season on a number of oat quality traits. Location was the most important factor in this test but the genotype x environment interaction, for all traits (including oil) was small, suggesting that few changes in rank order should result between locations. However, cool environmental conditions may influence both total oil content and oil composition. Welch (1975) found that oats sown in the winter produced slightly higher levels of total oil. The study also found that there is an increase in unsaturated fatty acids, specifically oleic and linoleic acids, under these conditions. Similar results were obtained by Saastamoinen et al. (1989). These results are consistent with a number of other studies conducted in other higher plant species where cool growing conditions have been found to cause increases in levels of C_{18} unsaturated fatty acids (Stymne and Stobbart, 1997).

2.6 Oat Genome

Cultivated oat (*Avena sativa* L.) is a hexaploid species with a chromosome number of n=3x=21 which are organized into three genomes A,C, and D (Rajathy and Thomas, 1974). Oat, like wheat (*Triticum aestivum* L.), is an allopolyploid with each of the three genomes arising from distinct diploid progenitors (Thomas, 1992). The oat genome falls between that of wheat and barley in terms of number of base pairs. Wheat contains approximately 16 billion base pairs, oat

11.3 billion base pairs, and barley only 4.9 billion base pairs (Arumuganathan and Earle, 1991). One copy of oat genomic DNA would weigh 23.4 pg.

Due to the hexaploid nature of oat (*Avena sativa* L.), certain levels of homology exists between homologous chromosomes of the A, C, and D genomes. There is evidence to suggest that the oat genome is organized in a similar fashion to that of bread wheat (Thomas, 1992). Studies on oat's tolerance to irradiation damage (MacKey, 1954) and chromosome deletion (McGinnis, 1966) showed that many genes with the same or similar function, are carried on more than one chromosome. This level of redundancy in the genome helps to explain why unusual segregation ratios often occur. A 3:1 ratio would be the expected ratio for a single locus, two allele gene, but oat often exhibits 15:1 and 63:1 segregation ratios (Jensen, 1961). Very complex inheritance patterns can arise in such systems, particularly for multi-genic and/or multi-allelic traits. Jensen (1961) also suggested that this genome organization is favourable for the production of novel genes as deleterious effects could be over-ridden by wild type alleles/genes.

Despite the homology of chromosomes within the oat genome, little non homologous chromosome pairing appear to occur. Rajhathy and Thomas (1974) reported the regular formation of 21 bivalent chromosomes at meiosis. Similar pairing patterns have been found to be under genetic control in wheat (Sears and Okomato, 1958; Riley and Chapman, 1958), where a locus on the 5B chromosome restricts pairing to homologous chromosomes.

2.7 Genetic Mapping

Botstein et al. (1980) postulated the use of genetic mapping for genetic counseling in humans. The same principles apply to plant breeding. Using DNA mapping and marker techniques early generation material can be screened based on true genotype rather than phenotypic expression. Genetic mapping involves collecting a set of marker data that shows differences between two (parental) types. Paterson et al. (1991) stated that a genetic marker must distinguish between two parents and be accurately reproduced among progeny. Linkage between markers is determined based on frequency/likelihood of changes between parental types.

Genetic maps allow linkages to be established between genetic markers and genes of interest (Tanksley et al., 1989). Such maps are particularly useful in resolving quantitative traits (polygenically inherited) (Paterson et al., 1988). The use of genetic maps allows traits that can only be measured as the sum of multiple gene effects, to be deciphered into the discrete genes involved. This allows breeders to better understand the genetic factors underlying quantitative traits and to employ more systematic approaches to their selection programs.

2.8 Genetic Marker Systems

A number of types of data can be used as markers. The earliest markers were morphological markers such as colour, dwarfism, or albinism. Morphological markers have a number of weaknesses. They tend to be very limited in number, suffer from genetic interactions (such as dominance and/or epistasis), and often carry deleterious effects (like albinism) (Tanksley et al., 1989). Modern genetic markers now tend to fall into two classes: protein based markers and DNA based markers.

One type of protein-based markers is known as isozymes. Proteins from tissue extracts are separated on starch gels using electrophoresis. A variety of staining techniques are used to visualize the proteins. Small differences in related proteins often result in a change in their mobility in the gel and cause shifts in the banding pattern (Tanksley and Orton, 1983).

DNA based markers come in a variety of forms, but all indirectly analyze DNA sequence differences. Three of the most common markers currently used are restriction fragment length polymorphisms (RFLP), randomly amplified polymorphic DNA (RAPD), and amplified fragment length polymorphisms (AFLP).

The RFLP technique uses restriction enzymes to digest genomic DNA. DNA fragments (restriction fragments) are separated electrophoretically on agarose gels. Radioactive DNA probes are hybridized to restricted fragments where homology exists between the probe and genomic DNA (Botstein et al., 1980). Banding patterns are visualized by exposing autorad film. Polymorphisms arise through single base changes that destroy or create restriction sites and through the insertion or deletion of DNA within a fragment sufficient to alter its mobility. RFLP markers have two strong advantages. One, is their ability to detect multiple alleles at a single

locus (co-dominant); and two, is their reproducibility (Waugh and Powell, 1992). A number of crop species have already been mapped using RFLP markers (Penner, 1996).

Williams et al. (1990) described a method of detecting DNA polymorphisms using polymerase chain reaction (PCR) and arbitrary primers. This technique uses ten base primers that randomly anneal to genomic DNA during the PCR reaction. A very similar method was developed by Welsh and McClelland (1990) using longer primers and radio-labelling for detection. The technique of Williams et al. (1990) has achieved wider acceptance. Amplification (copying by polymerase) occurs wherever primers anneal. Several cycles of temperature and time control denaturing, annealing, and extension within a thermocycler resulting in high copy numbers of amplified fragments. Much like RFLP, RAPD polymorphisms arise from single base changes in primer annealing sites (as opposed to restriction sites) either preventing a primer from annealing or creating new sites. Deletions, insertions or other substantial changes within fragments will also create polymorphisms. These randomly amplified polymorphic DNA (RAPD) fragments can be separated on gels using electrophoresis. RAPDs lend themselves to high throughput situations better than RFLPs, don't use radioactive labelling, and have been shown to be highly polymorphic in many different species (Penner, 1996). RAPDs do have two main drawbacks. Reproducibility both within and between labs has had to be addressed and the markers are mostly dominant in nature (heterozygotes will show the same banding pattern as the homozygote with the band) (Penner, 1996).

Amplified fragment length polymorphism* (AFLP) analysis is a new technique first described by Vos et al. in 1995. AFLP analysis combines different aspects of the RFLP and RAPD techniques described earlier. In AFLP analysis, genomic DNA is digested with restriction enzymes that leave sticky ends, adapters are ligated onto the digested fragments, fragments are amplified using primers matching the adapter sequence plus one to three bases into the genome. Amplified fragments are separated using polyacrylamide gel electrophoresis and visualized using either radio-labeling (Vos et al, 1995) or silver staining (Bassam et al., 1991). AFLPs are highly polymorphic and each reaction yields a high number of bands, both monomorphic and

---

* AFLP is a registered trademark of Keygene Inc.

polymorphic, making it very well suited for the rapid creation of large data sets (van Eck et al., 1995) although most of the bands are dominant. The technique is also described by Lin and Kuo, 1995.

Molecular techniques such as RFLP, RAPD, and AFLP have been successfully used to identify markers linked to traits of interest in a number of cereals and other crops. RFLP had been a widely used technique for over 20 years and many markers and maps have been created in many species. Maps containing RFLP loci have been created in maize (Edwards et al., 1992), soybean (Keim et al., 1990), barley (Kleinhofs et al., 1993) and oat (O'Donoughue et al., 1995). Several loci in each of these studies have been linked to traits of interest. Although RAPD markers have been included in mapping projects, several markers linked to specific traits have also been identified. Disease resistance genes for stem rust (Penner et al., 1993a) and crown rust (Penner et al., 1993b) have been linked to RAPD fragments. A RAPD marker linked to daylength insensitivity has also been found (Wight et al., 1994). AFLP markers have been used to expand existing maps in potato (van Eck et al., 1995), explore genetic diversity in sunflower (Hongtrakul et al., 1997), and characterize quantitative traits in barley (Hayes et al., 1997).

2.9 Existing Oat Maps

Three existing oat maps have been published. Two of them are maps of diploid relatives of *Avena sativa* L. (O'Donoughue et al., 1992; Rayapati et al., 1994). The relationship between *Avena sativa* and these diploid species is still not well understood (Thomas, 1992). In 1995 the first map of cultivated oat was published by O'Donoughue et al. This study mapped a cross between two species of cultivated oat – *Avena sativa* L. and *Avena byzantina* C. Koch. The map consisted of a total of 561 RFLP (537), RAPD (7), isozyme (10), seed coat (3), STS (3), and morphological (1) markers. The markers formed 38 linkage groups covering 1482 cM. Based on their results O'Donoughue et al. (1995) estimate the total hexaploid oat genome to be 2932 cM, indicating that the existing map covers about half of the genome.

2.10 Identification of Quantitative Trait Loci

The basic concept behind identifying quantitative trait loci (QTL) linked to markers is simply to separate classes based on phenotype (marker score) and determine if there is a significant difference in phenotype score between classes (Falconer and MacKay, 1996). There are two basic approaches to detecting QTLs: single marker analysis and interval mapping. Linkage to single markers is detected using regression analysis between genotype classes and phenotypic score. Partitioning the amount of variance explained by a QTL identified in this manner is complicated by the recombination between the marker loci and the QTL. An estimate of the genotypic value can be calculated but the true effect of a QTL cannot be determined unless its distance from the marker is known. As the recombination frequency increases so will the number of false negative and false positive scorings. This leads to underestimation of the genetic effect of a QTL (Lander and Botstein, 1989).

In order to establish this linkage distance and develop a reasonable estimate of genetic effect interval mapping analysis can be used. Interval mapping requires the use of a genetic linkage map with known distances between markers. Regions between markers are tested for different recombination frequencies and the resulting changes in variance and class means (Falconer and Mackay, 1996). As the distance from a marker to a QTL increases so does the possibility of a recombination event between them. If recombination occurs in an individual, that individual is put in the wrong class based on marker score. Interval mapping calculates maximum possible values for class means and genetic variation for a QTL as it "moves" between two markers. Changing distance between two markers with a known linkage allows the calculation of a recombination frequency between the markers and the potential QTL site. Using this arbitrary recombination frequency allows new calculations of genotypic class means and variances.

To determine if the new class means and variances are the result of the presence of a QTL, a likelihood estimate for the calculated results in the presence of a QTL versus the likelihood for achieving the same result in the absence of a QTL is determined. This likelihood estimate is called the LOD score (Lander and Botstein, 1989). The calculated LOD score is compared to a threshold test statistic. If the calculated LOD is greater than the test statistic then

a QTL at that locus is inferred. The test statistic should be calculated based on random permutations of the data set being used (Doerge and Churchill, 1994). By randomly shuffling marker and phenotypic data and then calculating LOD scores for each permutation, the averaged threshold should then take into account any chance associations between the marker and phenotypic data. Threshold statistics can be influenced by several experiment-specific factors which standardized thresholds cannot do. These factors include sample size, genome size, marker density, proportion of missing data, and segregation distortion (Doerge and Churchill, 1994).

The concepts and methodologies outlined above have been used successfully to identify many QTLs. Paterson et al. (1989) used a RFLP linkage map of tomato to locate six QTLs affecting fruit mass, four QTLs for soluble solids and five QTLs for fruit pH. They also suggested that the techniques (based on the above concepts) they used were broadly applicable to all higher plants and animals. In barley, Hayes et al. (1997) located QTLs for six traits in eleven different populations. Siripoonwiwat et al., (1996) used the oat map published by O'Donoughue et al. in 1995 to identify QTLs affecting grain yield, test weight, plant height, groat percentage, straw yield and days to maturity.

3.0 INHERITANCE OF OIL CONTENT AND COMPOSITION IN OAT (*Avena sativa* L.)

3.1 ABSTRACT

Total oil content and fatty acid composition are two important factors in food value and storage stability of oat (*Avena sativa* L.) and oat foods. The objective of this study was to examine the basic factors underlying oat oil inheritance through phenotypic assessment. Two hundred and twenty three lines from an F7 population of a cross between the high oil cultivar AC Marie (approximately 5 percent oil) and the low oil cultivar Cascade (approximately 3.5 percent oil) were analyzed for fatty acid composition. Oil was extracted using petroleum ether (35 to 60°C bp) as the solvent in a Labconoco lipid extractor. Fatty acid composition was determined using gas chromatography of fatty acid methyl esters (FAME). The mean value for palmitic acid was 19.5 percent, 41.8 percent for linoleic, and 30.6 percent for oleic. Lesser amounts of stearic acid (1.4 percent) and linolenic acid (1.5 percent) were found. Correlations (P=0.01) between total oil content and palmitic, oleic and linoleic fatty acids were found to be 0.92, 0.94, and 0.97 respectively. Results indicated that a single gene was responsible for control of each fatty acid and total oil content. Near infrared reflectance (NIR) spectroscopy was used to analyze an 80 line subpopulation of the AC Marie/Cascade lines. These lines were grown at Glenlea (1996) and Saskatoon (1997). The means for each year were not significantly different (P=0.01). The means for the 80 line population over the two site years was 3.95 percent total oil and 5.20 percent groat oil. Broad sense heritability for total oil was 0.19 and for groat oil (total oil content corrected for percent hull content), 0.12. These results also indicated that a single gene was responsible for the majority of the total oil content differences in the AC Marie/Cascade population. The results of this study will help to increase overall understanding of oil content inheritance in oat.

3.2 INTRODUCTION

The genetic control of oil content and composition have been characterized for a variety of crops. Oil content and composition of grains are important factors in food/feed quality and the stable storage of processed grain products. In oat (*Avena sativa* L.), a combination of high oil content and unstable constituent fatty acids may contribute to rancidity problems in oat food products. Oil content is also important in relation to total energy and fat content of foodstuffs and feed. Although stability of stored products can be improved by a variety of treatments, modifying the constituent fatty acids and total oil content in the seed itself offers an attractive option for end users.

Total oil content is a function of the accumulation of individual fatty acids, typically in the form of triglycerides (Stymne and Stobbart, 1987). The biosynthesis of lipids is a very complicated biochemical process. Many of the steps in biosynthesis have been characterized but gaps still exist and very few regulatory processes have been discovered (Stumpf, 1987; Jaworski, 1987). This indicates that the overall genetic control of oil content and composition could be quite complex.

Brown and Craddock (1972) examined over 4000 accessions from the world oat collection and found that total oil content ranged from 3.1 percent to 11.6 percent. Levels in cultivated oat were found to be similar – 4 percent to 11 percent (Sahasrabudhe, 1979; de la Roche et al., 1977) which indicates that substantial variation exists within current oat populations. Earlier studies on oil inheritance have shown that total oil content is a highly heritable trait (from 63 percent to 93 percent) and has potential for exploitation in breeding programs (Baker and McKenzie, 1972; Frey et al., 1975; Thro and Frey, 1985). These reports agree that inheritance of total oil content is likely polygenetic and the result of additive gene action, although Karow (1980) reported a heterotic interaction in a cross between the cultivars Dal (high oil) and Sauk (medium oil).

Investigations into the fatty acids composition of oat cultivars have been carried out in a number of studies including those reported by Youngs et al. (1977), Sahasrabudhe (1979), de la Roche et al. (1977), and Karow (1980). These studies found that the major constituent fatty acids

were the saturated acids palimitic (16:0) and stearic (18:0) and the unsaturated fatty acids oleic (18:1), linoleic (18:2) and linolenic (18:3). These studies reported that stearic and linolenic acids were present in fairly small amounts and that palmitic, oleic, and linoleic acids make up over 95 percent of the total oil content. Trace amounts of other fatty acids were also found to be present in some of the analyses.

Heritabilities for each fatty acid were reported to be 91 percent for palmitic, 97 percent for oleic, and 96 percent for linoleic acids (Youngs and Puskulcu, 1976). These were substantially higher than those reported by Karow (1980), which were 66 percent for oleic acid and 80 percent for linoleic acid. De la Roche et al. (1977) reported that total oil content was correlated to several of the constituent fatty acids. Total oil content was positively correlated with oleic acid and negatively correlated with both palmitic and linoleic acids. These findings were confirmed by Youngs and Puskulcu (1976) and Sahasrabudhe (1979). Genetic control of fatty acid composition appears to be dependent upon the cross(es) being examined. Additive control of palmitic, stearic, and linolenic acids was suggested by Karow (1980), while partial dominance influenced oleic and linoleic acid inheritance.

The goal of this study was to correlate differences in total oil of the cultivars AC Marie (high oil) and Cascade (low oil) to differences in the relative amounts of specific fatty acids. An increase in the understanding of the amount and type of phenotypic variation for oil content is an important step in the identification of molecular markers for the genes involved.

3.3 MATERIALS AND METHODS

3.3.1 Fatty Acid Composition Study

An F_7 population consisting of 223 bulked lines from a cross between the high oil cultivar AC Marie and the low oil cultivar Cascade was grown in three replicates at three sites (Glenlea, Portage la Prairie and Winnipeg) in Manitoba in 1995. The plots were planted in a randomized complete block design. Eight samples of each parent were also included as treatments. Seed was planted in 1.5 m rows with 22 cm spaces between rows. Fall rye was planted between each treatment row to provide additional physical separation.

Each site was treated with a pre-emergence application of glyphosate (1 L/acre) and a post-emergence application of chlorsulfuron at the recommended rate. Foliar applications of the systemic fungicide propiconazole (0.2 L/acre), were used as required to control crown rust (*Puccinia coronata* f. sp. *avenae* Eriks.) and stem rust (*Puccinia graminis* Pers. f. sp. *avenae* Eriks. & Henn.). Seed from complete rows was harvested and bulked for replanting and analysis.

In 1996, the same field program was used but the population was only planted in the Portage la Prairie and Glenlea locations. The rye spacer rows were also omitted in the 1996 field trials.

To determine the total oil content of these lines, approximately 20 g of whole oats were ground to a coarse flour texture. A 4 g sample was placed in a Labconoco lipid extractor where it was processed for 7 hours using petroleum ether (35 to 60°C bp) as a solvent. The solvent was evaporated off and the resulting oil weighed to determine total oil content.

Fatty acid composition was determined by converting a 50 µl oil sample to fatty acid methyl esters using 2 ml of a solution of 2 percent concentrated sulfuric acid in methanol (v:v) as a catalyst. This was allowed to react at room temperature for several hours. The resulting methyl esters were analyzed in a Hewlett Packard 5890 gas chromatograph using a Supelco Wax 10 column (15 m x 0.32 mm). Peaks were identified by comparison to known standards and the relative amounts of individual fatty acids were determined by comparing peak areas.

Total oil content and relative fatty acid composition data were collected and compiled in Quattro Pro V. 6.0 (Novell, 1995). This software was used to calculate means, variances, standard deviations, and frequencies for all data. Regression analysis was also used to determine relationships between fatty acids and total oil content.

3.3.2 Total Oil Content Study

During the fall and winter of 1995-96, a subpopulation of the AC Marie/Cascade population was created by selecting 20 high oil lines, 20 low oil lines and another 40 lines at random. This subpopulation was derived using a generation of single seed descent. Five seeds from each line were planted in 10 cm pots and grown to maturity in the greenhouse. A randomly selected single plant from each pot was harvested. This seed was used to plant two locations of

80 lines plus both parents at Glenlea and Portage la Prairie although, due to an error in field maintenance, the location at Portage was destroyed. The material was planted in a non-random design. The field techniques used were identical to those used for the 223 line population in 1996. The Glenlea location was harvested and kept for replanting and analysis. This 80 line subpopulation was also grown in 1997 at four locations – Glenlea, MB, Morden, MB, Saskatoon, SK, and Aberdeen, Idaho. Only one of these four locations was available for analysis. Dry conditions in Glenlea resulted in insufficient seed quantities for a number of lines. Flooding at the Morden site damaged early emergence causing this site to be abandoned. The Aberdeen material could not be returned to Canada due to import restrictions caused by a Karnal bunt outbreak in the US.

Samples from Glenlea - 1996 and Saskatoon - 1997 were analyzed for total oil content using Near Infrared Reflectance spectroscopy (NIR). A volume of whole grain was placed in a NIR Systems model 6500 scanning monochromator. The samples were double scanned and total oil content was automatically calculated using a calibration curve created specifically for oat. The NIR automatically determined an average of 64 scans for each sample.

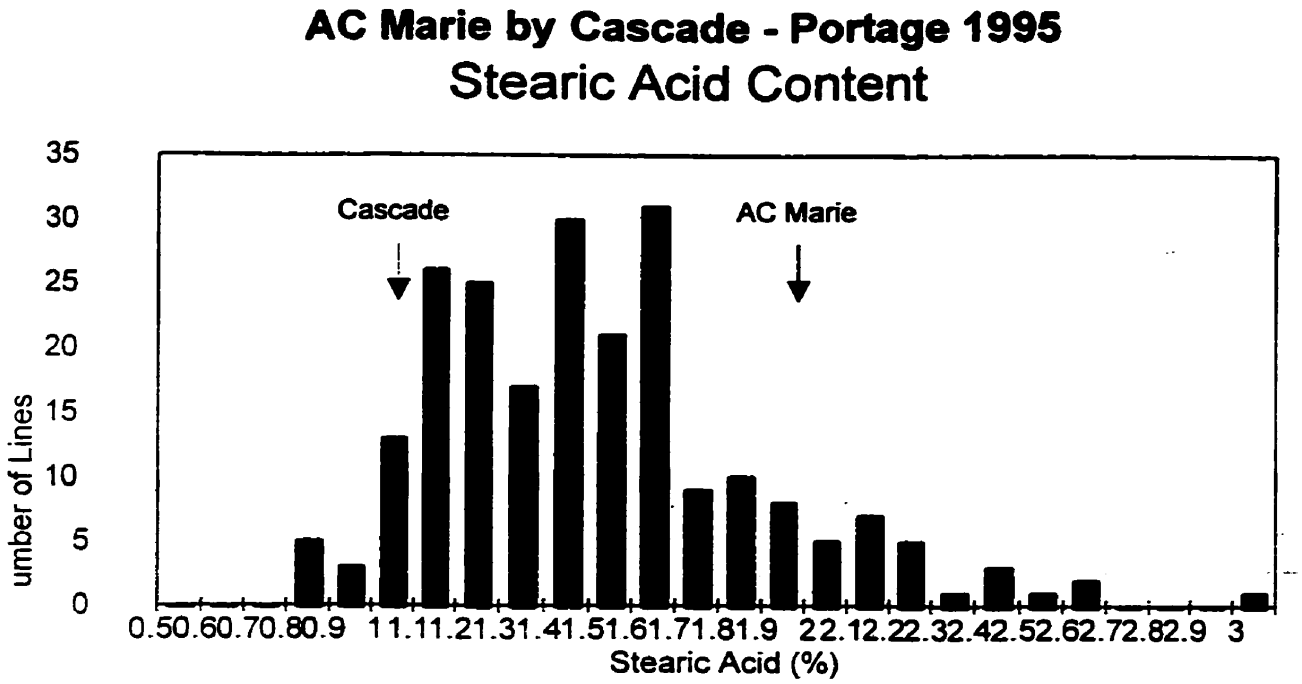
3.4 RESULTS

3.4.1 Fatty Acid Composition Study

Mean values and ranges for total oil and fatty acid composition of AC Marie, Cascade and the 223 F_7 lines grown at Glenlea in 1995 are shown in Table 3.1. Total oil content and relative amounts of fatty acids present for each of the segregants is shown in Figures 3.1 through 3.3. These distributions show a wide range of values for relative content of each of the fatty acids. In the distributions for palmitic, oleic, and linoleic acids, several individuals exhibit values beyond the parental range. For total oil content, only one individual is markedly outside of the parental values. This indicates trangressive segregation suggesting that more than one gene is involved in the accumulation of fatty acids and differing alleles are found in the two parents. Linolenic and stearic acids appear to be controlled by similar genetic factors in the two parents as values only vary within the parental range.

Table 3.1. Total oil content and specific fatty acid means, expressed as a percentage, for the parental cultivars AC Marie and Cascade, and 223 segregating lines. Means are plus or minus the standard deviation indicated. Values in parentheses indicate the range of values.

Oil Component	AC Marie	Cascade	Segregants
Total Oil	7.1±0.26% 282.4 mg ±10.3	3.3±0.4% 130.6 mg ±16	4.64±1.0% 185.5 mg ±40.2
Palmitic Acid	19.3±1.5% (17.287-22.142)	19.5±1.2% (17.796-21.143)	19.5±1.7% (15.646-25.502)
Stearic Acid	2.1±0.3% (1.709-2.622)	1.1±0.3% (0.708-1.580)	1.4±0.5% (0.724-6.505)
Oleic Acid	34.7±1.9% (32.455-37.852)	27.8±3.1% (20.590-31.129)	30.6±3.4% (20.995-38.062)
Linoleic Acid	39.8±1.1% (37.937-41.752)	43.4±2.3% (38.777-45.765)	41.8±2.2% (34.263-47.025)
Linolenic Acid	1.2±0.1% (1.018-1.258)	1.9±0.3% (1.466-2.26)	1.5±0.3% (0-2.271)



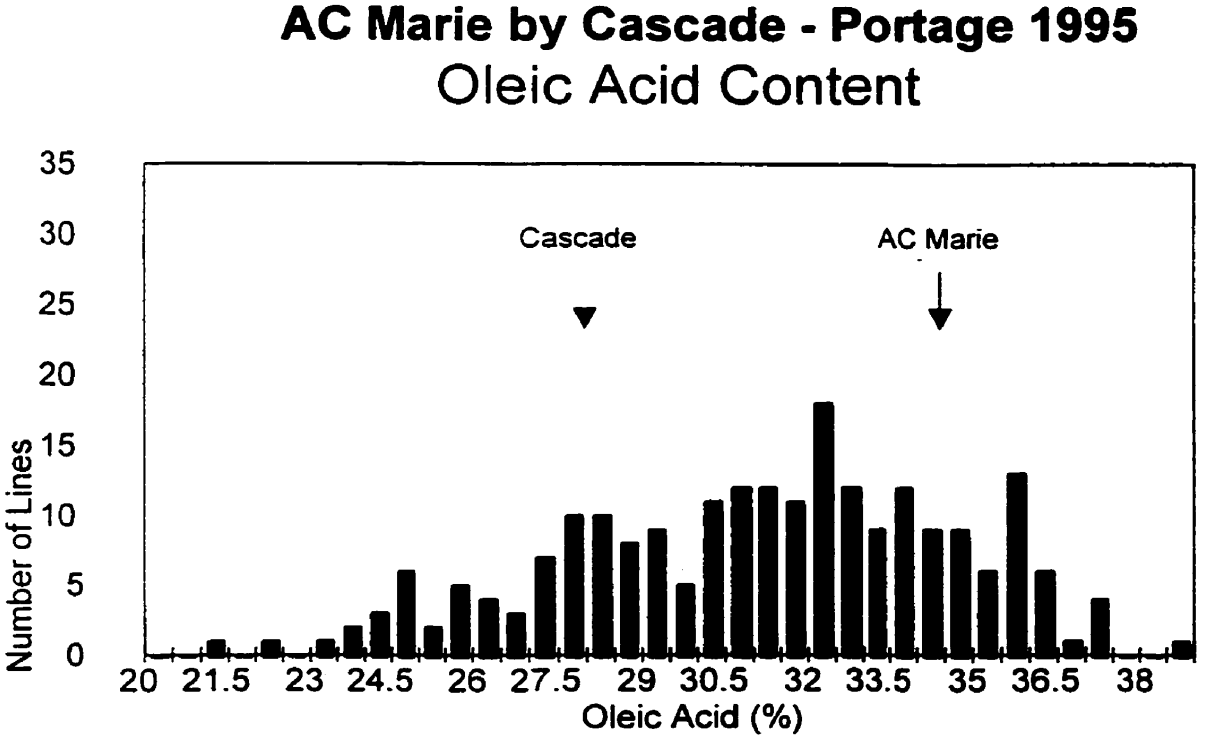
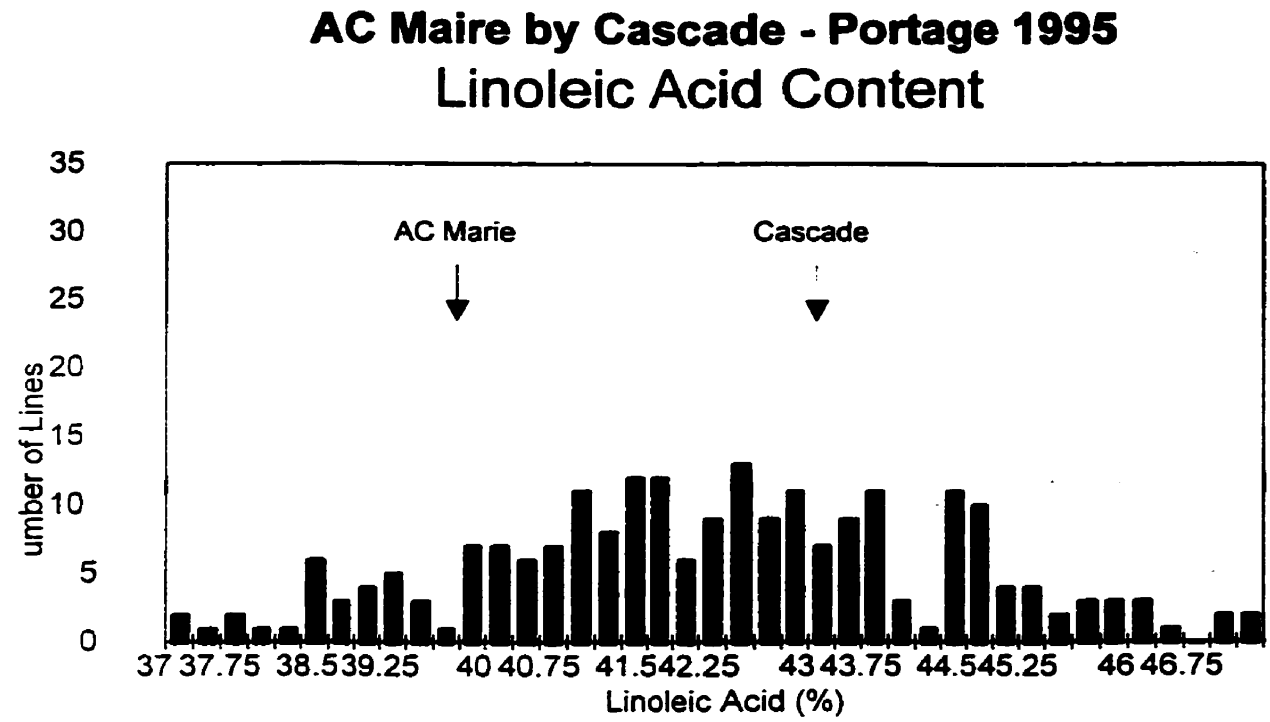


Figure 3.1. Frequency distributions of stearic and oleic acid content for 223 F_7 lines of AC Marie/Cascade grown at Portage la Prairie, MB in 1995. Fatty acid values are relative content shown in percent.



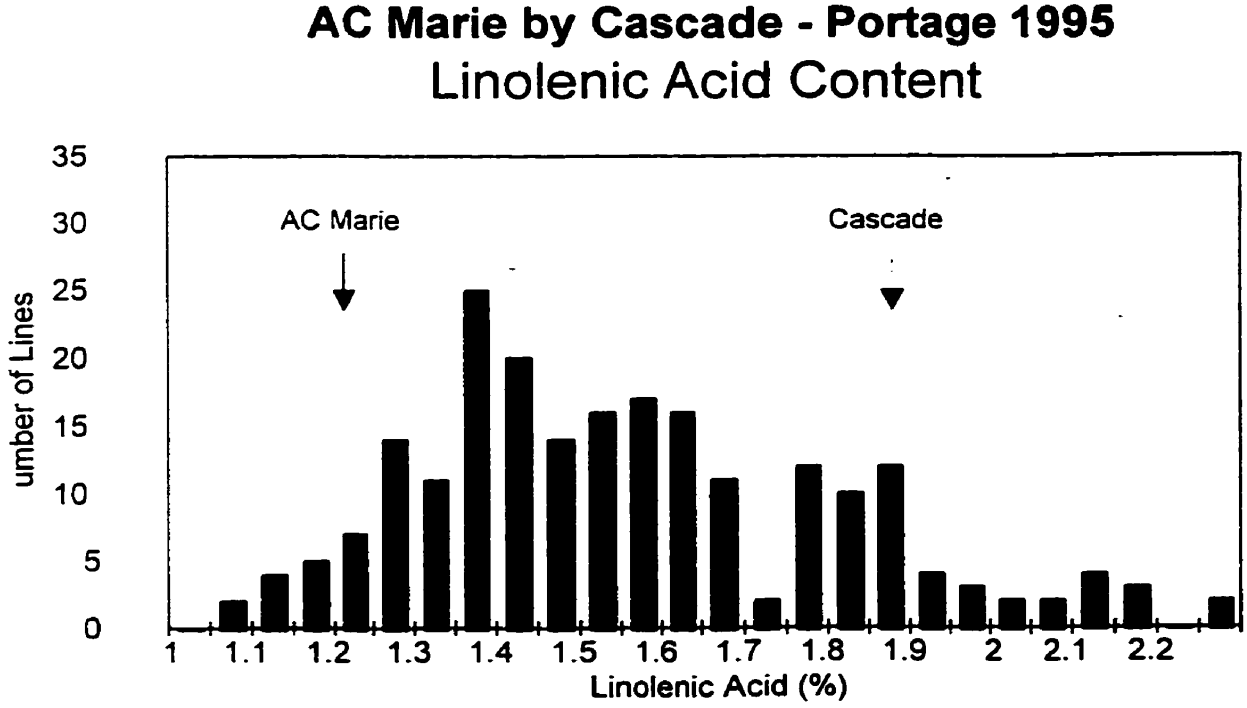
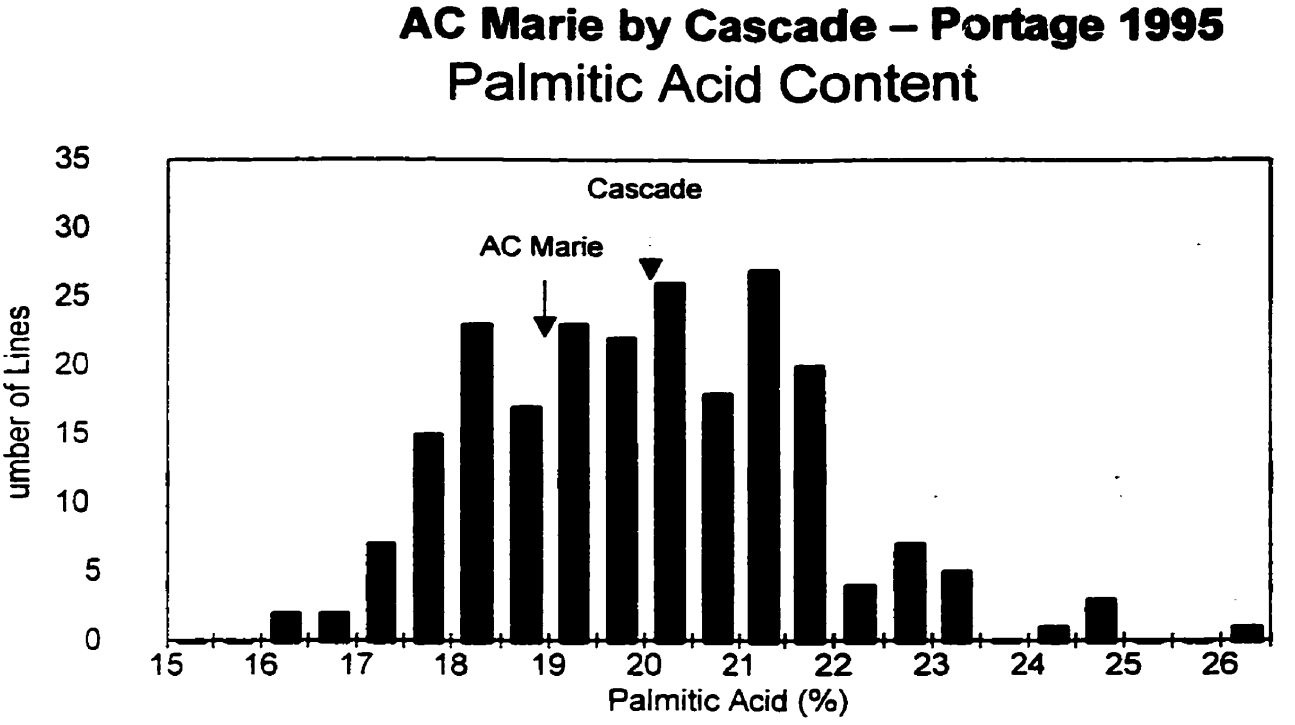


Figure 3.2. Frequency distributions of linoleic and linolenic acid content for 223 F_7 lines of AC Marie/Cascade grown at Portage la Prairie, MB in 1995. Fatty acid values are relative content shown in percent.



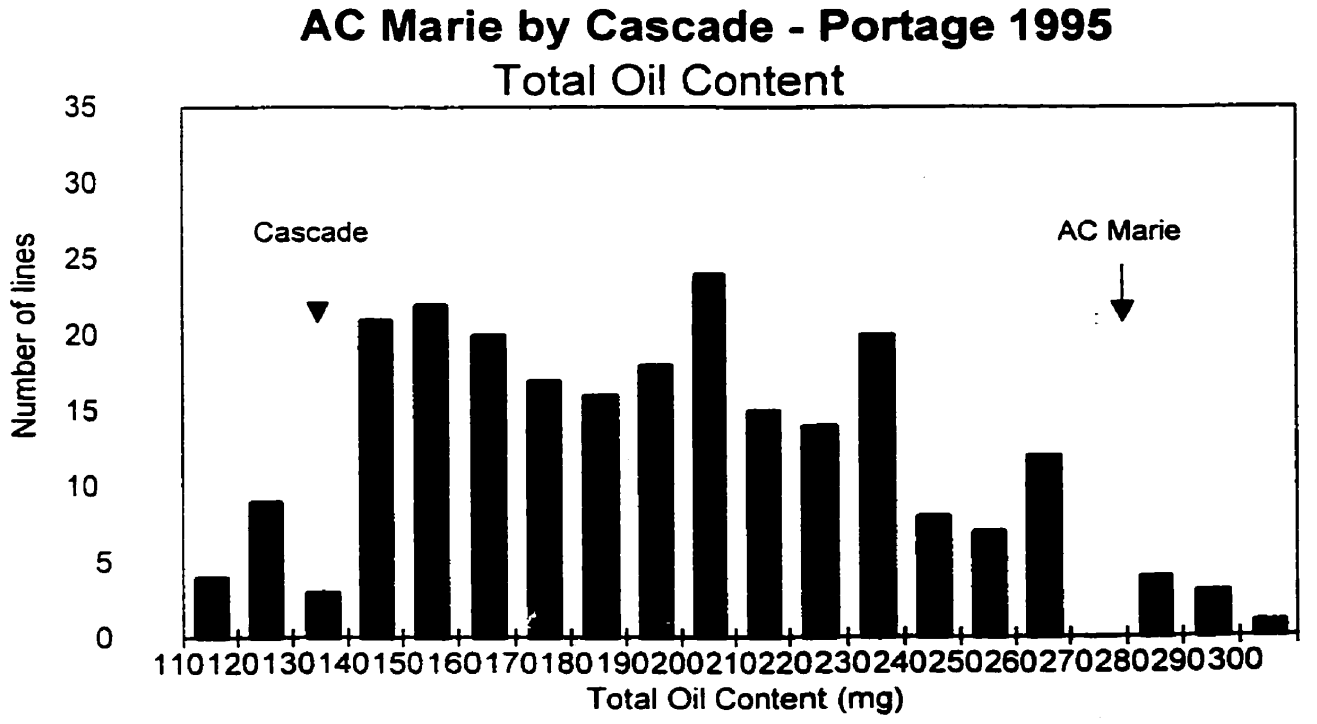


Figure 3.3. Frequency distributions of palmitic acid and total oil content for 223 F_7 lines of AC Marie/Cascade grown at Portage la Prairie, MB in 1995. Fatty acid values are relative content shown in percent. Total oil content is shown in mg of oil extracted from a 4 g sample of ground oat.

Regression analysis of oil content and composition based on actual mg of oil/fatty acid revealed significant ($P < 0.01$) positive relationships between total oil and each of the fatty acids (Table 3.2). The correlation between total oil content and palmitic, oleic and linoleic acids is very strong (0.92, 0.94 and 0.97 respectively). The correlations between total oil and stearic acid (0.65) and linolenic acid (0.33) are still significant if not as strong. Linolenic and stearic acids show the lowest correlation values with other fatty acids. The highest correlation between two fatty acids exists between oleic and linoleic acids at 0.93.

3.4.2 Total Oil Content Study

Mean values for total oil content and groat oil content for the parental cultivars AC Marie and Cascade and the 80 segregants of the derived subpopulation are shown in Table 3.3. Groat oil percentage was calculated by dividing total oil content by one minus the percent hull. Percent hull was determined by NIR. Frequency distributions for the segregants show a slight skewness toward low oil content in both 1996 (Glenlea) and 1997 (Saskatoon) (Figure 3.4). Similar trends are shown by the frequency distributions for groat oil content (total oil content corrected for percent hull) (Figure 3.5). No transgressive segregation is indicated by the NIR data.

The means between the two location years have a difference of 0.31 percent but Glenlea 1996 is not significantly different from the mean for Saskatoon 1997 ($P = 0.05$). Another indicator that the data between the two years is similar is Spearman's rank order coefficient. This statistic measures how well the order of the lines, with respect to total oil content, was conserved. A coefficient of one means the order is completely the same, while a coefficient of zero would mean the order is not conserved at all. The coefficient for total oil content was calculated to be 0.86.

Broad sense heritability of total oil content was calculated using the regression method where $h^2=(x)(b_{xy})/y$ (Mahmud and Kramer, 1951). The mean of the 1996 generation is denoted as x, the regression of 1997 segregants on the 1996 segregants is b_{xy}, and the mean of the 1997 segregants is y. The heritability was found to be only 0.19 for total oil content and 0.12 for groat oil content.

Table 3.2. Correlation coefficients between absolute amounts of fatty acids present in a 223 line segregating population of AC Marie/Cascade grown at Portage la Prairie, MB in 1995.

	Total Oil	Palmitic Acid	Stearic Acid	Oleic Acid	Linoleic Acid	Linolenic Acid
Total Oil	1					
Palmitic Acid	0.92*	1				
Stearic Acid	0.65*	0.77*	1			
Oleic Acid	0.94*	0.78*	0.50*	1		
Linoleic Acid	0.97*	0.85*	0.55*	0.93*	1	
Linolenic Acid	0.33*	0.18^{NS}	-0.03^{NS}	0.39*	0.37*	1

NS Nonsignificant
* Significant at P < 0.01

Table 3.3. Mean values for total oil content and groat oil content for the parental cultivars AC Marie and Cascade and 80 segregating lines. Means are expressed as percentages and are plus or minus the standard deviations indicated. Conversion to groat oil percentage was accomplished by removing the percent hull as indicated by NIR measurement. Means for the two locations are not significantly different (P=0.01).

Location-Year	Line	Total Oil Mean (%)	Groat Oil Mean (%)
Glenlea-1996	AC Marie	4.73±0.07	5.99±0.13
	Cascade	3.31±0.07	4.57±0.06
	Segregants	3.81±0.32	5.11±0.42
Saskatoon-1997	AC Marie	5.11±0.12	6.46±0.16
	Cascade	3.67±0.02	5.08±0.08
	Segregants	4.12±0.40	5.28±0.97

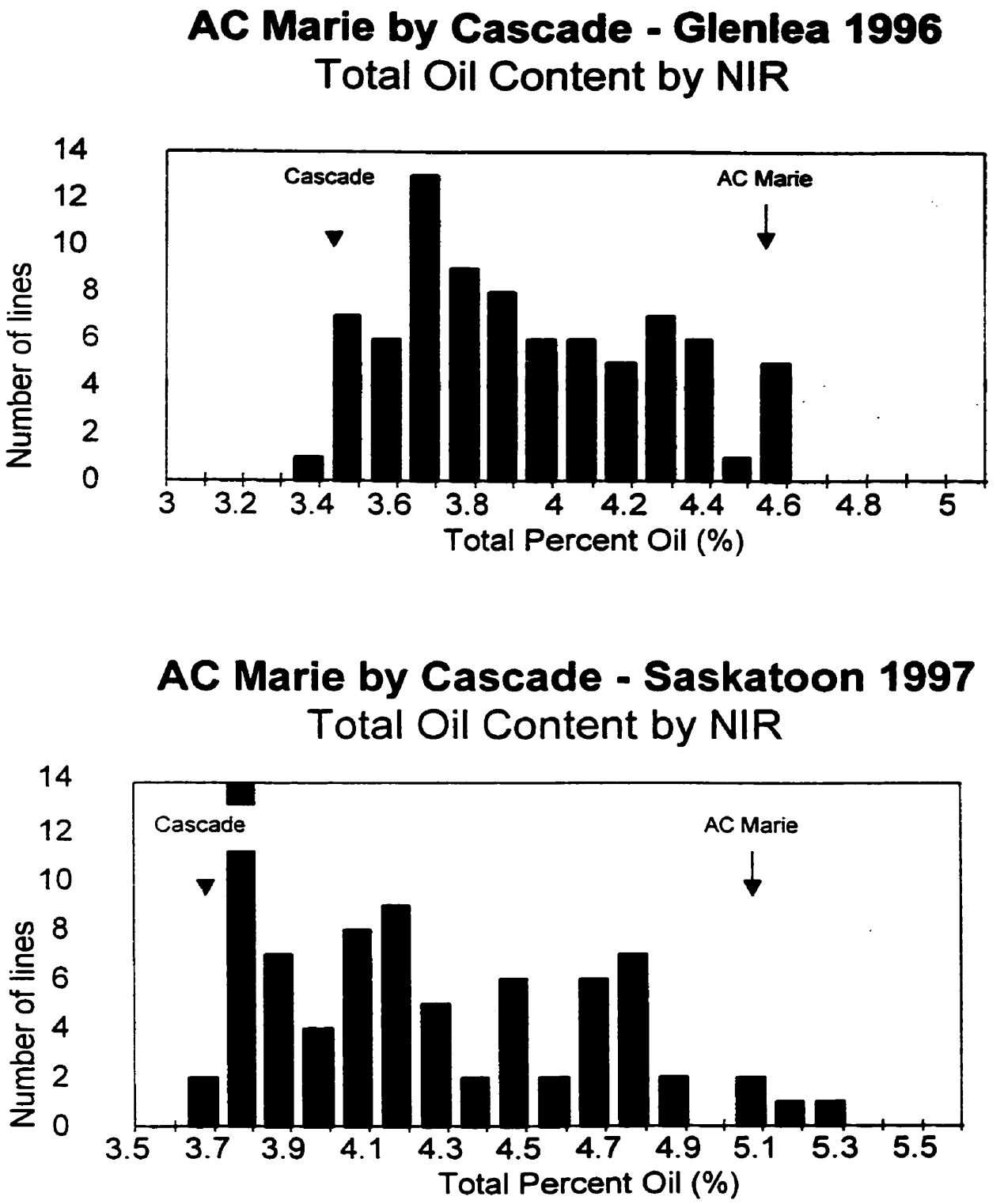


Figure 3.4. Frequency distributions showing total oil content expressed as a percent for 80 F10 and F11 lines of an AC Marie/Cascade population grown in Glenlea, MB in 1996 and Saskatoon, SK in 1997. Oil content was assessed using near infrared reflectance spectroscopy.

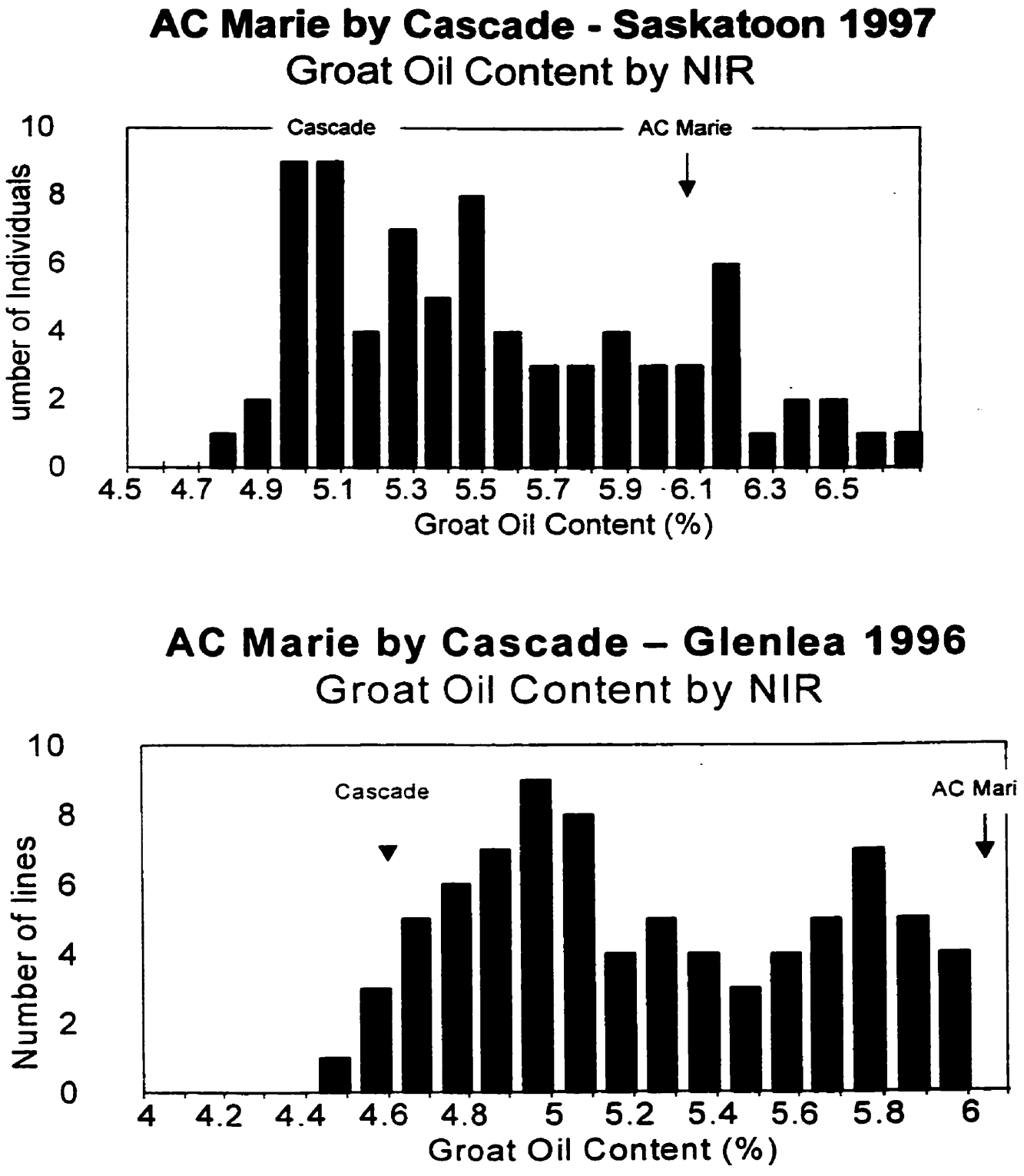


Figure 3.5. Frequency distributions showing groat oil content expressed as a percent for 80 F_{10} and F_{11} lines of an AC Marie/Cascade population grown in Saskatoon, SK in 1997 and in Glenlea, MB in 1996. Oil content and hull content were assessed using near infrared reflectance spectroscopy.

3.5 DISCUSSION

3.5.1 Fatty Acid Composition Study

The goal of this study was to determine the inheritance of a variety of fatty acids and their relationship to total oil inheritance using an advanced generation (fixed) population of lines segregating for oil inheritance. The population chosen for this study was derived from an F_2 bulk and each subsequent generation was also bulked. The result of this treatment was that each F_2 plant that was bulked was either homozygous or heterozygous for any given locus in a 1:2:1 ratio (AA:Aa:aa). As these lines were propagated and bulking continued, loci became fixed, but lines that were derived from individuals that were heterozygous in the F_2 generation consisted of individuals that were homozygous at a locus but the allele ratio of 1:1 (A:a) still existed within the line. This means that by the eighth generation all of the individuals within those lines are virtually homozygous but for each locus there is an equal number of individuals that are fixed for each allele at that locus. The lines of the population as a whole will be one half homozygous and homogeneous and one half homozygous and heterogeneous for any given locus.

This level of heterogeneity within the lines results in a population that is unsuitable for inheritance studies. Each heterogeneous line is actually a composite of individuals fixed for either of the alleles for each locus. The result is that each line more closely resembles a heterozygous individual (overlooking dominance and heterotic effects) rather than an individual that is homozygous for each locus. This means that lines that are heterogenous at a locus will exhibit midparent values for that locus.

Adding an extra level of complexity is the consideration that for each given locus one half of the lines will be heterogeneous. Therefore, even if fatty acid composition is controlled by only a single gene for each major fatty acid and one for total oil accumulation (six genes) only 1.6 percent (0.5^6) of the individuals will actually be homogeneous for those traits. In a population of 223 lines only 3 would be expected to be homogeneous for all of these loci.

Initial inspection of the data from this population indicates that the inheritance of each fatty acid is controlled by multiple genes. Normal distributions are exhibited by each of the fatty acids and total oil content in Figures 3.1 to 3.3. Such smooth curves would usually be associated

with multiple gene inheritance. However, when one considers the heterogeneous nature of this population, it seems more likely that the even distributions are the result of "noise" generated by the heterogeneous lines. Essentially, the difference between parental genotypes is obscured by the presence of a heterogeneous mixture of genotypes in half of the lines.

Each of these curves could be the result of what are essentially three overlapping curves for each trait. Consider, for example, an individual fatty acid with only one gene controlling inheritance. In an advanced generation (approaching complete homozygosity) there would be three "genotypes". Two fixed for parental type (one for each parent) and one a 50:50 mixture of the two parental types or the midparent genotype. If the distributions for the phenotypic measurement of each of the "genotypes" are close enough together, the overlap causes these three distinct curves to appear as one continuous curve. The parental curves each only represent 25 percent of the total and are therefore "shorter" curves while the midparent values would be expected in about 50 percent of the lines making this curve twice at "tall" as the parental ones. This almost ensures that the curves would appear normal.

Closer inspection of the data, keeping the above in mind, would indicate that only a single locus may be responsible for the accumulation of each fatty acid. This does not preclude the possibility of multiple gene control for each fatty acid but the single gene theory is at least as plausible. This theory is also supported by the lack of transgressive segregation for individual fatty acids, at least beyond the ranges shown for the parents (Table 3.1). If multiple loci were involved, greater evidence of transgressive segregation would be expected as favourable alleles or genes were collected and fixed within individual lines.

The amounts and ranges of fatty acids described in this study are consistent with those reported in earlier literature (Youngs et al., 1977, Sahasrabudhe, 1979; de la Roche et al., 1977; and Karow, 1980). This was expected as these studies covered a wide range of germplasm and reported fairly consistent trends in oil composition for oat. The correlations between total oil and each of the fatty acids were found to be significant with very strong correlation between total oil and the major fatty acids (palmitic, oleic, and linoleic) and the weakest between total oil and linolenic acid. These results are similar to those reported by de la Roche (1977) and Karow

(1980), although both of these studies found negative correlations between total oil content and oleic acid. The strong correlations are surprising given the high number of mid-parent type values but it does make sense that as the total amount of oil increases there is also an increase in the absolute amount of each fatty acid.

In order to make more accurate estimates of fatty acid composition, future work could include analysis of individual plants from the bulks. This would make it possible to confirm whether a single gene is affecting the accumulation of a fatty acid or if multiple genes are involved. This would increase the power of the study in two ways. First, it would give an estimate of variability within lines (although it would overestimate variation in heterogeneous lines). Second, it would effectively increase the population size, as each individual could potentially be considered as a separate genotype.

3.5.2 Total Oil Content Study

The generation of the 80 segregant subpopulation by selecting single individuals from the original lines was designed to alleviate the heterogeneity issues that arose in the 223 line population. Selecting a single seed as the source for each of the 80 lines should have resulted in 80 homozygous and homogeneous lines segregating for total oil content.

The NIR derived oil content data (Figure 3.4) show a slight skewness toward low oil content but there is a consistent "dip" in the middle of the curve, indicating a possible bimodal distribution. When the data are corrected for hull content (oil content/(1 – percent hull)), a potential bimodal distribution becomes more noticeable (Figure 3.5), particularly for the Glenlea 1996 data set. A bimodal distribution is indicative of a single gene controlling total oil content rather than multiple gene inheritance as was first indicated by the fatty acid composition study.

Correcting oil content for the presence of the hull (generating groat oil content) is valid as the hull content in the AC Marie/Cascade population can range between 22 to 33 percent (Ronald, 1997). This population is also segregating for hull percentage and the magnitude of this variation could easily obscure differences in oil content between lines which only range between about 3.3 to 5.5 percent (total oil content).

A number of different modes of inheritance have been suggested by other researchers (Baker and McKenzie, 1972; Frey et al., 1975; Thro and Frey, 1985) but the data from this study indicates that a single gene dominates oil content inheritance in the cross AC Maire/Cascade. This means that selection for oil content in this cross should be successful but would not likely yield increases higher than the high oil parent (AC Marie) or lower than the low oil parent (Cascade).

The low broad sense heritabilities reported in this study are in direct contrast with those reported in earlier studies (as above). This is due in part to the use of very late generation material (F_{10} and F_{11}) and the limited scope of this study. Results reported by other researchers were primarily done on early generation material (F_2 to F_4) where segregation was still taking place.

The low heritability result is contrasted by the high Spearman's rank order coefficient. This statistic indicates that rank order of the lines was largely conserved between the two site years for the 80 line subpopulation (the high lines stay high and the low lines stay low). The conservation of rank order infers that oil content is consistently inherited across generations indicating that it is actually quite heritable.

Despite the low heritability evidenced in this cross, successful selection for oil content (either higher or lower) should be attainable since only one gene appears to be controlling inheritance of total oil content. Future work could include development of molecular markers for oil content genes (such as the one in the AC Marie/Cascade cross). The 80 line subpopulation is well suited for molecular studies because each line should be fixed and homogeneous.

4.0 A MOLECULAR LINKAGE MAP OF OAT (*Avena sativa* L.) USING AFLP* MARKERS

4.1 ABSTRACT

A linkage map of the cross between the oat (*Avena sativa* L.) cultivars AC Marie and Cascade was created using amplified fragment length polymorphism* (AFLP) analysis. The cross had a good level of polymorphism, with a mean number of polymorphisms per gel of 14.1. A total of 591 markers were identified. One hundred and seventeen of these markers were not segregating at a 1:1 ratio and were discarded from the data set. The remainder were organized into 35 linkage groups using the software packages JoinMap and MapMaker. Twenty-two markers were unlinked and the overall genome size was estimated to be approximately 2900 cM. The map covered a total of 1077 cM. This linkage map is useful for identification of quantitative trait loci, increasing knowledge of oat genome structure and provides a framework for the construction of a more detailed map of *Avena sativa*.

* AFLP is a registered trademark of Keygene Inc.

4.2 INTRODUCTION

Identifying genetic markers for complexly inherited traits can be a difficult and time consuming process. Map based marker identification has been successfully used in the identification of a number of traits in other cereal crops (Heun et al., 1991; Rayapati et al., 1994; Tanksley et al., 1989). Despite the importance of oat as a crop, mapping in oat has lagged behind wheat and other cereals.

Oat mapping studies have been conducted on other *Avena spp* (O'Donoughue et al., 1992; Rayapati et al., 1994 and O'Donoughue et al., 1995) but none of these studies were conducted on *A. sativa* by *A. sativa* crosses. These maps were constructed using a variety of markers systems including RFLP, RAPD, isozymes, morphological markers and others. These maps have not yet achieved full coverage of the genome.

Oat has a large nuclear genome that is organized into A, C and D genomes. *Avena sativa* is a hexaploid oat that has a chromosome number of n=3x=21 (Rajhathy and Thomas, 1974). One oat genome contains approximately 11.3 billion base pairs, or about 23.4 pg per complete nuclear genome (Arumuganathan and Earle, 1991). This compares to about 16 billion base pairs for wheat and only 4.9 billion base pairs for barley. The large number of linkage groups (21) and an estimated size of 2900 cM (O'Donoughe et al., 1995) indicate that a large number of markers would be needed to create even a skeletal map.

The large size of the genome together with its hexaploid state make oat (and other polyploids such as wheat) difficult to map. Three complete sub-genomes exist within the oat genome. This means that there are a large number of linkage groups and that, for a given loci, there could be one, two, or three homologous copies. The possibility of a single marker being associated with more than one homologous linkage group (chromosome) could create false linkages. The size of the genome also increases the random chance that two markers from unlinked loci will co-migrate, obscuring marker data and increasing the number of false positives in scoring marker data.

Amplified fragment length polymorphism* (AFLP) analysis has been shown to generate large numbers of polymorphic markers, even in large genomes with large segments of repetitive DNA and polyploid structure (Vos et al., 1995; Lin and Kuo, 1995; Hongtrkul et al., 1997; Hayes et al., 1997). AFLP analysis can generate a high number of polymorphisms per run in a wide range of genome sizes and types. The ability of the AFLP technique to detect many polymorphisms across a large number of individuals makes it an ideal tool for rapidly creating maps of large genomes.

The objective of this study was to build a linkage map of the cross AC Marie/Cascade. A genetic map would be valuable in identifying markers for traits of all kinds, assisting in the development of marker assisted selection techniques for oat and further the understanding of patterns of inheritance in *Avena sativa*.

4.3 MATERIALS AND METHODS

4.3.1 Mapping Population

During the fall and winter of 1995-96 a mapping population was created by selecting 20 high oil lines, 20 low oil lines and 40 lines at random from an existing population of 223 F_7 segregating lines of an AC Marie/Cascade population. The original population was derived through F_2 bulking. Due to the way it was created, the F_7 population consisted of lines that were essentially homozygous but, for a given locus, 50 percent of the lines were heterogeneous – a mixture of individuals fixed for either parental allele. The subpopulation was derived by selecting a single seed from each of the 80 lines described above and multiplying them in the greenhouse for one generation. The resulting seed was bulked and used to continue propagating the line for phenotypic assessment in the field and as a mapping population. The resulting population was both homozygous and homogeneous.

Five seeds from each line of the mapping population plus five seeds of each of the parents, AC Marie and Cascade, were grown in 10 cm pots in the greenhouse. At the one to two leaf stage tissue samples for each line were collected from each plant, bulked and stored in liquid nitrogen. Samples were lyophilized and then stored at -20°C prior to DNA extraction.

---

* AFLP is a registered trademark of Keygene Inc.

4.3.2 DNA Extraction

DNA extraction was accomplished using a scaled down version of Kleinhofs et al. (1993) CTAB extraction protocol. After re-suspension, samples were spun at 13,000 rpm for 15 minutes to pellet non-soluble contaminants and the supernatant was placed into a clean tube. The extracted DNA was quantified using a spectrophotometer at 260nm. After quantification, a 10 µg sample of DNA from each line was removed for further analysis. The remainder was precipitated and placed under ethanol at -20°C for long term storage.

4.3.3 Amplified Fragment Length Polymorphism

A modified version of Keygene's AFLP protocol was used to generate polymorphisms. A one µg sample of template DNA was digested using five units Mse I in 1X React 1 buffer (Gibco-BRL) in a total volume of five µl. Digestion was allowed to continue for two hours at 37°C. Five units of Pst I and H Buffer (Boeringher-Mannheim) were then added to make a total volume of 50 µl. This was also incubated for two hours at 37°C. Enzymes were deactivated by a 70°C treatment for 15 minutes.

Double stranded adapters were ligated onto the digested DNA using T4 DNA Ligase (Gibco-BRL) and ligation buffer at 20°C for two hours at room temperature. The Pst I adapter had a sequence of:

5' GACTGCGTAGGTGCA
3' GAGCATCTGACGCATCC (reverse)

The Mse I adapter had a sequence of:

5' GACGATGAGTCCTGAG
3' TACTCAGGACTCAT (reverse)

Fifty pmoles of Mse I adapter and five pmoles of Pst I adapter were added to one µg of digested DNA for the ligation reaction.

Preamplification of the DNA was accomplished by making a PCR cocktail of 37.5 ng of each preamp primer (reverse sequence of each primer plus A for Pst I primer and plus C for Mse I primer), 0.8 µm dNTP's, 1X Perkin Elmer PCR buffer, 0.6 µm MgCl and one unit of Taq polymerase (Perkin Elmer). Twenty four µl of PCR cocktail were added to 10 ng of template DNA

and placed in an MJ DNA Engine thermocycler programmed for 20 cycles of 94°C for 30 seconds followed by 56°C for one minute followed by 72°C for one minute.

The selective amplification PCR cocktail contained 30 ng each of Pst I and Mse I primers, 0.6 μm of MgCl, 0.8 μm of dNTP's, 1X Perkin Elmer PCR buffer, and one unit of Taq polymerase (Perkin Elmer). Eighteen μl of cocktail were added to two μl of preamp DNA (diluted 1:6) and run in an MJ DNA Engine thermocycler programmed as follows :

One cycle of: 94°C for 1 minute - denature
65°C for 1 minute - annealing
72°C for 90 seconds – extension

Nine cycles of the same denaturing and extension conditions but a 1°C reduction in annealing temperature per cycle.

Twenty three cycles of: 94°C for 30 seconds - denature
56°C for 30 seconds - annealing
72°C for 60 seconds - extension

All PCR reactions were carried out in 96 well micro-titer plates with thermal seals. Following amplification, 20 μl of stop solution (98% formamide, 10 mM EDTA, bromophenol blue, orange G, xylene cyanol) were added to the PCR samples. All samples were denatured for five minutes at 95°C prior to loading on sequencing gels.

The PCR products were run on six percent (w/v) acrylamide, 0.4 mm thick sequencing gels (6% 20:1 acrylamide:bis, 7.5 M urea, 1X TBE buffer) using a Biorad apparatus. Three or four μl of sample were loaded into each well. Gels were run until the last dye band was at least two thirds of the way down the gel.

Bands were resolved using a commercially available silver staining technique (Promega). Gels were scored on a light box by two independent researchers and the results compared. Mismatched scores were discarded. The presence or absence of a band (that was polymorphic between AC Marie and Cascade) was recorded for each line. Data was recorded as a dominant marker system. Due to the advanced generation of the population, bands that migrated the same distance as a parental band were scored as homozygous for the positive parent and the absence of a band was considered equivalent to the negative parent. Co-dominant markers were scored

as two separate markers in order to avoid the complication of heterozygous loci. Migration distances of polymorphic fragments and pGem markers were recorded for each gel.

4.3.4 Map Construction

A combination of the software programs JoinMap Version 2.0 (Stam and van Ooijen, 1995) and MapMaker Version 3.0 (Lander et al. 1987) was used in the construction of this map. Different modules or features of each program were used to perform specific tasks but marker order within linkage groups was tested using both programs as a check. The same LOD and recombination criteria were used for both programs.

Segregation ratios for the marker data were checked using JoinMap and markers which showed deviation from a 1:1 segregation ratio ($p<0.05$) were removed from the mapping data set. Groups were determined at a LOD of 5 and a maximum recombination level of 30 percent using both JoinMap and MapMaker. Specific marker orders within groups were determined at a LOD of 2. After groupings and specific marker orders were determined, linkage between groups was tested at a LOD of 1 and a maximum recombination of 50 percent.

4.4 RESULTS

A total of 42 gels were scored resulting in the identification of 591 polymorphic fragments across the 80 lines of the mapping population and the two parents. A mean of 14.1 polymorphisms per gel were identified. Table 4.1 shows the number of polymorphisms identified per specific primer combination. The most scoreable polymorphisms on a gel were 24, the least were 6.

A total of 117 markers were found (using a chi-squared test) to be skewed away from a 1:1 segregation ratio ($p<0.05$). These markers were discarded from the mapping database. A further 22 markers were unlinked after grouping at an LOD of 5 and a maximum recombination of 30 percent. The map (Figure 4.2) covers approximately 1077 cM in a total of 35 linkage groups. The average distance between markers (not including unlinked markers or markers showing skewed segregation) is 2.4 cM.

Table 4.1. AFLP specific amplification primer combinations and the number of polymorphisms recorded for the 80 line mapping population of AC Marie/Cascade.

Primer Combination	Number of Polymorphisms
P-ACT + M-CAC	10
P-AAC + M-CAC	18
P-ACC + M-CAC	12
P-ACG + M-CAC	20
P-ACA + M-CAC	20
P-AAG + M-CAC	18
P-ACG + M-CAA	18
P-ACA + M-CAA	12
P-AGG + M-CAA	20
P-AAC + M-CAA	7
P-ACG + M-CAG	14
P-ACA + M-CAG	14
P-AAC + M-CAG	15
P-AGG + M-CAG	18
P-ACT + M-CAG	17
P-ACT + M-CAT	16
P-ACG + M-CAT	7
P-AAC + M-CAT	10
P-ACA + M-CAT	13
P-ACT + M-CTG	9
P-ACG + M-CTG	12
P-AGC + M-CTT	16
P-ACC + M-CTT	15
P-AAT + M-CTT	13
P-ACA + M-CTA	16
P-AGT + M-CTA	20
P-ATG + M-CTA	16
P-ATT + M-CTA	22
P-AAA + M-CGT	23
P-AGG + M-CCC	11
P-AGT + M-CGT	6
P-AGC + M-CCC	8
P-ACA + M-CCG	12
P-AAA + M-CAG	24
P-AAC + M-CAG	6
P-ATA + M-CCT	15
P-ACG + M-CCG	13
P-AGG + M-CTT	11
P-ATA + M-CCG	12
P-ATA + M-CGA	15
P-ACT + M-CTT	9
P-ACT + M-CGT	8
Mean	14.1

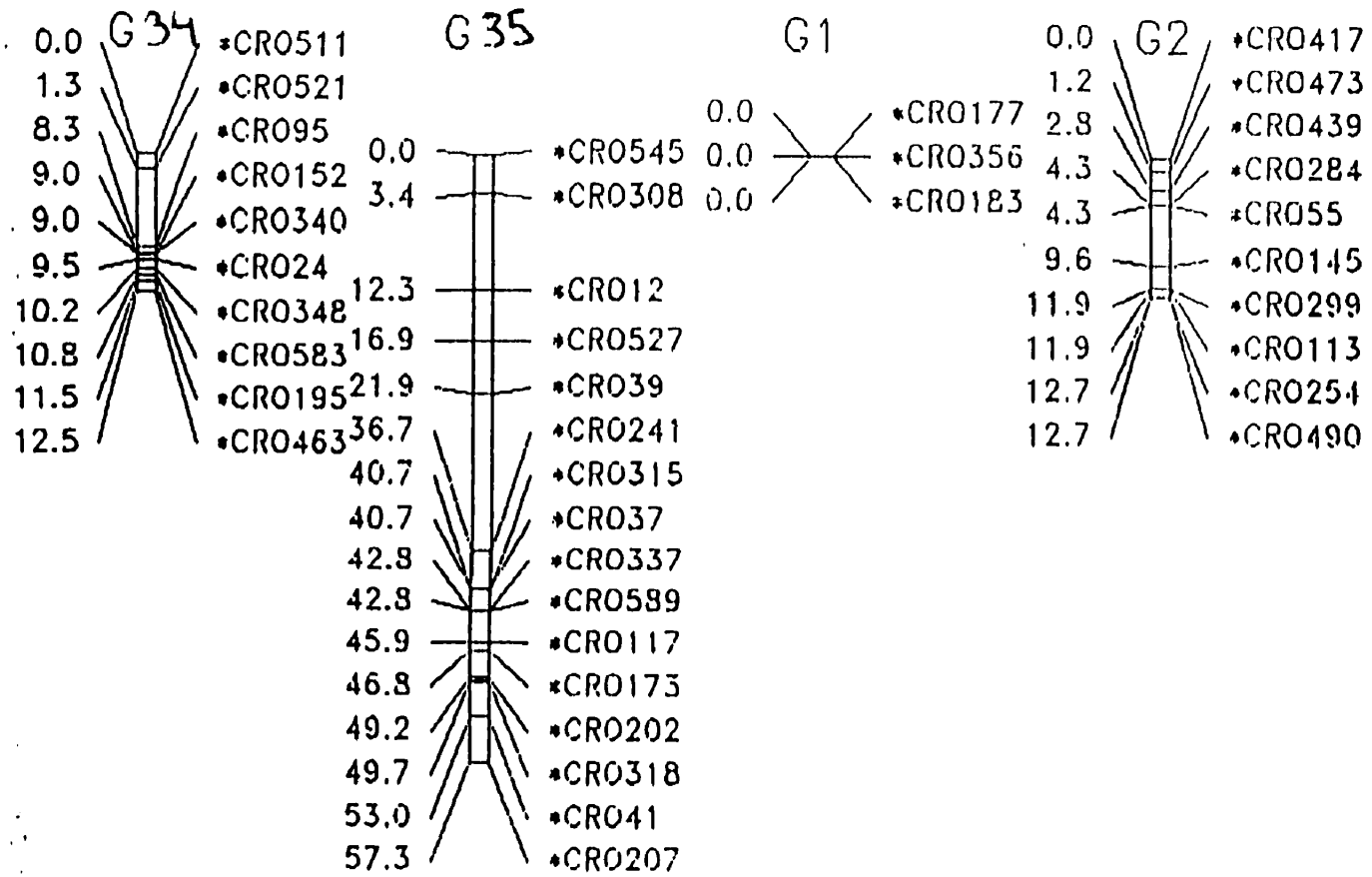


Figure 4.1. AFLP linkage map of the cross AC Marie/Cascade. Cumulative distances in cM are shown to the left of the linkage groups. Individual marker names are indicated to the right of the linkage groups.

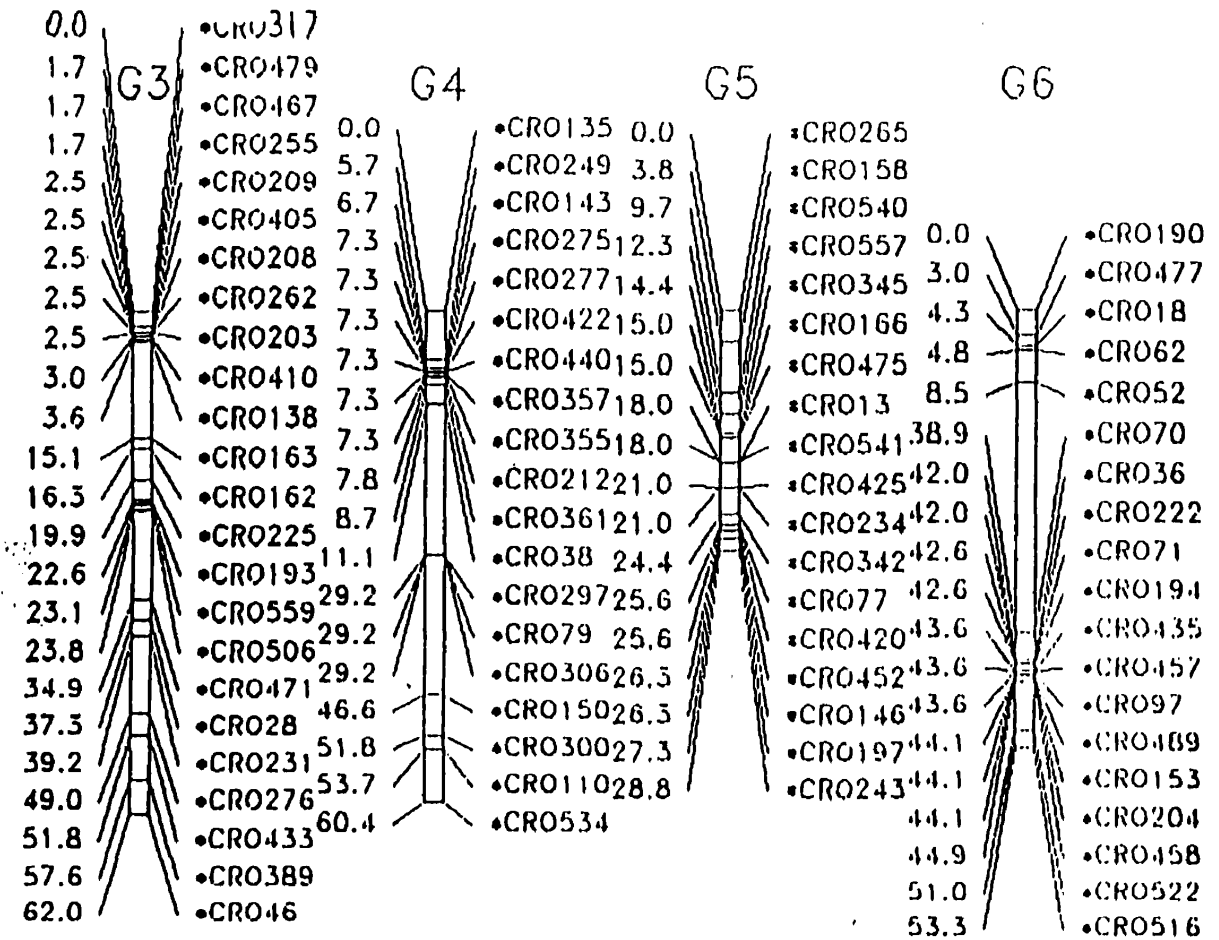


Figure 4.1. Continued

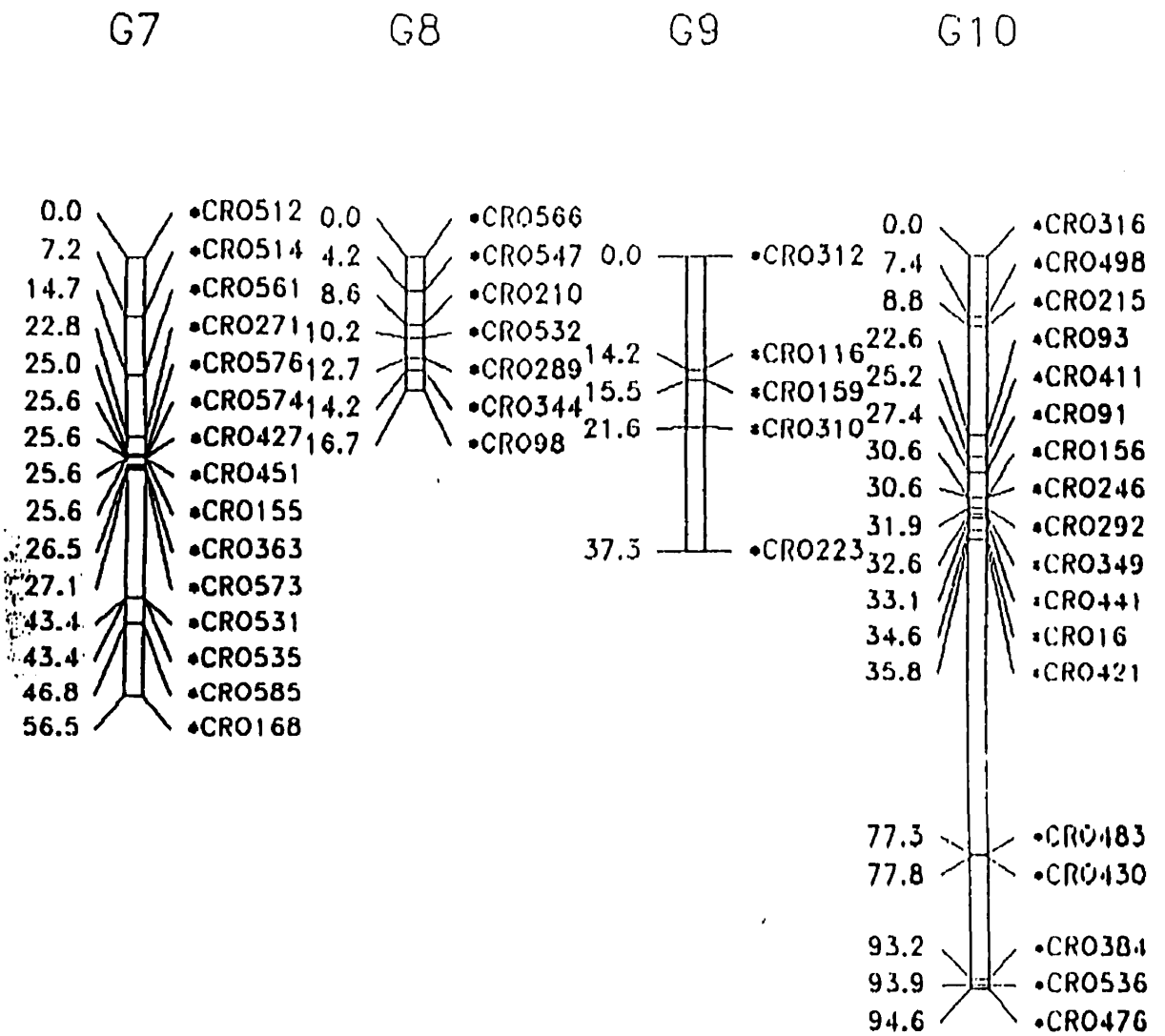


Figure 4.1. Continued

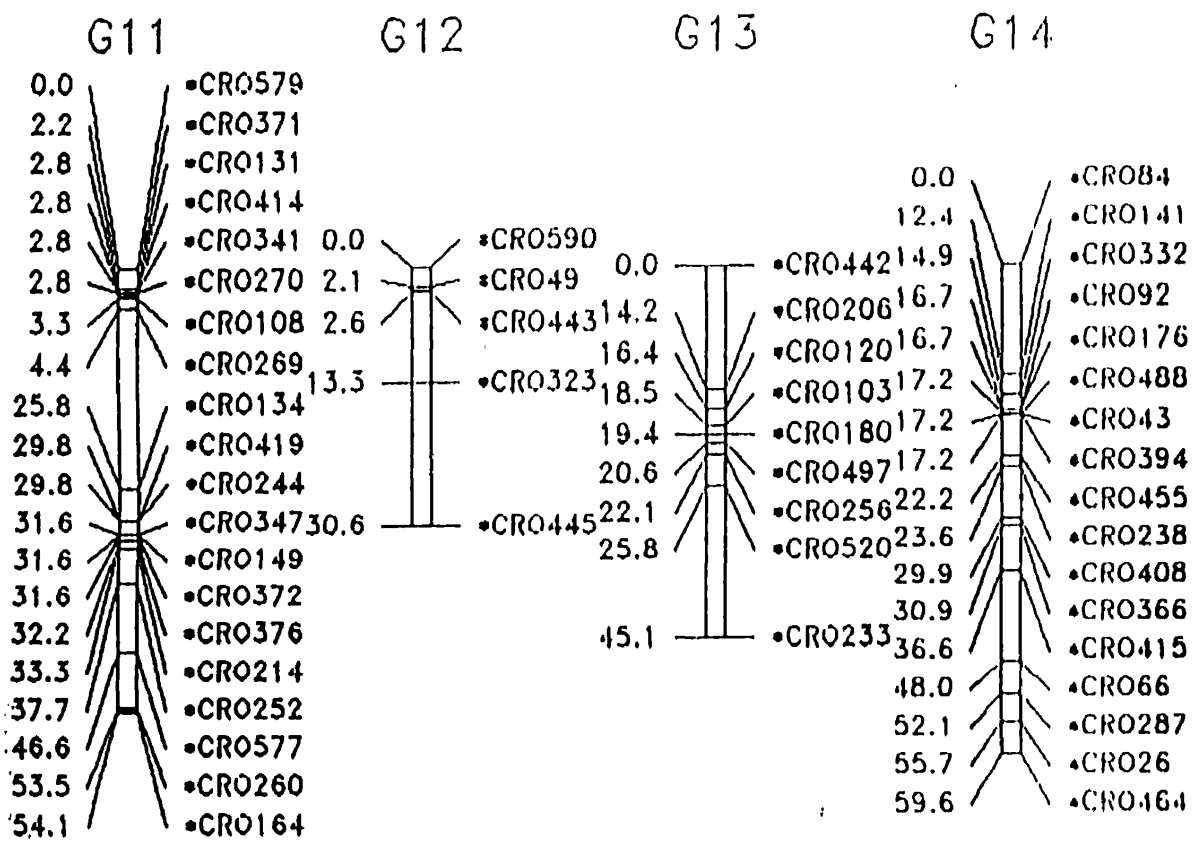


Figure 4.1. Continued

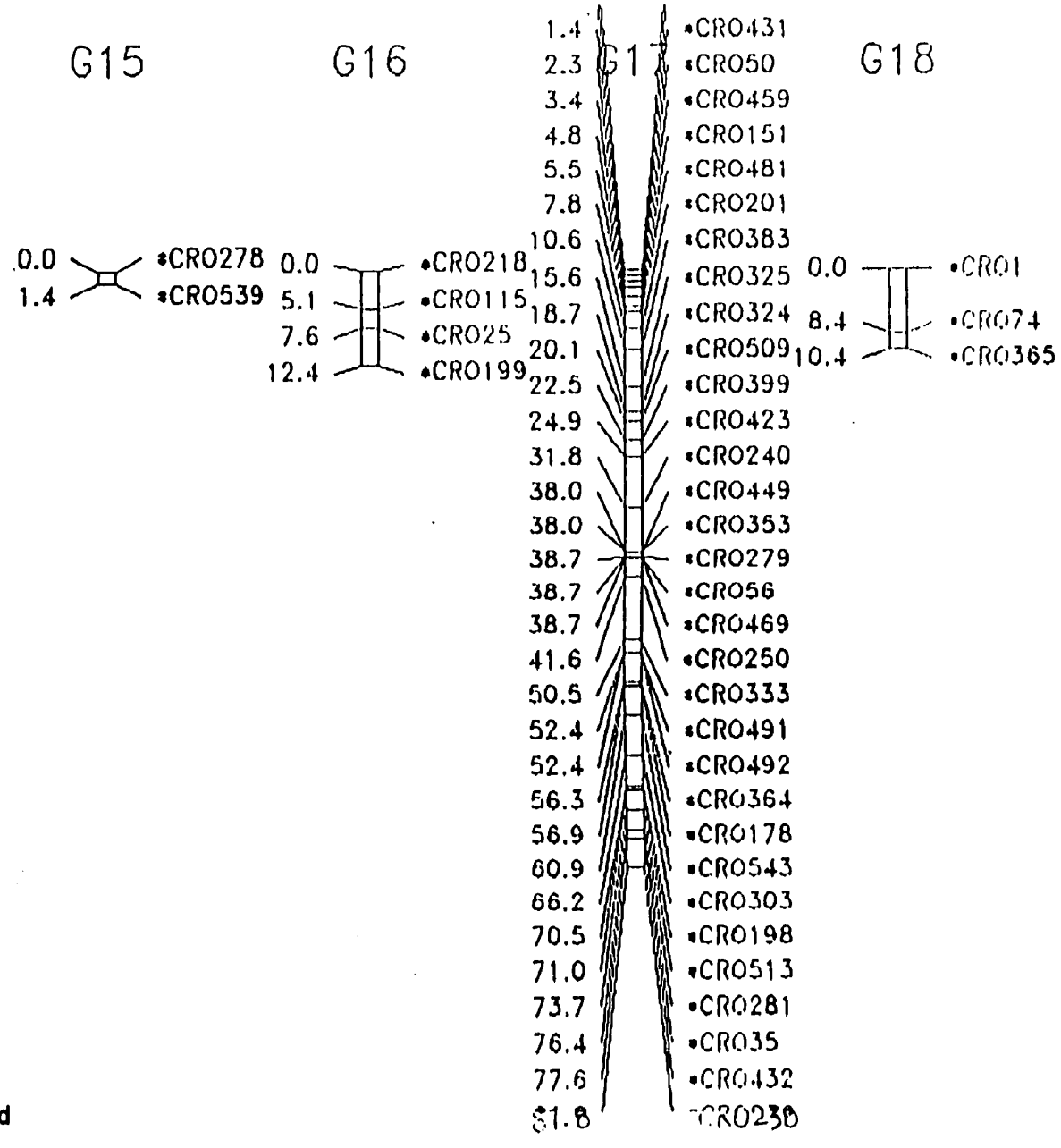


Figure 4.1. Continued

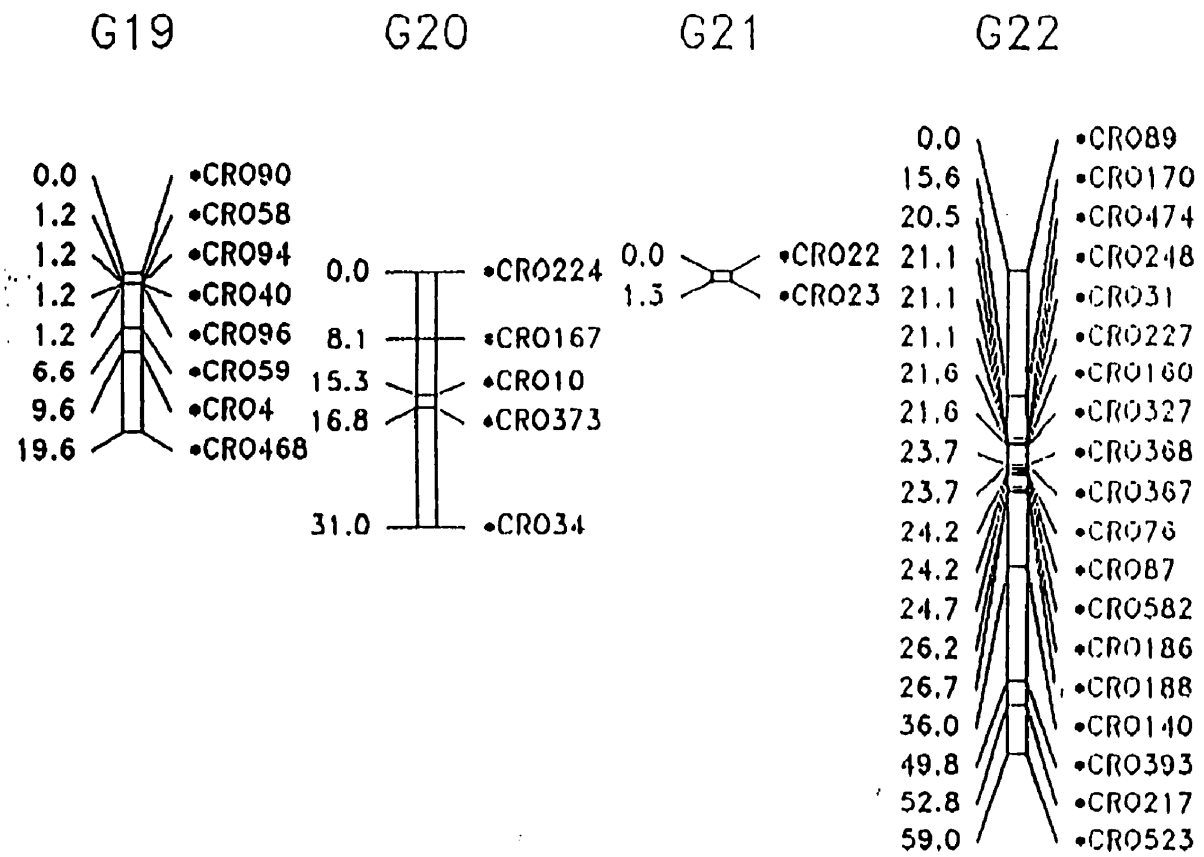


Figure 4.1. Continued

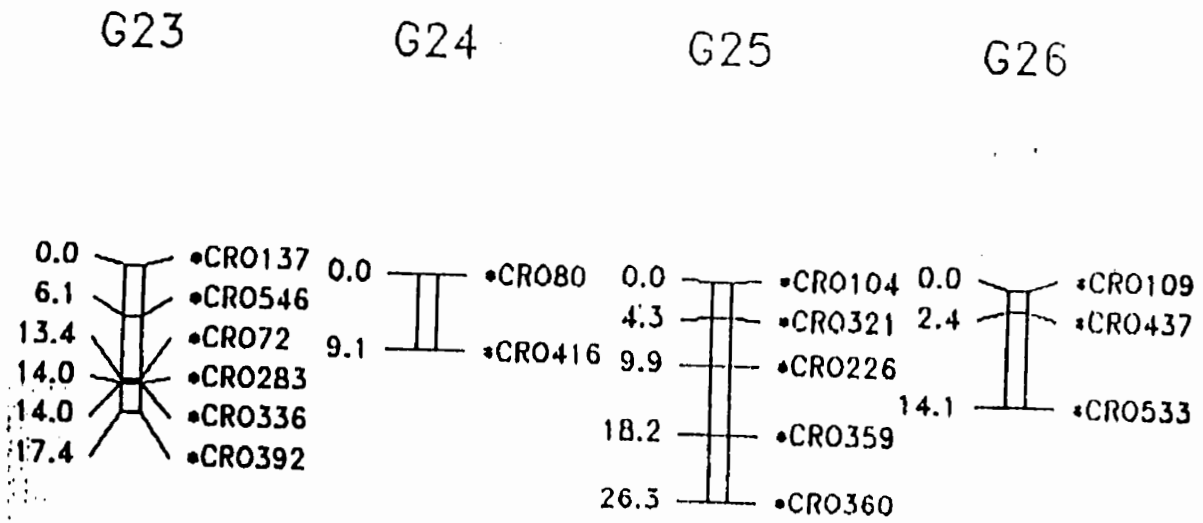


Figure 4.1. Continued

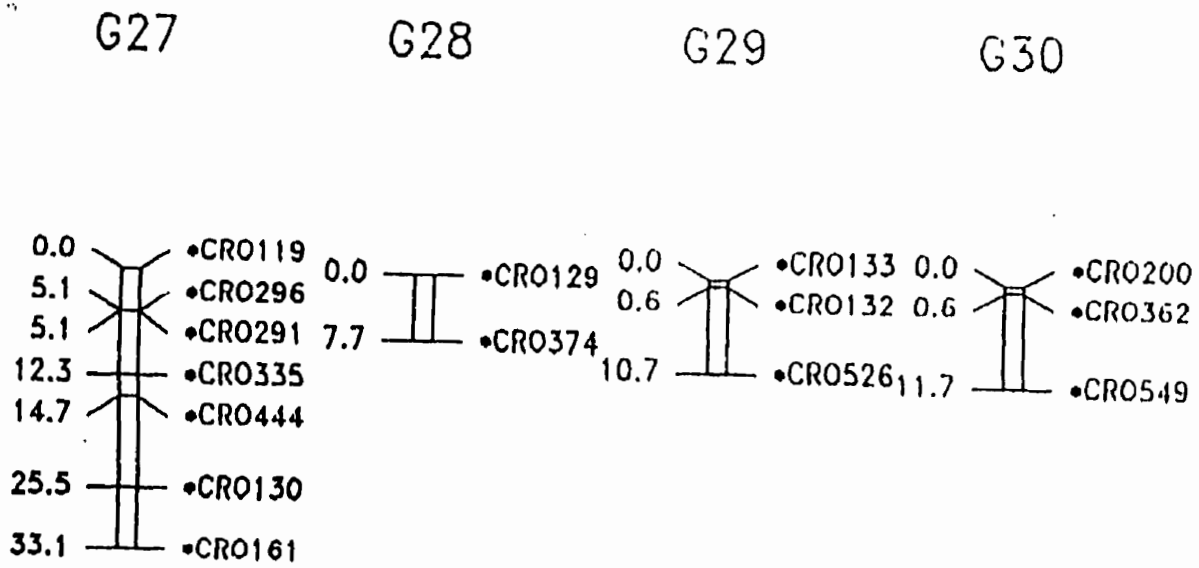


Figure 4.1. Continued

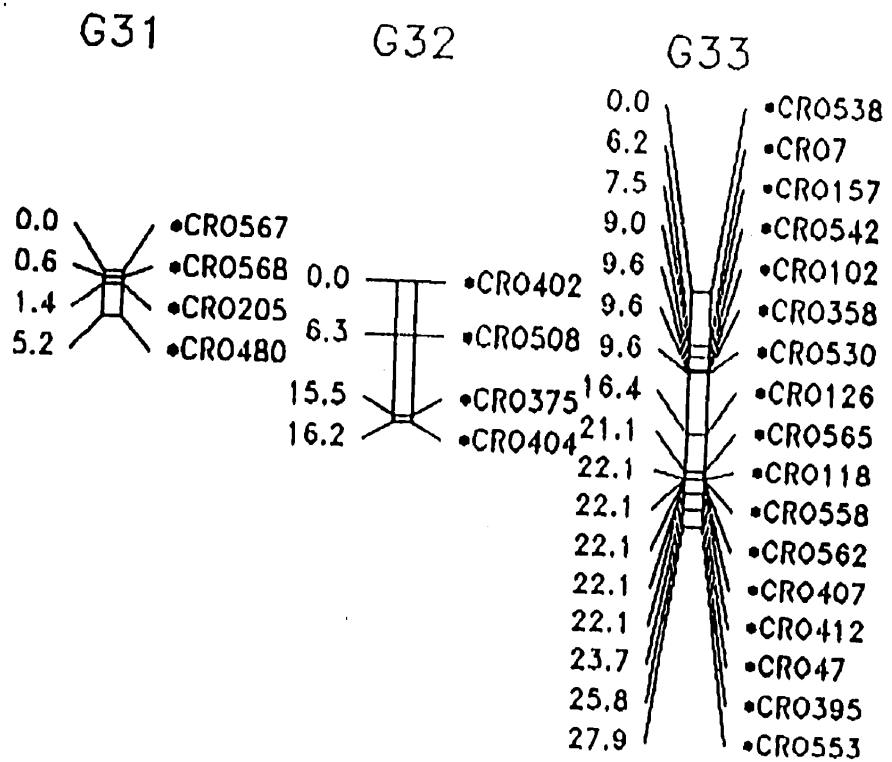


Figure 4.1. Concluded

4.5 DISCUSSION

4.5.1 Marker System

The AFLP marker system yielded a very high level of polymorphism per run. This made for relatively fast marker identification and allowed the development of a skeleton map in just over one year. Other marker systems like RFLP's and RAPD's and isozymes rarely yield more than one or two markers per gel/run (although multiplexing could increase this efficiency) (O'Donoughue et al., 1992; Rayapati et al., 1994 and O'Donoughue et al., 1995). The AFLP technique yielded more than 14 polymorphisms per PCR/gel which allowed rapid development of a marker database.

Independent double scoring was used to minimize the error in data reading. Only clear and well defined polymorphisms were used in the data set. One of the complicating factors is the effect of differing numbers of copies of a locus. Hexaploid species can contain up to three copies of homeologous loci. Differing band intensity due to one, two or three copies of a locus being amplified can blur the line between positive and negative scoring for an individual. Trying to estimate the number of copies of an allele base on band intensity would have been difficult because lines also differed in intensity due to the PCR process.

4.5.2 Segregation Ratios

The high number of markers that did not segregate according to the expected 1:1 ratio could have been influenced by the hexaploid nature of the *A. sativa* genome. Since any given marker could actually be amplified from a homologous locus on three separate linkage groups (independently assorting chromosomes), skewed segregation ratios could be the result of two or three independently assorting homologous loci. If two independently assorting homeologous loci were amplified by the same primer combination a 15:1 segregation ratio would result in the F_2 generation but as lines move toward homozygousity the ratio would become 3:1. For three independently assorting loci the ratio would be 63:1 in the F_2 generation but as lines moved toward homozygousity the ratio would become 7:1. Fifty one of the 117 markers removed from the data set due to improper segregation ratios could fit into either a 3:1 or a 7:1 ratio.

It is still important to remove these markers from the data set as they will show linkage between independently assorting linkage groups. This would lead to falsely mixing linkage groups and make marker ordering within a linkage group extremely difficult. Markers showing unexpected segregation ratios were included in the oat map published by O'Donoughue et al. (1995).

This leads to a potential problem with visually scoring the gels. This problem could be somewhat overcome by accounting for relative band intensity to indicate "dose". A dose effect for certain markers was noted in the study by O'Donoughue et al. (1995) but was not taken into account when examining segregation ratios. No dose effect was taken into account during data scoring in this study. Scoring a dose effect would be impossible to do with any accuracy without the use of expensive gel reading equipment that could take readings of band intensity. Even if accurate dose effect readings could be taken, it would still be difficult to assign loci to specific linkage groups.

4.5.3 Map

The map presented in this paper is by no means complete. The AC Marie/Cascade map currently stands at 35 linkage groups instead of the expected 21 and only covers 1077 cM of the expected 2900 cM (O'Donoughue et al., 1995). By considering the 22 unlinked markers and the 14 extra linkage groups, this study would estimate the genome to be just under 2900 cM. This is almost exactly the same figure estimated in other studies, although it should be noted that the cM distance of the map is dependant upon the parents and species used to create the map as well as the mapping technique.

There are a variety of factors that have an impact on map construction and add complexity to the mapping effort. The impact of changing the conditions or default values in mapping software cannot be understated. Small changes in LOD scores or maximum recombination values at just about any stage can have great impact on groupings and/or marker order. Added to this is the impact of including or excluding markers which fit poorly in the map or map onto the same loci.

A portion of the markers that mapped to the same locus were created when co-dominant markers were scored as two separate markers, although those that indicated heterozygosity would become separated by a distance determined by the number of heterozygous individuals identified. In hind sight it would have been more accurate to include the heterozygosity in the data set. The level of heterozygosity observed was surprising given the advanced generation of the material.

Other markers that mapped to the same location could have been the result of either homologous loci or random chance co-segregation. However, the probability of random chance co-segregation is extremely small.

Although markers that map onto apparently the same loci would seem to be redundant, excluding one or the other can impact marker order. This does raise the possibility that some of the apparently co-segregating bands are only exhibiting psuedo-linkage. Including pseudo-linked markers could artificially force separate linkage groups together and/or fracture linkage groups by pulling markers into the wrong groups. The full impact of including psuedo-linked markers within a linkage group is unknown, but would seem to be of great importance in mapping any polyploid species.

Another challenge in creating linkage maps is regions of monomorphic DNA in the parents. Although AC Marie and Cascade come from different breeding programs there is a strong possibility that there is some common germplasm in their respective backgrounds. Since some traits (such as yield, hull colour, height etc.) are commonly selected for in breeding programs, certain regions could become fixed for the same groups of alleles.

Since there are no polymorphic fragments within monomorphic regions, only the flanking markers can be used to determine if cross-overs have occurred within the monomorphic region. The effect of this is that it may be difficult to find markers within those monomorphic regions. Although the recombination distance of the region can be determined, marker saturation cannot be achieved. Additionally, large monomorphic regions may interfere in bringing linkage groups together. Long recombination distances between linked markers can reduce LOD scores making linked groups appear unlinked.

4.5.4 Future Directions

This map creates a good foundation for developing a complete map of *A. sativa*. The level of polymorphism between AC Marie and Cascade is encouraging when considering the effort involved in bringing the rest of the unlinked markers into the map. Assuming that each unlinked marker and extra linkage group represents an additional distance of 50 cM, the remainder of the map is about 1800 cM. This would bring the total genome to about 2900 cM in this cross. The addition of more markers is needed to increase genome coverage and bring together the unlinked markers and extra linkage groups.

Although the current map covers only about half of the genome in detail, it represents a powerful tool in identified markers linked to certain traits. The cross used is segregating for a wide range of traits including oil content, oil composition, hull content, height and disease resistance. Identification of markers linked to these traits would be beneficial in increasing the speed and accuracy of an oat breeding program. It must be acknowledged that markers identified from this study would need to be tested for cross applicability in other cultivars before widespread implementation.

Before a great deal of effort is put into identifying markers to bring linkage groups together and link the unlinked markers to the map, some investigation into chromosomal location of linkage groups and position on chromosomes may be useful. This type of identification will be very difficult until a well characterized set of aneuploids is available for oat.

5.0 IDENTIFICATION OF A QTL AFFECTING OIL CONTENT IN OAT (*Avena sativa* L.)

5.1 ABSTRACT

A single quantitative trait loci (QTL) affecting total oil content in oat (*Avena sativa* L.) was identified in a cross between the high oil parent AC Marie (5 percent oil content) and the low oil parent Cascade (3.5 percent oil content). Eighty F_9 and F_{10} lines grown in Glenlea (1996) and Saskatoon (1997) were evaluated for oil content using near infrared reflectance (NIR) spectroscopy. Means for the two years were not statistically different (P=0.05). A linkage map of the cross consisting of 591 amplified fragment length polymorphism (AFLP) markers was used. The map was organized into 35 linkage groups covering 1077 cM. The QTL, closely linked to the marker CRO217, was responsible for a 0.58 percent increase in oil content when comparing the class means for those lines carrying CRO217 with the class mean for lines not carrying CRO217. Threshold values for testing for the presence of a QTL were calculated using 1000 random permutations of the data set. The identified QTL had a test statistic of 82.6 compared to a threshold value of 9.6. The marker CRO217 may be useful in marker assisted selection during the development of improved oil cultivars.

5.2 INTRODUCTION

Raising the oil content of oat (*Avena sativa* L.) has been suggested as one possible way to increase the value of the crop (Frey and Hammond, 1975). Brown and Craddock (1972) demonstrated that oil contents ranged between three and eleven percent within the world oat collection although most cultivars only range between five and nine percent. An oat line containing over 16 percent oil was created by Schipper and Frey (1991) showing that high oil lines could be created. Although increasing the oil content of oat would increase the energy value, demands in today's market place for low fat foods and foods with low levels of unsaturated fats may change the emphasis of such work.

Oil content in oat is a complexly inherited trait (Brown et al., 1974; Frey et al., 1975; Thro and Frey, 1985) that can exhibit additive, dominant and heterotic gene interactions (Karow, 1980). Traits that are under multi-genic control (quantitative traits) can be difficult to manipulate in breeding programs (Falconer and Mackay, 1996). One of the reasons for this is that parental types may carry alleles that have the opposite effect. For example, a high oil content parent may also carry some genes for low oil while a low oil content parent may carry genes that contribute to high oil. By identifying loci that affect oil content and determining their actual genetic contribution to phenotype, a better understanding of all the factors affecting oil content will result. Identified QTLs should help improve the efficiency of selective breeding and expand knowledge of how complexly controlled traits are inherited.

Genetic analysis of the oat genome is complicated by its large size and hexaploid nature. Oat is an allopolyploid containing A, C, and D genomes with a basic chromosome number of n=3x=21 (Rajhathy and Thomas, 1974). The polyploid nature of oat complicates mapping due to the large number of linkage groups, detection of homologous loci with a single marker, and the co-migration of unlinked fragments (O'Donoughue et al., 1995).

O'Donoughue et al. (1995) published a map of cultivated oat based on the cross Kanota/Ogle. This map contained 561 marker loci, organized into 38 linkage groups that covered 1482 cM. It was estimated that this map covers over half of the total oat genome which was estimated to be 2932 cM. Siripoonwiwat et al. (1996) successfully used this map to identify QTLs

for agronomic traits such as grain yield, straw yield, plant height, test weight, groat percentage and days to maturity. QTLs affecting quality parameters in barley were identified by Hayes et al. (1997) using an AFLP map. Loci affecting diastatic power, grain protein, malt extract and α-amylase activity were mapped across eleven populations.

The purpose of this study was to determine the existence of one or more QTL affecting the inheritance of oil content in the cross AC Marie/Cascade using an existing AFLP linkage map of the genome.

5.3 MATERIALS AND METHODS

5.3.1 Plant Material

The population used in this study consisted of a subpopulation of an AC Marie/Cascade cross. The original population consisted of 223 segregating lines that were generated from F_2 bulks. First generation seed was allowed to self and then single seeds from the F_2 generation were used to advance the lines further. After the F_2 generation all harvested seeds were bulked. A subpopulation was created by selecting a single seed from the F_7 bulk of 80 of the original 223 lines. These 80 lines consisted of 20 lines showing high oil content, 20 lines showing low oil content, and 40 random lines. The plant materials were grown and harvested in Saskatoon, SK and Glenlea, MB during the summer of 1997. The plots were planted in 1.5 m rows with 9 inch spacing between rows.

5.3.2 Oil Content Measurement

Whole grain samples were analyzed for total oil content using near infrared reflectance spectroscopy (NIR). A volume of whole oat was placed in a NIR Systems model 6500 scanning monochromator. The samples were double scanned and a total of 64 scans were automatically averaged to determine a reading for each sample. A calibration curve created specifically for oat was used to determine total oil content. Data were collected and compiled in Quatro Pro V. 6.0 (Novell, 1995).

5.3.3 Genetic Markers and Genome Map

A map of the cross AC Marie/Cascade was created by Anderson et al. (1999) using the amplified fragment length polymorphism (AFLP) technique. A total of 591 AFLP markers were

identified and used to create the map. The map was organized into 35 linkage groups covering an estimated 1077 cM. The map data were used as one of the data sets for QTL analysis using MQTL (Tinker and Mather, 1995).

5.3.4 Detection of QTLs

Possible QTLs were identified using the software package MQTL. Marker loci and linkage distances from the AFLP map and phenotypic information for oil content from two environments were used as input for the software. Simple interval mapping (SIM) and composite interval mapping (CIM) were used to test for the presence of QTLs affecting oil content. Threshold values were determined using 1000 permutations of the marker and phenotype data.

5.4 RESULTS

The results of the MQTL SIM analysis indicated that only one locus was found to be affecting oil content in the AC Marie/Cascade population. This locus was determined to be at or very closely linked to the marker CRO217 which was located on linkage group 22 (Figure 5.1). The SIM test statistic for this locus was 82.6 versus the threshold statistic of only 9.6. Figure 5.2 shows how the SIM test statistic increases along the linkage group, peaks at the CRO217 locus and begins decreasing. The same QTL was identified using CIM. Three other loci were found to be approaching the threshold level but were not significant using either the SIM method or the CIM method.

AC Marie carries the positive allele of the CRO217 marker and the band is absent in Cascade. When the population is segregated based on presence or absence of the CRO217 marker, the mean oil content of the class carrying the positive allele increases to 4.31 percent from 3.97 percent. The class lacking the CRO217 band has a mean oil content of 3.73 percent. Means for these two classes were found to be significantly different to a probability level of <0.001, based on two way analysis of variance with replication. Figure 5.3 shows the frequency distribution of oil content for both genotypes.

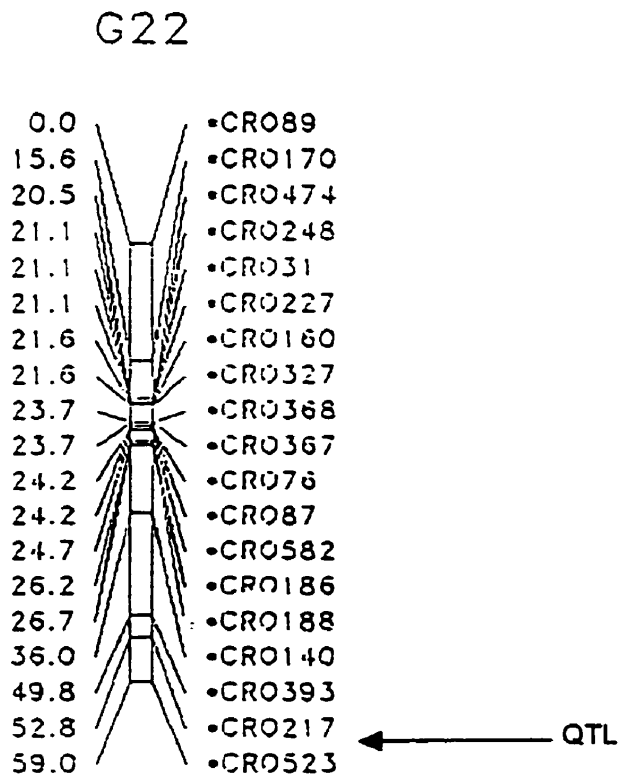


Figure 5.1. Linkage group 22 of the cross AC Marie/Cascade showing position of marker CRO217 linked to a QTL affecting oil content.

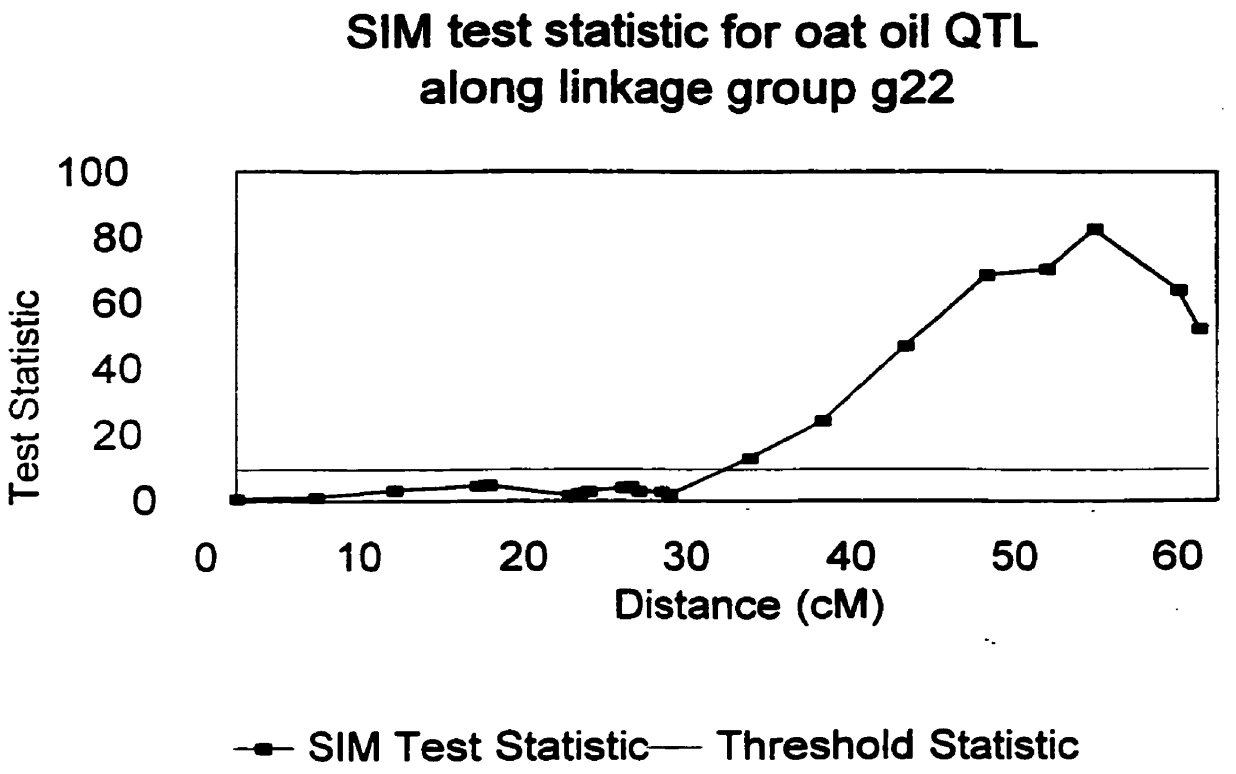


Figure 5.2. Graph of simple interval mapping (SIM) test statistic for oil content as it was calculated at various loci along the linkage group g22 of the cross AC Marie/Cascade. The threshold statistic, calculated based on 1000 random permutations of the data set, is shown as a dashed line.

5.5 DISCUSSION

The single QTL identified explained a large portion of the variation for oil content in this cross. The chromosomal region linked to the CRO217 marker is responsible for an average increase of about 0.6 percent in total oil content. While this represents a large portion of variation in a cross where the parents only differ by about 1.5 percent in oil content, it does not explain all of the variation observed. Although no other QTLs were discovered using either the SIM or sCIM techniques, additional loci may be affecting the inheritance of oil content, explaining more of the variation observed. Some of the unexplained variation may also be the result of environmental effects.

The MQTL software package uses a multiple regression approach to interval mapping. Simple interval mapping tests genotype classes for significant phenotypic differences at each marker locus and at each interval between markers. SIM testing does not take into account the presence of other QTLs in the data set and thus the presence of large QTLs can interfere with the detection of QTLs with lesser effects (Doerge, et al., 1994). Composite interval mapping is designed to take into account additional QTLs by performing regressions on both phenotype score versus the specific marker/interval being tested and phenotype score versus the rest of the marker data less the specific marker/interval. The regression of phenotype score onto the rest of the marker set effectively removes the effects of other QTLs in the genome allowing for better detection of a QTL linked to a specific marker (Jansen, 1993). Neither approach is able to distinguish QTLs that are closely linked to each other.

Since only a single large effect QTL was identified by either approach (SIM or CIM) any additional QTLs affecting oil content are expected to be either very small or linked to the CRO217 marker where their effect is masked by larger QTL. Additionally, the resolution of the phenotypic data may be insufficient to distinguish small effect QTLs. Oil content determination using NIR may carry too high an error to identify lesser QTLs.

Doerge and Churchill (1994) demonstrated the value of using multiple permutations of the data set to calculate threshold statistics. The presence of a large effect QTL(s) in the

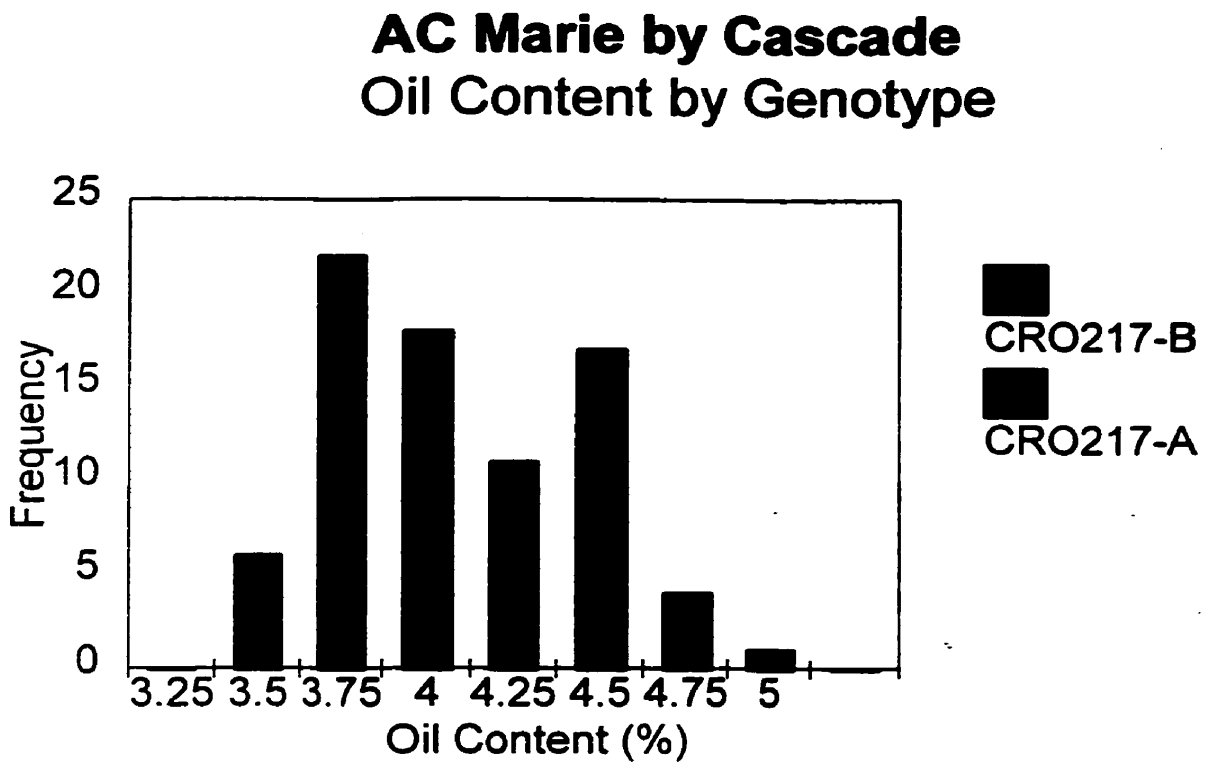


Figure 5.3. Frequency distribution showing oil content frequency and genotype of an 80 line population of the cross AC Marie/Cascade. The CRO217-B genotype corresponds to the Cascade genotype and the CRO217-A genotype corresponds to the AC Marie genotype. Oil content data shown are mean values for samples grown in Glenlea, MB in 1996 and Saskatoon, SK in 1997.

phenotypic data does increase the overall values for the threshold statistics. Perhaps the variation explained by identified QTLs should be removed from the phenotypic data set and thresholds recalculated when looking for additional QTLs.

The identification of this QTL is a positive step toward increasing understanding of oil inheritance in oat. In order to utilize this marker several additional developmental steps must take place. Since AFLP is a dominant marker system the Cascade allele is represented simply by the absence of a band. Oil content is a trait that breeders may wish to either increase or decrease, depending on the desired outcome. For a marker to be useful, through marker assisted selection, in a breeding program, both alleles should be identifiable. To accomplish this the CRO217 band must be sequenced, locus specific primers designed (these are primers which amplify a monomorphic band), then the monomorphic band sequenced, and finally, allele specific primers (for both the AC Marie and Cascade alleles) would need to be created.

With allele specific markers it would be possible to quickly assess early generation breeding populations for the presence of either high oil or low oil alleles. A DNA based test for oil content would eliminate the need to increase seed supplies to allow phenotypic testing and increase breeding efficiency by reducing the volume of material being advanced unnecessarily.

Before additional development work should take place this QTL should be assessed in other populations in order to determine its cross-applicability. It is also important to remember that the expected gain (or reduction) in oil content from the presence (or absence) of this allele is only 0.6 percent in this cross.

6.0 GENERAL DISCUSSION

This investigation looked into the inheritance of oil content and composition in oat and genetic control of total oil content using a map based approach. Total oil content and compositional data were measured on a segregating population of F_8 and F_9 progeny of the cross AC Marie/Cascade. This phenotypic data was used to determine inheritance of total oil content, relationships between the accumulation of different fatty acids, and, in conjunction with the mapping data, to detect the presence of a QTL affecting total oil content. Molecular marker data were generated for this population using AFLP. The marker data were organized into a linkage map, which led to the identification of a marker linked to a QTL for oil content.

The phenotypic data suggested that total oil content is likely controlled by more than one gene and that each fatty acid may also be under poly-genic control. This pattern could have been overstated as a result of individual lines being heterogeneous. Heterogeneous lines are essentially mixtures of individuals fixed for parental alleles at each locus. This resulted in a much higher frequency of mid-parent type values than would have been expected in such advanced material.

The heritability of oil content was found to be quite low in this cross with a broad sense heritability of only 0.19. The low heritability and lack of transgressive segregants in the total oil content study would suggest that substantial changes in total oil content beyond the parental values would be difficult with this population. This is supported by the genetic data, which revealed that only one locus with a major effect was controlling oil content inheritance in this population.

From the data generated by this study, selection for oil content should be highly successful, despite the low heritability. Gains (or reductions) in oil content will be dependent on genetic variation. NIR was a valuable technique for identifying relative levels of oil, although further work would need to be done to determine its accuracy at determining absolute oil content. The availability of NIR and the identification of markers for a locus with a strong effect on oil content may increase breeders' ability to make advances in their materials. No significant

environmental effect was found which suggests that breeders could make gains in selection for total oil content using a fairly small number of locations.

AFLP proved to be a powerful tool for the construction of linkage maps, allowing the rapid identification of almost 600 polymorphic markers. Despite being a dominant marker system and showing some clustering of loci, these markers were formed into a map covering almost 40 percent of the total genome. This compares very favourably with some other reported mapping efforts (O'Donoughue, et al., 1995; Kleinhofs, et al., 1993; Heun, et al., 1993).

The linkage map created in this study provides a framework for the identification of additional QTLs for any traits that are also segregating in the AC Marie/Cascade population. This map agrees reasonably well with that created by O'Donoughue et al. (1995). Further development is needed to increase the coverage of the entire genome. Data from this study would estimate total genome size at about 2900 cM. This assumes an additional 50 cM distance for each of the 22 unlinked markers and condense the 35 existing linkage groups down to the expected 21 linkage groups.

Development of the linkage map by adding additional markers and/or markers of known chromosomal location would enhance its usefulness. A more complete map would be valuable for a number of studies, including further identification of oil related QTLs, identification of QTLs for other traits, comparison mapping between other oat crosses, and refining understanding the inheritance of traits of interest.

The QTL linked to the marker CRO217 was found to be responsible for 0.6 percent increase in oil content in the AC Marie/Cascade population. The usefulness of this marker for marker assisted selection will depend upon its cross-applicability to other populations and the conversion of the CRO217 marker into allele specific markers. Such markers would be useful in selecting high or low oil progeny from early generation crosses. Early generation selections may allow for more rapid development of specific oil content cultivars in oat.

7.0 REFERENCES

Arumuganathan, K., and E.D. Earle. 1991. Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9:208-218.

Baker, R.J., and R.I.H. McKenzie. 1972. Heritability of oil content in oats, *Avena sativa* L. Crop Sci. 12:201-202.

Bassam, B.J., G. Caetano-Anolles, and P.M. Gresshoff. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. 196:80-83.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Gen. 32:314-331

Branson, C.V., and K.J. Frey. 1989. Recurrent selection for groat oil content in oat. Crop Sci. 29:1382-1387.

Brown, C.M., D.E. Alexander, and S.G. Carmer. 1966. Variation in oil content and its relation to other characters in oats (*Avena sativa* L.). Crop Sci. 6:190-192.

Brown, C.M., and J.C. Craddock. 1972. Oil content and groat weight of entries in the world oat collection. Crop Sci. 12: 514-515.

Brown, C.M., A.N. Aryeetey, and S.N. Dubey. 1974. Inheritance and combining ability for oil content in oats (*Avena sativa* L.). Crop Sci. 14:67-69.

Burnette, D., M. Lenz, P.F. Sisson, S. Sutherland, and S.H. Weaver. 1992. Marketing , processing and uses of oat for food. P. 247-263. *In* H.G. Marshall and M.E. Sorrells (eds.) Oats: Science and technology. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin.

Canada Grains Council. Canadian grains industry statistical handbook, 1996 edition. Canada Grains Council, Winnipeg, Manitoba. 1993, pp. 4-69.

De la Roche, I.A., D.E. Alexander, and E.J. Weber. 1971. Inheritance of oleic and linoleic acids in *Zea mays* L. Crop Sci. 11:856-859.

De la Roche, I.A., V.D. Burrows, and R.I.H. McKenzie. 1977. Variation in lipid composition among strains of oats. Crop Sci. 17:145-148.

Doerge, R.W., Z.B. Zeng, and B.S. Weir. 1994. Statistical issues in the search for genes affecting quantitative traits in populations. P. 15-26 *In* Analysis of molecular marker data. Joint Plant Breeding Symposia Series – August 1994, Corvallis, Oregon.

Doerge, R.W., and G.A. Churchill. 1994. Issues in genetic mapping of quantitative trait loci. P. 27-32 *In* Analysis of molecular marker data. Joint Plant Breeding Symposia Series – August 1994, Corvallis, Oregon.

Edwards, M.D., T. Helentjaris, S. Wright, and C.W. Stuber. 1992. Molecular-marker-facilitated investigations of quantitative trait loci in maize. Theoretical and Applied Genetics 83:765-774.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. Longman Group Ltd., Essex, England.

Forsberg, R.A. and D.L. Reeves. 1992. Breeding oat cultivars for improved grain quality. P. 751-775. *In* H.G. Marshall and M.E. Sorrells (eds.) Oats: Science and technology. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin.

Frey, K.J., and E.G. Hammond. 1975. Genetics, characteristics, and utilization of oil in caryopses of oat species. Journal of the American Oil Chemists' Society 52:358-362.

Frey, K.J., E.G. Hammond, and P.K. Lawrence. 1975. Inheritance of oil percentage in interspecific crosses of hexaploid oats. Crop Sci. 15:94-95.

Hayes, P.M., J. Cerono, H. Witsenhoer, M. Kuiper, M. Zabeau, K. Sato, A. Kleinhofs, D. Kudrna, A. Kilian, M. Saghai-Maroof, D. Hoffman, and the North American Barley Genome Mapping Project. 1997. Characterizing and exploiting genetic diversity and quantitative traits in barley (*Hordeum vulgare*) using AFLP markers. Journal of Quantitative Trait Loci 3:Article 2.

Heun, M., A.E. Kennedy, J.A. Anderson, N.L.V. Lapitan, M.E. Sorrells, and S.D. Tanksley. 1991. Construction of a restriction fragment length polymorhphism map for barley (*Hordeum vulgare*). Genome. 34:437-447.

Hongtrakul, V., G.M. Huestis, and S.J. Knapp. 1997. Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theoretical and Applied Genetics 95:400-407.

Jansen, R.C. 1993. Interval mapping of multiple quantitative trait loci. Genetics. 135:205-211.

Jaworski, J.G. 1987. Biosynthesis of monoenoic and polyenoic fatty acids. P. 159-174. *In* P.K. Stumpf (ed.) The biochemistry of plants. Volume 9 Lipids: Structure and function. Academic Press, Orlando, Florida.

Jensen, N.F. 1961. Genetics and inheritance in oats: Inheritance of morphological and other characters. P.125-206. *In* F.A. Coffman (ed.) Oats and oat improvement. The American Society of Agronomy, Madison, Wisconsin.

Kalbasi-Ashtari, A. and E.G. Hammond. 1977. Oat oil: Refining and stability. Journal of the American Oil Chemists' Society. 54: 305-307.

Karow, R.S. 1980. Oil composition in parental, F_1, and F_2 populations of two oat (*Avena sativa* L.) crosses. MSc Thesis. University of Wisconsin, Madison.

Karow, R.S. and R.A. Forsberg. 1984. Oil composition in parental, F_1, and F_2 populations of two oat crosses. Crop Sci. 24:629-632.

Keim, P., B.W. Diers, T.C. Olson, And R.C. Shoemaker. 1990. RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126:735-742.

Kibite, S. and M.J. Edney. 1992. Effects of genotypes x environment interaction on the expressions of grain quality traits in oat. P. 101-103. *In* A.H. Barr (ed) The changing role of oats in human and animal nutrition, Proceedings of the 4th International Oats Conference, Adelaide, Australia.

Kleinhofs, A., A. Kilian, M. Saghai-Maroof, R Biyashev, P. Hayes, F. Chen, N. Lapitan, A. Fenwick, T. Blake, V. Kanazin, E. Ananiev, L. Dahleen, D. Kudrna, J. Bollinger, S. Knapp, B. Liu, M. Sorrells, M. Heun, J. Franckowiak, D. Hoffman, R. Skadsen, and B. Steffenson. 1993. A molecular, isozyme, and morphological map fo the barley (*Hordeum vulgare*) genome. Theoretical and Applied Genetics 86:705-712.

Knowles, P.F. 1965. Variability in oleic and linoleic acid contents in safflower oil. Econ. Bot. 19:53-62.

Krishnan, P.G., W.J. Park, K.D. Kephart, D.L. Reeves, and G.L. Yarrow. 1994. Measurement of protein and oil content of oat cultivars using Near-Infrared Reflectance spectroscopy. Cereal Foods World. 39:105-108.

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

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

Lin, J., and J. Kuo. 1995. AFLP: a novel PCR based assay for plant and bacterial DNA fingerprinting. Focus 2:66-70.

Mahmud, K., and H.H. Kramer. 1951. Segregation for yield, height, and maturity following a soybean cross. Agronomy Journal 43:605-609.

MacKey, J. 1964. Mutation breeding in polyploid cereals. Acta Agri. 4:549-557.

McGinnis, R.C. 1966. Establishing a monsomic series in *Avena sativa* L. p. 87-96. *In* R. Riley and K.R. Lewis (ed.) Chromosome manipulations in plant genetics. Oliver and Boyd, London, England.

O'Donoughue, L.S., S.F. Kianian, P.J. Rayapati, G.A. Penner, M.E. Sorrells, S.D. Tanksley, R.L. Phillips, H.W. Rines, M. Lee, G. Fedak, S.J. Molnar, D. Hoffman, C.A. Salas, B. Wu, E. Autrique, and A. Van Deynze. 1995. A molecular linkage map of cultivated oat. Genome 38:368-380.

O'Donoughue, L.S., Z. Wang, M. Roder, B. Kneen, M. Leggett, M.E. Sorrells, and S.D. Tanksley. 1992. An RFLP-based linkage map of oats based on a cross between tow diploid taxa (*Avena atlantica* x *Avena hirtula*). Genome 35:765-771.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Paterson, A.H., S.D. Tanksley, and M.E. Sorrells. 1991. DNA markers in plant improvement. Advances in Agronomy 46:39-91.

Penner, G.A. 1996. RAPD analysis of plant genomes. P. 251-268. *In* P.P. Jauhar (ed.) Methods of genome analysis in plants. CRC Press, Boca Raton, Florida.

Penner, G.A., J. Chong, M. Levesque, S.J. Molnar and G. Fedak. 1993. Identification of a RAPD marker linked to the oat stem rust gene *Pg3*. Theoretical and Applied Genetics 85:706-712.

Penner, G.A., J. Chong, C.P. Wight, S.J. Molnar and G. Fedak. 1993. Identification of a RAPD marker for the crown rust resistance gene *Pc68* in oats. Genome 36:818-820.

Peterson, D.M. 1992. Composition and nutritional characteristics of oat grain and products. P. 265-292. *In* H.G. Marshall and M.E. Sorrells (eds.) Oats: Science and technology. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin.

Price, P.B., and J.G. Parsons. 1975. Lipids of seven cereal grains. Journal of the American Oil Chemists' Society 52:490-493.

Rajhathy, T., and H. Thomas. 1974. Cytogenetics of oats (*Avena* L.). Genet. Soc. Canada Misc. Publication No. 2. Genetics Society of Canada, Ottawa, Ontario.

Rayapati, P.J., J.W. Gregory, M. Lee, and R.P. Wise. 1994. A linkage map of diploid *Avena* based on RFLP loci and a locus conferring resistance to nine isolates of *Puccinia coronata* var. *avenae*. Theoretical and Applied Genetics 89:831-837.

Riley, R., and V. Chapman. 1958. Genetic control of the cytologically diploid behaviour in hexaploid wheat. Nature. 182:713-715.

Ronald, P. 1997. Genetic analysis of percent hull in oat (*Avena sativa* L.). MSc thesis. University of Manitoba, Winnipeg.

Sahasrabudhe, M.R. 1979. Lipid composition of oats (*Avena sativa* L.). Journal of the American Oil Chemists' Society 56:80-84.

Sears, E.R., and M. Okomato. 1958. Intergenomic relationships in hexaploid wheat. Proceedings of the 10th International Congress on Genetics. 2:258-259.

Schipper, H. and K.J. Frey. 1991. Observed gains from three recurrent selection regimes for increased groat oil content of oat. Crop Sci. 31:1505-1510.

Schipper, H. and K.J. Frey. 1992. Agronomic and bioenergetic consequences of selection for high groat oil content and high protein yield in oat. Plant Breeding 108:241-249.

Siripoonwiwat, W., L.S. O'Donoughue, D. Wesenberg, D. Hoffman, J.F. Barbosa-Neto, and M.E. Sorrells. 1996. Chromosomal regions associated with quantitative traits in oat. Journal of Quantitative Trait Loci, Volume 2,

Stam, P., and J.W. van Ooijen. 1995. JoinMap version 2.0: Software for the calculation of genetic linkage maps. CPRO-DLO, Wageningen.

Stumpf, P.K. 1987. The biosynthesis of saturated fatty acids. P. 121-136. *In* P.K. Stumpf (ed.) The biochemistry of plants. Volume 9 Lipids: Structure and function. Academic Press, Orlando, Florida.

Stymne, S. and A.K. Stobart. 1987. Triacylglycerol biosynthesis. P. 175-215. *In* P.K. Stumpf (ed.) The biochemistry of plants. Volume 9 Lipids: Structure and function. Academic Press, Orlando, Florida.

Tanksley, S.D. and T.J. Orton. 1983. Isozymes in plant genetics and breeding. Eslevier, Amsterdam.

Tanksley, S.D., N.D. Young, A.H. Paterson, and M.W. Bonierbale. 1989. RFLP mapping in plant breeding: New tools for an old science. Biotechnology 7: 257-264.

Thomas, H. 1992. Cyotgenetics of *Avena*. P. 473-508. *In* H.G. Marshall and M.E. Sorrells (eds.) Oats: Science and technology. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin.

Thro, A.M. and K.J. Frey. 1985. Inheritance of groat-oil content and high oil selection in oats (*Avena sativa* L.). Euphytica 34:251-263.

Thro, A.M., K.J. Frey, and E.G. Hammond. 1985. Inheritance of palmitic, oleic, linoleic and linolenic fatty acids in groat oil of oats. Crop Sci. 25:40-44.

Tinker, N.A., and D.E. Mather. 1995. MQTL: Software for simplified composite interval mapping of QTL in multiple environments. Journal of Quantitative Trait Loci. Volume 1.

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

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

Welsh, J., and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18:7213-7218.

Wight, C.P., G.A. Penner, L.S. O'Donoughue, V.D. Burrows, S.J. Molnar, and G. Fedak. 1994. The identification of random amplified polymorphic DNA markers for daylength insensitivity in oat. Genome 37:910-914.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535.

Youngs, V.L. and M. Puskulcu. 1976. Variation in fatty acid composition of oat groats from different cultivars. Crop Sci. 16:881-883.

Youngs, V.L., M. Puskulcu, and R.R. Smith. 1977. Oat lipids. I. Composition and distribution of lipid components in two oat cultivars. Cereal Chem. 54:803-812.

Youngs, V.L. 1986. Oat lipids and lipid related enzymes. P. 205-226. *In* F.H. Webster (ed.) Oats: Chemistry and technology. Am. Assoc. Cereal. Chem., St. Paul, Minnesota.