# FINE MAPPING OF BIOMASS YIELD QUANTITATIVE TRAIT LOCI IN *LOLIUM PERENNE* L.

Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester

> by Céline TOMASZEWSKI Department of Biology University of Leicester

> > September 2011

## Declaration

I hereby declare that none of the work presented in this thesis has been submitted for another degree in this or any other University. All the work recorded in this thesis is original unless otherwise acknowledged in the text or by references.

The thesis is based on work conducted in the Department of Biology at the University of Leicester mainly during the period between April 2008 and September 2011

Date:

Céline TOMASZEWSKI

## ABSTRACT

#### Fine mapping of biomass yield quantitative trait loci in *Lolium perenne* L.

#### Céline Tomaszewski

Biomass yield is a complex quantitative trait controlled by many environmental and genetic factors. Therefore its study relies on QTL mapping. In a precursor study, a genetic map of *L. perenne* was constructed on an inbred-derived F2 population and three major biomass QTL have been found on linkage groups (LGs) 2, 3 and 7.

In this study, a fine map of the QTL positions was developed by mapping additional ryegrass specific SSR, rice Sequence Tagged Site and Diversity Array Technology markers. A total of 153 markers were added to the existing map leading to a map density of 3.5 cM. The QTL positions were recalculated for dry weight, fresh weight, dry matter and leaf width and in accordance to the preliminary analysis biomass QTL were localized on LGs 2, 3 and 7 but despite the fine map the QTL intervals were not reduced.

In order to analyze the QTL regions, the screening of a *L. perenne* BAC library was performed using the markers flanking the QTL and several clones were isolated. After analysis using the AFLP fingerprinting method, five clones were send for full sequencing to perform a gene prediction and annotation using the *Ab initio* approach. The annotation revealed for one of the gene structures predicted homology to the *lg1*-like gene and four other showed homology to regions flanking genes of interest suggesting the possible presence of the genes within the biomass QTL region. The four genes were: *L. perenne* heading date (*Hd1*) gene, *Avena strigosa* beta-amyrin synthase (*Sad1*) and cytochrome P450 CYP51H10 (*Sad2*) genes and *Lolium multiflorum* gene for cold responsive protein.

Dedicated to my parents for their patience, support and love Thank you for always being there for me.

### **ACKNOWLEDGEMENTS**

#### I would like to show my gratitude to:

My supervisors Dr Susanne Barth and Professor Pat Heslop-Harrison for their support and guidance throughout this project.

Dr Kerstin Diekmann for the helpful and constructive discussions throughout this study particularly during the writing time.

Dr David Kopecky for providing the DArT markers sequences and Dr Stephen Byrne for his help with the DArT markers analysis.

The Teagasc Oak Park staff and the University of Leicester staff who helped me on many occasions.

The friends I met in Ireland, especially Ulrike Anhalt, Bicheng Yang, Marialaura Destefanis, Jeanne Mehenni-Ciz and Mariateresa De Cesare, for their help, friendship and support that made the PhD time easier.

Teagasc for the Walsh Fellowship and the National Development Plan for financial contributions.

# **ABBREVIATIONS**

AFLP	amplified fragment length polymorphism		
BAC	bacterial artificial chromosome		
BC	backcross		
bp	base pair		
BSA	bovin serum albumin		
cDNA	complementary deoxyribonucleic acid		
CIM	composite interval mapping		
cM	centiMorgan		
СТАВ	cetyltrimethylammoniumbromide		
DArT	diversity array technology		
DM	dry matter		
DNA	deoxyribonucleic acid		
dNTP	nucleotides		
DW	dry weight		
EDTA	ethylenediaminetetraacetic acid		
EST	expressed sequence tag		
FW	fresh weight		
g	gram		
IM	interval mapping		
LB	lysogeny broth		
L. multiflorum	Lolium multiflorum		
L. perenne	Lolium perenne		
LG	linkage group		
LOD	logarithm of odds		
LW	leaf width		
М	molar		
MAS	marker-assisted-selection		
Mbp	mega base pairs		
mL	milliliter		
ML	maximum likelihood		
mM	millimolar		

MQM	multiple QTL model
NaCl	sodium chloride
NIL	near isogenic line
PCR	polymerase chain reaction
QTL	quantitative trait loci
RFLP	restriction fragment length polymorphism
RIL	recombinant inbred line
rpm	rotations per minute
RNase	ribonuclease
S	second
SD	segregation distortion
SDS	sodium dodecyl sulfate
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
STS	sequence tagged site
Temp	temperature
t/ha	ton per hectare
U	unit
v/v	volume added to volume
μg	microgram
μl	microliter
%	percentage
°C	degree Celsius
$\chi^2$	chi-square

# TABLE OF CONTENTS

List of tables1
List of figures
Chapter 1: General introduction to "Fine mapping of biomass yield quantitative trait loci in <i>Lolium perenne</i> L."
1.1 Breeding in perennial ryegrass7
1.1.1 Lolium perenne L7
1.1.2 The trait biomass yield7
1.1.3 Breeding in <i>Lolium perenne</i>
1.2 Positional cloning analysis to dissect complex traits10
1.2.1 Population structure for mapping11
1.2.2 Linkage map
1.2.3 QTL analysis and mapping13
1.2.4 Positional cloning14
1.3 Context and objectives15
Chapter 2: Materials and methods17
2.1 Fine mapping17
2.1.1 Plant material
2.1.2 DNA extraction
2.1.3 SSR markers analysis
2.1.4 STS markers design and analysis
2.1.5 Linkage map construction
2.2 DArT analysis22
2.3 QTL analysis23
2.3.1 Phenotypic data
2.3.2 QTL calculation
2.4 BAC library
2.4.1 Characteristics of the BAC library
2.4.2 Replication of the library25
2.4.3 Pooling of the BAC DNA

	2.4.4 Screening of the BAC library	28
	2.4.5. AFLP fingerprinting	30
	2.4.6 BAC clones sequencing and gene prediction	31
Ch	apter 3: Mapping of ryegrass specific SSR markers and rice STS markers	.33
3	3.1 Introduction	33
	3.1.1 Fine mapping	. 33
	3.1.2 Synteny	. 35
	3.1.3 Context and objectives of this chapter	37
3	3.2 Results	38
	3.2.1 Evaluation of the SSR and STS markers	38
	3.2.2 Construction of the linkage map	43
	3.2.3 Synteny	46
	3.3 Discussion	. 48
~		
Ch	apter 4: Fine mapping with DArT markers	57
4	4.1 Introduction	57
	4.1.1 Principle of Diversity Array Technology (DArT) and its application	. 57
	4.1.2 Mapping of DArT	58
	4.1.3 Objectives	60
4	4.2 Results	60
	4.2.1 Construction of the linkage map	. 60
	4.2.2 Characteristics of the map	. 64
4	4.3 Discussion	. 67
Ch	anter 5. Fine manning of biomass OTL	72
CII 4	5.1 Introduction	72
	5.1.1 Methods for OTL detection	72
	5.1.2 Factors to consider for OTL mapping	<i>7 2</i> 74
	5.1.3 Context and Objectives	75
4	5.2 Results	78
•	5.2.1 OTL positions after addition of SSR and STS markers	78
	5.2.2 OTL positions after addition of the DArT markers	

5.3 Discussion	92
Chapter 6: Identification and description of <i>Lolium perenne</i> genomic seque	ences
underlying biomass QTL	100
6.1 Introduction	100
6.1.1 BAC libraries as tool for identifying genomic sequences	100
6.1.2 Amplified Fragment Length Polymorphism (AFLP) fingerprinting	102
6.1.3 Context and Objectives	103
6.2 Results	104
6.2.1 Screening of a <i>L. perenne</i> BAC library	
6.2.2 AFLP fingerprinting	107
6.2.3 Gene prediction	
6.2.4 Mapping of the four genes	
6.2.5 Heading date QTL	119
6.3 Discussion	120
Chapter 7: General discussion	128
7.1 Fine mapping	
7.2 QTLanalysis	129
7.3 Description of the genomic sequences underlying the biomass QTL	
7.4 Further work	
Annexes	138
References	166

# LIST OF TABLES

<b>Table 1</b> : Polymerase chain reaction programs for simple sequence repeat markers from different sources used for the construction of the linkage map.       20
<b>Table 2</b> : Polymerase chain reaction program for the rice STS markers
<b>Table 3</b> : Polymerase chain reaction program for markers used for the screening of the BAC library
<b>Table 4</b> : AFLP amplification program
<b>Table 5</b> : List of the clones sent for sequencing
<b>Table 6</b> : Summary of the number of <i>Lolium perenne</i> specific SSR primers tested for thedifferent sources with the amplification rate and the polymorphism degree obtainedafter testing on the parental lines, the F1 and a subset of F2 genotypes
<b>Table 7</b> : Summary of the number of rice STS primers designed for each linkage groupwith the amplification rate and the polymorphism degree obtained after testing on theparental lines, the F1 and a subset of F2 genotypes
<b>Table 8</b> : Alignment between L. perenne and rice. L. perenne (Lp) is represented by thesequences from non-polymorphic amplicons from rice STS designed from ricechromosome 4. Rice is represented by genes showing orthology with the sequences 46
<b>Table 9</b> : Alignment between L. perenne and rice. L. perenne is represented by thesequences from non-polymorphic amplicons from rice STS designed from ricechromosome 1. Rice is represented by genes showing orthology with the sequences47
<b>Table 10</b> : Review of Quantitative trait loci analyses reported in <i>Lolium perenne</i>
<b>Table 11</b> : List of the markers identified as cofactors for MQM mapping for the map         with the SSR and STS markers
<b>Table 12</b> : QTL positions identified by IM, MQM and CIM, for the traits fresh weight,dry weight, dry matter and leaf width in the field and in the greenhouse experiment. Foreach test the QTL are displayed by LOD scores and percentage of explained variance.83
<b>Table 13</b> : QTL positions after addition of the DArT markers. QTL were identified byIM, MQM and CIM, for the traits fresh weight, dry weight, dry matter and leaf width inthe field and in the greenhouse experiment. For each test the QTL are displayed byLOD scores and percentage of explained variance
Table 14: Primers developed from the DArT sequences
Table 15: List of the clones isolated for each marker used to screen the BAC library.         107

<b>Table 16</b> : Characteristics of the five BAC clone sequence assemblies.       110
<b>Table 17</b> : Features of the five genomic sequences studied
<b>Table 18</b> : Annotation of the predicted genes using BLASTn and BLASTp for the five sequenced BAC clones. The five interesting genes are highlighted in blue
<b>Table 19</b> : Primers designed from gene sequences that were mapped onto the <i>L. perenne</i> linkage map.         118
Table 20. OTL registers identified by IM and MOM manning for the trait heading date

**Table 20**: QTL positions identified by IM and MQM mapping for the trait heading date

 in the field experiment. The QTL are displayed by LOD score and percentage of

 explained variance
 120

## LIST OF FIGURES

**Figure 2**: Pooling strategy combining a "plate pool" screening (dimension 1 and 2) followed by the screening of rows and columns assembled in pools (dimensions 3 to 6).

**Figure 5**: Chromatogram (top) of the sequences obtained with the marker LOC\_Os01g16152 for the maternal line (M) and the paternal line (Fa). The black arrow showed the SNPs. Amplicons were digested with the enzyme Hinf I and the reactions were separated by electrophoresis on 3% agarose gel (bottom). The primer was tested on the mother (M), the father (F), the F1 generation (F1) and five genotypes from the F2 generation (F2) to confirm the polymorphism before scoring the entire population...... 42

**Figure 8**: Mapping of DArT markers onto the genetic map developed by Anhalt et al. (2008) to generate a high resolution genetic map of the entire genome of *L. perenne*...62

# Chapter 1: General introduction to "Fine mapping of biomass yield quantitative trait loci

### in Lolium perenne L."

Plant improvement started around 10 000 years ago when humans domesticated some species and selected the most interesting cultivars. Modern breeding objectives are to create varieties with traits and characteristics that are of interest for the farmers and for the consumers. While breeding was first mainly based on an evaluation of the phenotypic values of individuals, the tendency is more and more to combine conventional breeding schemes with molecular genetic. The use of new biotechnology methods, such as marker assisted selection (MAS) can help to implement more efficient breeding strategies and permit rapid progress in the creation of superior varieties. Marker assisted selection is based on the establishment of a relation between a gene and an identifiable genetic marker. During the breeding program, it will be easier to select this marker than the trait itself (Hayward et al. 1994). As for many crops, Lolium perenne was selected for improvement. However it is rather recent. The first variety developed from a breeding program dated from around 1920 (Wilkins 1991). Improving yield remains the major objective of *Lolium* breeding programs. In 50 years the gain in dry matter yield has been of 4-6% per decade in some areas (Wilkins and Humphreys 2003). The average annual dry matter yield in Europe is around 17t/ha but one expects to develop more productive varieties to reach a yield of 25t/ha (Humphreys et al. 2010). Nowadays current forage breeding programs are using conventional breeding schemes to increase biomass yield and constant improvements are being made. However there is an increasing interest for the new biotechnology methods as a complement.

#### **1.1 Breeding in perennial ryegrass**

#### 1.1.1 Lolium perenne L.

Lolium perenne L. (L. perenne) or perennial ryegrass belongs to the Poaceae family which contains the most important crops of the world such as *Oryza sativa* (rice), Triticum aestivum (wheat) or Zea mays (maize). Within the Poaceae family, the Lolium genus comprises eight species including the two major fodder crops L. perenne and L. multiflorum (Italian ryegrass). The origin of these species is believed to be the Mediterranean area from where they spread to North and West Europe, Asia and North Africa (Humphreys 2003). L. perenne is now widely used in North and South America, Australia and New-Zealand. It is a diploid (2n=14) species with a haploid genome size of 2034 Mbp (Farrar et al. 2007) and a two locus self-incompatibility system. This obligate outbreeding permits maintenance of a high degree of genetic diversity (Gill et al. 2006). Lolium perenne is the most important perennial crop in temperate regions in the world (Wilkins 1991). With a long growing season from early spring to the beginning of winter combined with a good grazing tolerance, it is adapted for permanent pasture and forage production. The high level of digestibility and palatability makes L. perenne highly valuable for use in dairy production and sheep forage systems. However it is subject to cold and drought stress (Yamada et al. 2005). In addition to the cattle feeding utility, it is also commonly used as lawn in gardens or sport fields and in other amenity areas.

#### 1.1.2 The trait biomass yield

In forage grasses like *L. perenne*, farmers are expecting a high biomass yield throughout the year. This criterion, together with forage quality, remains the main objective of forage grass breeding and selection (Humphreys et al. 2005, Wilkins 1991).

The biomass production tends to be exponential and is directly linked to the growth of the leaves (Jones and Lazenby 1988). Plants have the ability to transform via the process of photosynthesis the energy from sunlight into organic compounds necessary for their growth. Knowing that 90% of plant dry matter consists of organic compounds, dry matter production is therefore directly related to photosynthesis (Marschner 2002). Two environmental variables are particularly important during the process of photosynthesis: the light (or irradiance) and the temperature. During low irradiance, less carbon is fixed having as a consequence thinner and more etiolated leaves (Jones and Lazenby 1988). Also the incorporation of the carbon into new tissues is highly dependent on the temperature. High temperature enhances the photosynthetic fixation of carbon in leaves but too high temperature can result in the decline in net photosynthesis with the rate of  $CO^2$  evolution increasing more than the rate of incorporation (Marschner 2002, Jones and Lanzenby 1988). For species of temperate climates such as L. perenne, high temperature can damage the tissues and too low temperature might result in the death of the plant (Wilkins 1991). Several other abiotic factors, mainly nitrogen deficiency and water availability (excess or drought) regulate the plant development. However the biomass production is also dependent on genetic control. Individuals within the same species differ in their growth most often due to small changes in several genes (Maloof 2003). Biomass yield is a trait controlled by multiple genes and by the interaction between these genes and the environment. Many studies demonstrated these relations by looking for biomass or growth Quantitative Trait Loci (QTL) in different environments (level of nitrogen, disease, drought,...). And some studies show that different QTL regions can contribute to plant development but that most regions seem specific to a single environment (Rauh et al. 2002). Both genetic and environmental factors determine the final biomass production.

#### **1.1.3 Breeding in** *Lolium perenne*

Until 30 years ago, the main aim of breeding programs was to provide sufficient forage for animal production. Nowadays, it is to reduce the economic and environmental costs of animal production and more traits are taken into account for the selection of new varieties (Wilkins and Humphreys 2003). Although significant improvement in persistency and yield production have been achieved, less progress was made on traits associated with nutritional value, disease resistance, seed yield, drought, freezing temperature and several other environmental stresses. The breeding method is based on recurrent selection consisting of the evaluation of individuals within a population followed by an intercross of the superior individuals to create a new population (Conaghan 2010). Usually the breeding scheme consists of a phenotypic recurrent selection combined with progeny testing and subsequently varieties are constructed and evaluated. However the selection intensity is low since the plot performances can take place only every 10 years (Wilkins and Humphreys 2003). To increase the selection efficiency, full-sib and half-sib progeny tests were incorporated. Another breeding strategy commonly used the doubling of the chromosome number. Tetraploid varieties have a higher yield potential, crown rust and snow mould resistance but less persistency than the diploid ones (Wilkins and Humphreys 2003). A complement to the traditional breeding approach is marker assisted selection based on the use of molecular markers to precisely position the genes. MAS can accelerate the breeding process but will not decrease the time to develop a new variety although for the introgression of a specific recessive gene it will divide the number of cycles by two (Wilkins and Humphreys 2003). Marker assisted selection simplifies phenotyping particularly for traits difficult to measure and decreases the amount of lines tested. In traditional breeding, a large greenhouse or field area is needed, incurring important costs. But the main advantage of MAS is the better precision to select the targeted genotypes (Collard and Mackill 2008, Lamkey and Lee 1993). The introgression of genes controlling quantitative trait such as biomass yield is the principal interest of MAS in *L. perenne*.

#### 1.2 Positional cloning analysis to dissect complex traits

Many important traits in plants are quantitative and under the control of several genes interacting together and with the environment. The trait biomass yield belongs to this category. The dissection of such traits consists of a QTL analysis based on genetic linkage map with the ultimate goal of cloning the genes at the QTL (Figure 1).



Figure 1: Overview of the major steps involved in the dissection of a complex trait with the map-based QTL cloning approach to isolate candidate genes (CG). Adapted from Salvi and Tuberosa 2005.

#### **1.2.1 Population structure for mapping**

The prerequisite of a QTL analysis is the production of a mapping population. The choice of the mapping population is dependent on the objectives of the mapping project, the markers used, the species, and the traits to be mapped (Young 2000). For the construction of the population several criteria have to be considered. The parents selected for the crossing have to have different traits (Collard et al. 2005). Different types of crosses are possible and will mainly depend on the pollination pattern of the plant species. For self-pollinating species the population derives generally from highly homozygous parents (Collard et al. 2005). In cross-pollinating species such as perennial ryegrass there are more options. A linkage mapping study can be performed on an F1 population based on two heterozygous parents providing the possibility to build a map for each parental line when dominant markers are available (Cogan et al. 2006, Faville et al. 2004). But the most common population structures are F2 populations derived from the F1 or from a backcross (BC) population resulting of a cross between the F1 with one of the parents. Their main advantages lie in the ease of construction and the short time of production (Collard et al. 2005). Compared to the BC population, the F2 presents three genotypes at every marker position and thus enables the estimation of dominance components of a QTL (Slate 2005). The F2 segregating population is also the base of the Recombinant Inbred Line (RIL) population. RILs are derived from a cross between two inbred strains followed by repeated selfing over several generations to create new inbred lines (Broman 2005). RILs have less linkage disequilibrium but being based on homozygous lines they are particularly useful in QTL studies since the population can be propagated without further genotyping and the phenotypic evaluation can be replicated in different environments (Tan et al. 2004, Tanksley 1993).

#### 1.2.2 Genetic linkage map

A genetic linkage map provides a simplified representation of the chromosomes showing the position of molecular markers relative to each other (Doerge 2002). The construction requires an appropriate mapping population and one or more DNA marker resource such as Simple Sequence Repeat (SSR), Amplified Fragment Length Polymorphism (AFLP) or Diversity Array Technology markers (DArT). Markers are assigned to a linkage group and put into an order along the group (Mollinari et al. 2009). The distance between two markers is evaluated in centimorgans (cM) and is proportional to the intra-chromosomal recombination events caused by crossing over taking place during meiosis (Peters et al. 2003). The lower the frequency of recombination between two markers, the closer they are and conversely the higher the frequency is, the further away the two markers are supposed to be (Collard et al. 2005). Due to the considerable amount of markers employed to characterize large populations, nowadays the calculation of recombination frequencies is performed with specific software providing the two most common mapping functions. The Haldane mapping function (Haldane 1919) assumes that there is no interference which would increase or decrease the proportion of double crossovers. The Kosambi mapping function (Kosambi 1944) assumes that recombination events influence the occurrence of adjacent recombination events. The linkage between loci is usually calculated using the logarithm of odds (LOD) score. The LOD score compares the likelihood that the two loci are linked, to the likelihood that they are not linked. A significance level of LOD>3 is generally considered as an acceptance level of linkage. The order of markers is usually defined using multipoint linkage analysis that allows an analysis of several markers simultaneously (Mester et al. 2003). The segregation of the markers compared to the Mendelian segregation is tested by the test of chi-square ( $\chi^2$ ) and a deviation of the genetic segregation ratio from their expected Mendelian fraction is called segregation distortion (SD) (Lyttle 1991). Genetic linkage maps help to understand the genome structure and evolution and are essential for genome analysis including mapbased cloning and physical mapping making them useful tools for marker assisted selection (Alm et al. 2003, Bert et al. 1999). A mapping approach is a first necessary step in the identification of QTL.

#### 1.2.3 QTL analysis and mapping

Most important agronomic traits like biomass yield or grain yield are under the control of several genes. Therefore the study of these traits relies on QTL mapping. The concept of QTL was first described by Sax (1923) when he discovered that the complex trait 'seed size' in bean was associated with the single trait 'seed coat colour'. A QTL is a region within a genome containing genes responsible for variation of quantitative traits influenced both by multiple genetic and environmental factors (Doerge 2002). The analysis is based on the detection of an association between the quantitative trait and molecular markers within a population (Slate 2005, Kearsey and Farquhar 1998). QTL mapping involves few basic steps: the construction of a linkage map, the collection of trait data in different environments and finally the combination of the marker data with the phenotypic data by statistical analysis. QTL are detected by the non-random association of alleles at different loci, called linkage disequilibrium (Skøt et al. 2007), between markers and trait values (Mackay 2001). With the development of molecular markers, many genetic maps were generated giving the possibility to search for QTL in many crop species (Young 1996). In parallel the detection of QTL became easier with the elaboration of algorithms allowing the statistical analysis of the association between phenotypic and genotypic data (Doerge 2002). The most common methods of detection are: single marker mapping, interval mapping (IM), composite interval mapping (CIM) and multiple QTL model mapping (MQM). The presence vs. the absence of a QTL at the testing position is statistically evaluated by the LOD score (Doerge 2002). A large LOD score indicates the probable presence of a QTL (Broman 2001). The LOD value for declaring a significant QTL is defined by the permutation test. Nowadays QTL analysis is processed for several applications in genetic research and genetic improvement. The most obvious profit is for MAS by using the markers flanking the QTL intervals but it is also of benefit for gene positioning and cloning. QTL analysis is also powerful for the dissection of complex traits or to understand the heritability of phenotypic variations and genetic adaptations (Asins 2002, Mauricio 2001).

#### **1.2.4 Positional cloning**

Once the QTL is localised to a sufficiently small region, the next step toward the identification of putative candidate genes consists of the identification of the sequences responsible for the QTL, or in other words, the cloning of the QTL. Several strategies allow identification of genes tightly linked to the genetic locus of interest. When there is prior knowledge of the biochemical and signalling pathways involved, a candidate gene approach is an alternative to target known genes probably associated to the trait and is the most adapted for QTL cloning (Kloosterman et al. 2010, Pflieger et al. 2001). But biomass yield has been little studied and nothing is known about the biological functions of the genes implicated. Although less efficient, positional cloning is in that case the most appropriate. Despite many QTL studies reported in plants, only a few QTL have been cloned. It is a long process and getting a map fine enough to reduce the number of genes under the QTL remains difficult. Nevertheless, for the genes *Hd6* in rice (Takahashi et al. 2009) and fw2.2 in tomato (Frary et al. 2000) the number of genes

explaining the observed QTL was narrowed down to a single gene facilitating the process of cloning. And several QTL were successfully cloned inspite of the accumulation of genes within the QTL regions (Salvi and Tuberosa 2005).

Usually a genome Bacterial Artificial Chromosome (BAC) library is screened with the markers underlying the QTL and the isolated clones are sequenced. To identify genes in the new sequences, the synteny between closely related species can be used to identify conserved genes. In addition, the *ab initio* method will permit prediction of genes only on the basis of the local sequence characteristics. The gene prediction programs rely on the statistical qualities of the exons to distinguish a gene from its surrounding sequence (Rust et al. 2002). To assign a function to the predicted gene, databases of known gene sequences are scanned with the sequence of the new gene. Genes functionally related to the trait can be selected as candidates and used for testing (Salvi and Tuberosa 2005).

#### 1.3 Context and objectives

Despite its economic importance, until recently *L. perenne* remained poorly studied from a genetic point of view. The development of a genetic linkage map started at the end of the 1990's and a total of around 20 linkage maps in *L. perenne* have been published so far using all together eight mapping populations of different types (F1, F2, F2 inbred line derived and BC populations). A large proportion of the linkage maps were developed for further QTL analysis. Several studies on QTL mapping were carried out in a large range of morphological and physiological traits such as vernalization response (Jensen et al. 2005a), heading date (Skøt et al. 2005) or water-soluble carbohydrate (Turner et al. 2010). But to date there is a lack of information concerning the trait biomass yield. The traits fresh weight, leaf length and leaf width were first analysed (Yamada et al. 2004) and recently QTL for the traits fresh weight, dry weight,

dry matter and leaf width were identified (Anhalt et al. 2009). Co-localization of these traits was found on linkage groups 2, 3 and 7. The data available are not sufficient for the identification of markers linked to biomass or for the detection of candidate genes. Therefore the project that formed the basis of this thesis "Fine mapping of biomass yield quantitative trait loci in *Lolium perenne* L." aimed to refine and describe the chromosomal regions linked to the trait biomass.

In details the objectives are:

(1) to refine the preliminary map developed by Anhalt et al. (2008) by the addition of more molecular markers of different types on the selected chromosomal regions of LGs 2, 3 and 7.

(2) to map STS markers developed from rice sequences to study the synteny between *L*. *perenne* and rice.

(3) to better define biomass QTL positions with a reduction of the QTL intervals to a few cM

(4) to screen a *L. perenne* BAC library to select clones covering the QTL regions for the identification of *L. perenne* genomic sequences underlying biomass QTL

(5) to sequence BAC clones underlying the QTL to perform a gene prediction and annotation in order to identify potential candidate genes.

#### **Chapter 2: Materials and methods**

#### 2.1 Fine mapping

#### 2.1.1 Plant material

Perennial ryegrass inbred lines were created and used as parental plants for the construction of an F2 inbred line derived population. The parental inbred lines were developed as part of the cytoplasmic male sterility program in Teagasc, Ireland (Connolly and Wright-Turner 1984) and result of a cross between *Festuca pratensis* (female parent) and a *Lolium perenne* cultivar (paternal parent). To produce the maternal inbred line of the F2 population, the ryegrass cultivar 'S24' (IGER) was used while for the paternal inbred line the crossing was made with the ryegrass cultivar 'Premo' (Mommersteeg international BV). The material obtained was backcrossed for several generations to the ryegrass parent, then selfed for ten generations to produce inbred lines. The two inbred lines with a different genetic background have been crossed to produce the F1 generation which was selfed to generate the F2 mapping population. This population is diploid with seven chromosomes in the haploid set and consists of 360 individuals. Using the Genomic In Situ Hybridization approach, it has been demonstrated that no detectable chromosome segments from *Festuca* remained in the population (Anhalt et al. 2008).

#### 2.1.2 DNA extraction

Leaf material was collected from the parents, the F1 and the 360 F2 genotypes. The samples were frozen in liquid nitrogen and ground to a fine powder further transferred to a 50 ml falcon tube. DNA was isolated using a modified CTAB protocol (Doyle and Doyle 1987) as follows: The tubes were filled with 15 ml of CTAB composed of 5 ml of 2X CTAB buffer (2%CTAB, 200mM Tris pH 8, 20mM EDTA pH 8, 1.4M NaCl, 1% polyvinylpyrrolidone, 0.28 mM mercaptoethanol) and 10 ml of 1X CTAB. The tubes were incubated for 1.5 hours at 65°C with the tubes being gently mixed periodically. Once the samples had cooled down to 25°C, 10 ml of chloroform: isoamyl alcohol (24:1) were added and the tubes were vigorously mixed for 60 minutes. After centrifugation at 3750 rpm for 15 minutes, the upper phase was transferred into a new 50 ml tube. The chloroform cleaning step was repeated with 4 ml of chloroform: isoamyl alcohol. The recovered supernatant was incubated in a water bath at 37 °C for 30 minutes with 75 µl of RNase (10 mg/ml). Then, 4 ml of ice cold isopropanol were slowly added to precipitate the DNA which was collected with a sterile hooked glass Pasteur pipette and placed into 1 ml Ethanol/sodium acetate solution (76% Ethanol, 0.2M NaAc) for around 20 minutes. The DNA was transferred for a very short time in 1ml ethanol/ammonium acetate (76% ethanol, 10 nM NH<sub>4</sub>Ac). Ethanol was removed and the pellet dried at room temperature was resuspended in 1ml TE buffer (1X). The quantity and the quality of the DNA were measured by spectrophotometry.

#### 2.1.3 SSR markers analysis

Primers used in this study are ryegrass specific SSR markers chosen from a number of public and non-public sources (Gill et al. 2006, Van Daele et al. 2008, King et al. 2008). A set of 130 primer pairs was selected to map on LGs 2, 3 and 7. In addition 23 EST-SSR markers developed for a consensus map were tested (Studer et al. 2010).

All primer pairs were initially evaluated for amplification efficiency and for polymorphism using the parental and F1 genotypes. Polymorphic markers were then screened in the entire F2 population. PCR reactions were conducted in 10 µl reaction volume containing 40 ng of DNA, 1 µl of 10X buffer, 0.2 µl of dNTP (10 mM), 0.25 µl forward and reverse primer (10 µM) and 0.06 µl of *Taq* DNA polymerase (New England Biolabs). The forward primers were fluorescently 5'-labelled with FAM (Metabion, Germany). PCR was carried out using a programmable thermocycler (Biometra Thermocycler, Germany) with the program adapted to the markers source (Table 1, Annex II). GeneScanTM500 LIZ as a size standard and formamide were added to the PCR product. After denaturation at 95 degrees during five minutes, the samples were analysed on an ABI3130XL using a 36 cm capillary array with POP-4 polymer (Applied Biosystems, Warrington, UK). The software GeneMapper V3.7 was used to score the data and inconsistent amplification was scored as missing data.

Primer source	Temp	Time (s)	Cycles	Markers mapped
	96°C	300	Initial denaturation	
ABERS, UK <sup>*</sup> (King et al. 2008)	96°C AT 72°C	15 30 30	35 cycles	LpHCA17C11
	72°C	240	Final extension	
	94°C	300	Initial denaturation	
Studer et al. (2010)	94°C AT+12 °C 72°C 94°C	30 60 60 30	- 12 cycles (touch down -1°C) G01-034	
	AT 72°C	60 60	30 cycles	001-010
	72°C	300	Final extension	
Van Daele et al. (2008)	96°C	300	Initial denaturation	
	96°C AT 72°C	15 30 30	35 cycles	
	72°C	240	Final extension	
	96°C	15	Initial denaturation	
ViaLactia, New Zealand <sup>*</sup> (Gill et al. 2006)	AT+10 °C 72°C	30 30	10 cycles (touch down -1°C)	rv0347-rv0122 C) rv1282-rv0959 rv0027 rv1212
	72°C	240	Final extension	rv1068-rv1144
	94°C AT 72°C	30 30 30	rv0433-rv04 25 cycles rv0459-rv11 rv1060-rv13	
	72°C	600	Final extension	

Table 1: Polymerase chain reaction programs for simple sequence repeat markers from different sources used for the construction of the linkage map.

AT: marker specific annealing temperature

(<sup>\*</sup>): Marker origin via licence agreement

#### 2.1.4 STS markers design and analysis

Rice STS markers were designed by aligning rice genomic and cDNA sequences with expressed sequence homologues from two to three different Poaceae species in order to identify conserved nucleotide sequences. The sequences were obtained on the rice genome annotation project webpage (http://rice.plantbiology.msu.edu/). Following the synteny found between rice and ryegrass, STS primers were designed from sequences of rice chromosomes 4 and 7 to map on ryegrass LG 2 and from rice chromosomes 1 and 6 to map on LGs 3 and 7, respectively. The multiple sequence alignments were performed with the software Macaw 2.0.5. Primers were designed to be anchored in exons and to cover an intronic region of 600-800 bp to increase the possibility of identifying polymorphisms (Lem and Lallemand 2003). The primers consist of 17 to 20 nucleotides with a GC percentage between 35 and 55%. Primers were named after the corresponding rice locus. After an initial evaluation for amplification efficiency and polymorphism in the parental and F1 genotypes, the polymorphic primers were screened in the entire F2 population. The PCR reaction volume was similar to the conditions for the SSR markers but the PCR was carried out with a different program (Table 2).

515 markers		
Temperature	Time (s)	Cycle
94°C	60	Initial denaturation
94°C	30	10 avalas
AT+10 °C	60	(touch down 1°C)
72°C	60	(touch down -1 C)
94°C	30	
AT	60	30 cycles
72°C	60	
72°C	60	Final extension

 Table 2: Polymerase chain reaction program for the rice

 STS markers

AT: marker specific annealing temperature

The amplicons of non-polymorphic STS primers were sequenced for the parents and the F1 to detect SNPs using the forward primer. Sequencing was outsourced to a company (AGOWA GmbH, Germany). The software NEBcutter V2.0 (Vincze 2003) was used to determine the restriction enzyme specific to the SNP. 15  $\mu$ l of PCR product was digested for 3 hours at 37 °C in 20  $\mu$ l reaction volumes using one unit of the appropriate restriction enzyme and 2  $\mu$ l of 10X buffer specific to the enzyme. For the STS and SNP markers, the reactions were separated by electrophoresis on 3% agarose gel stained by adding 0.5  $\mu$ g/ml of ethidium bromide, electrophorated at 100 volts for at least 1.5 hours in 0.5X TBE buffer.

#### 2.1.5 Linkage map construction

SSR and STS markers were used to generate genotypic data for the F2 population. The data were combined with previous data collected for this population (Anhalt et al. 2008). Linkage analysis and map construction were conducted using JoinMap 3.0 (Van Ooijen and Voorrips 2001). The markers were classified in two segregation types: the co-dominant markers were expected to segregate in a 1:2:1 pattern while the dominant in a 1:3 ratio. The determination of linkage groups was carried out with a logarithm-of-odds (LOD) score threshold between 4.0 and 6.0 using the data for the seven linkage groups. The ordering of markers within the linkage group was performed with a LOD score larger than 1.0, a maximum recombination value of 0.4 and a jump threshold of 5.0. The Kosambi's mapping function was applied to calculate the map distances in cM (Kosambi 1944). Marker segregation ratios were checked for deviation from Mendelian expectation by chi-square ( $\chi^2$ ) analysis to reveal regions with significant segregation distortion. The graphical display of the map was drawn using MapChart V2.2 (Voorips 2002).

#### 2.2 DArT analysis

Genomic DNA from the parents, the F1 and the 360 F2 genotypes was sent to the Diversity Array Technology P/L (Canberra, Australia) for analysis with their hybridization based markers. The markers were generated using the DArT array developed for the *Lolium/Festuca* complex (Kopecky et al. 2010). Polymorphic DArT markers were scored as present (1) or absent (0) in the different genotypes. The presence vs. absence DArT scored were converted into codes compatible with Joinmap 4.0 (Van Ooijen 2006) by comparison with the appropriate parental DArT assay. Thus for markers present in the paternal line, the value 0 was converted into a and 1 into c. When the markers were present in the maternal line, the genotypes scored 0 were converted into b and the genotypes scored 1 were converted into d. To reduce the risk of errors due to missing data, genotypes without information for the DArT markers were removed and a total of 325 genotypes remained. All dominant markers were grouped according to the parent they were segregating from and codominant markers were in both subsets. The two sets of markers in coupling phase (a/c scores combined with a/h/b and b/d scores combined with a/h/b, with a/h/b corresponding to the codominant markers) were loaded separately into JoinMap 4.0. A first determination of the seven linkage groups was carried out but it was not possible to get seven groups. A first selection of markers was performed by removing the markers with a  $\chi^2$  value above 25 and looking at the similarity of loci, identical markers above 0.98 were excluded (only one was kept). The assignment to the linkage groups was then performed a second time and all markers fell into seven groups. The longest map was kept as fixed order for a subsequent round of calculation with the entire set of markers. The estimated and the final maps were calculated using the Maximum Likelihood mapping function with three rounds of map optimization. The graphical representation of the map was drawn using MapChart V2.2 (Voorips 2002)

#### 2.3 QTL analysis

#### 2.3.1 Phenotypic data (Anhalt et al. 2009)

Phenotypic data were obtained from a field and greenhouse experiment performed during the period of December 2005 to August 2007. The experiment consisted of five replications. Each replication comprised 45 incomplete blocks of eight F2 genotypes and one check (either the maternal, the paternal or the F1 line). Two replications were planted in a field environment. In the greenhouse a trial in three replicates was set up with two out of the three replications in the following conditions: average nightly and daily temperatures of 11°C and 19°C respectively. The third replication was planted in a separated greenhouse in the following conditions: average nightly and daily temperatures of 6°C and 19°C respectively Three harvests from the greenhouse experiment were carried out in December 2005, February 2006 and April 2006. For the field experiment, four harvests were done in August 2006, October 2006, May 2007 and August 2007. The traits fresh weight and dry weight (in g) were measured and the trait dry matter (in %) calculated as follows: (dry weight/fresh weight)\*100. The heritability of traits and the distribution of data for each trait and each harvest were also calculated.

#### 2.3.2 QTL calculation

Data for each trait and for each experiment were combined with the genetic map information using the software MapQTL 6.0 (Van Ooijen 2009) and WinQTL cartographer 2.5 (Wang et al. 2010) to detect the QTL. The operation was performed with the two maps previously generated (without and with the DArT markers) to allow a comparison and see the effect of the addition of a high number of dominant markers. The statistic test Interval Mapping (IM) was first applied to estimate the QTL position. This test was performed on the seven linkage groups to detect possible new QTL after the addition of markers.

A second test, Multiple QTL Model (MQM) mapping specific to MapQTL was then used to get a more precise localisation of the QTL position. The cofactors were initially selected manually by taking the markers showing the highest LOD score in IM. The significance of the cofactors was tested by the automatic cofactor selection analysis and only the markers with a high significance were kept as cofactors. MQM was run on a trait by trait basis using all of the identified cofactors for each trait. To confirm the position of the QTL, an equivalent test, Composite Interval Mapping (CIM) from WinQTL cartographer 2.5, was processed. The selection of the cofactors was automatic. Both tests were performed with a walk speed of 2 and a number of neighbouring markers of 5. Significant (P $\leq$ 0.05) QTL were declared at a genome wide LOD thresholds level determined by permutation testing (Churchill and Doerge 1994) with 1000 permutations.

#### 2.4 BAC library

#### 2.4.1 Characteristics of the BAC library

The *L. perenne* BAC library used for this project was developed by the University of Arizona (US). The BAC library was constructed from a single genotype of an inbred line population using the restriction enzyme Hind III and the vector pAGIBAC1. The BAC library has an average insert size of 134 kb and consists of a total of 120 960 clones (315 plates) for a 5 genome equivalents coverage.

#### 2.4.2 Replication of the library

To avoid any loss and risk of contamination, a "work" copy of the library was made. For every microtiter plate in the master copy of the library, a microtiter plate was filled with sterile freezing medium composed of 2.5% LB broth, 13 mM KH<sub>2</sub>PO<sub>4</sub>, 36 mM K<sub>2</sub>HPO<sub>4</sub>, 1.7 mM sodium citrate, 6.8 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 4.4% (v/v) glycerol and

2.5% (v/v) of Chloramphenicol. The replication was performed using a metal hand-held replicator. The plates were incubated at  $37^{\circ}$ C overnight then stored at  $-80^{\circ}$ C.

#### 2.4.3 Pooling of the BAC DNA

The pooling strategy combined a screening of library plates followed by a screening of rows and columns within positive plates assembled in pools. For the construction of DNA "plate-pools" of the library, each of the 315 plates was replicated on individual LB agar plates using a 384 pin replicator. Agar plates were incubated at 37°C for a minimum of 12h until the resulting bacterial colonies were around 1 mm diameter. Subsequently, 4 ml of LB broth was added to each plate, into which the colonies were scraped off and removed to two 2 ml eppendorf tubes. The tubes were centrifuged at 13,000 rpm for 2 minutes. The supernatant was removed and one tube was stored at -80°C as stock for further analyses while the second was used to process the BAC DNA isolation. The pellet was resuspended in 200 µL of chilled buffer I (50 mM glucose, 10 mM EDTA, 25 mM Tris-cl, pH 8.0) with 50 µg/mL of RNAse and the tubes were kept on ice. 200 µL of fresh solution II (200 mM NaOH, 1% SDS) were added and after few inversions of the tubes for mixing, 200 µL of chilled solution III (5M KOAc, pH 4.8) was incorporated. The tubes were inverted again to mix and centrifuged at 13000 rpm for 15 minutes. The supernatant was transferred into a new tube with 400 µL of cold isopropanol and centrifuged at 13,000 rpm for 20 minutes. The pellet was resuspended in 100 µL of TE (1X), 10 µL of 3M NaAc and 250 µL of absolute ethanol (-20°C) before being placed for 1 hour at -20°C. After centrifugation at 13,000 rpm for 20 minutes, the pellet was washed with 500 µL of 70% ethanol before being resuspended in 30  $\mu$ L of TE (1X). The DNA concentration was then measured with a Qubit (Invitrogen).
For the plates showing a positive signal during the screening, a modified pooling strategy was applied to identify the position of the positive clones. This strategy is based on the assumption that any single copy sequence in the genome will occur at a maximum of once in the positive plate. The pooling strategy is illustrated graphically in Figure 2. The first 2 dimensions correspond to the step described previously. The third dimension consists of 4 pools, each comprising 4 consecutives rows of the positive plate. Similarly the fourth dimension has 4 pools of 6 consecutive columns. The fifth dimension is composed of 4 pools, each comprising one row from each pool of the third dimension. Finally the last dimension consists of 6 pools, each comprising one column of each pool in the fourth dimension. The next steps are similar to the plate pooling.



Figure 2: Pooling strategy combining a "plate pool" screening (dimension 1 and 2) followed by the screening of rows and columns assembled in pools (dimensions 3 to 6).

# 2.4.4 Screening of the BAC library

The BAC library was screened against the SSR and DArT markers flanking the QTL regions using a PCR-based screening procedure. Before the addition of the DArT markers to the linkage map, the screening was performed on the modified 6D pooling described previously (part 2.4.3: Pooling of the BAC DNA). PCR reactions were conducted in 10  $\mu$ l reaction volumes containing 10 ng of DNA, 1  $\mu$ l of 10X buffer, 1  $\mu$ l

of dNTP, 0.2  $\mu$ l forward and reverse primer (10uM) and 0.2  $\mu$ l of Taq polymerase. PCR were carried out using a programmable thermocycler (Biometra Thermocycler) using the program shown in Table 3.

used for the screening of the DAC horary.						
Temperature	Time (s)	Cycle				
94°C	180	Initial denaturation				
94°C	45	5 cycles				
AT + 5 °C	45	(touch down 1°C)				
72°C	90	(touch down -1 C)				
94°C	45					
AT*	45	30 cycles				
72°C	90					
72°C	600	Final extension				

 Table 3: Polymerase chain reaction program for markers used for the screening of the BAC library.

Once the DArT markers were added to the linkage map, a pooling of the BAC library was available from Amplicon Express (US). Our pooling strategy was efficient for the few SSR markers we had initially. But with the high number of DArT markers, this strategy would have taken too much time and the pooling elaborated by Amplicon Express was better adapted. However, all markers without positive signal with this pooling were tested a second time with the modified 6D pooling. Prior to the screening, primers for the DArT markers had to be designed. Sequences of the DArT were obtained from David Kopecky (Institut of Experimental Botany, Czech Republic). Primers were designed with Primer 3' (Rozen and Skaletsky 2000) and consisted of 17 to 20 nucleotides with a percentage of CG between 40 and 60%. All primers were tested for amplification with genomic DNA from *L. perenne*. PCR was performed in 10  $\mu$ l reaction volumes with 1  $\mu$ l of DNA, 1  $\mu$ l of 10X buffer, 1  $\mu$ l of dNTP (10mM), 0.5  $\mu$ l forward and reverse primer (10uM) and 0.5  $\mu$ l of Taq polymerase. The amplification conditions were similar to the modified 6D pooling. For both pooling, the PCR products

were analysed in a 1% agarose gel with a sample of genomic DNA from the *L. perenne* variety Cashel as a positive control.

# 2.4.5. AFLP fingerprinting

The selected BAC clones were fingerprinted using AFLP to observe the presence or absence of contigs between clones from different markers. BAC DNA was isolated using the protocol described previously (part 2.4.3: Pooling of the BAC DNA). DNA was digested overnight at 37°C in a final volume of 20  $\mu$ L containing 250 ng of DNA, 2.5 units of Eco RI and 1.25 units of Mse I, 1X enzyme buffer, and 1X of BSA. After digestion 5  $\mu$ L of ligation mixture containing 1 $\mu$ M of Eco RI adapter, 10  $\mu$ M of Mse I adapter, 5X of ligase buffer and 40U of T4 DNA ligase (New England Biolab) were added to the digested reaction before incubation for 2 hours at 20°C. The sequences of the adapters were as follow:

# Mse I adapter: 5' GACGATGAGTCCTGAG 3' 5' TACTCAGGACTCAT 3'

# Eco RI adapter: 5'CTCGTAGACTGCGTACC 3' 5'AATTGGTACGCAGTCTAC 3'

Amplification of the ligated products was performed using non-selective primers complementary to the adapters.

# MseI primer: 5'-GATGAGTCCTGAGTAA-3' EcoRI primer: 5'-GACTGCGTACCAATTC-3'

About 5  $\mu$ L of the ligated products were used as template DNA and amplification reaction was performed in a 20  $\mu$ L final volume containing: 30 ng of primer MseI, 50 ng of FAM-labelled primer EcoRI, 0.1  $\mu$ L of dNTP (2 mM), 0.1  $\mu$ L of 10X buffer and

0.1  $\mu$ L of Taq polymerase (5U/  $\mu$ L). The reaction was carried out in a programmable thermocycler (Biometra Thermocycler) with the program presented in Table 4.

Table 4: AFLF amplification program						
Time (s)	Cycle					
30	12 ovelos					
30	15 Cycles					
60						
30						
30	23 cycles					
60						
	30         30         30         30         60         30         30         30         60         30         30         60         30         30         60<					

Table 1. AFI D amplification program

The amplification products were checked on a ABI3130XL (Applied Biosystems, Warrington, UK) and fragment sizes were determined with the software GeneMapper V3.7. The output fragment size data were input to a GenoProfiler to produce a functional dataset for the software Fingerprinted Contig (FPC, Soderlund et al. 1997) which was used for automatic contig assembly of the selected clones. In the FPC analysis, a fixed tolerance of seven was used and contig assembly was performed at a cut-off value of 1e-12.

#### 2.4.6 BAC clones sequencing and gene prediction

Five BAC clones were sent for sequencing using a Roche GS FLX (GATC, Germany; Table 5). The BAC clones were identified with markers underlying the biomass QTL peak on LGs 2 and 3. For the marker LoPt 355957, only one clone could be identified. For G04-054 and M15185, using the results obtained with FPC contig, two clones were selected per marker to cover the largest genomic region. The assembly of the BAC sequences was performed by GATC, with the GS De novo Assembler (Newbler) software.

Table 5: List of the clones sent for sequencing				
Plate Address	Marker			
20I19 and 222J11	G04-054			
180D4 and 30C14	M15185			
153A21	LoPt 355957			

Genes were predicted using the program Rice Genome Automated Annotation System (RiceGAAS, http://ricegaas.dna.affrc.go.jp/) which combined different *ab initio* gene finders. A gene function was assigned to the predicted gene structures by BLASTn and BLASTp alignments. The cutoff threshold applied was a E value of 10<sup>-5</sup> or less and a

sequence identity above 60%.

For the genes identified known to have potentially an effect on the biomass yield, markers related to the genes were designed on the gene sequences using Primer 3' with the following conditions: 17 to 20 nucleotides with a GC percentage between 35 and 55%. Primers were named after the corresponding genes. A PCR amplification was initially performed on the parental and F1 lines using the conditions described previously for the STS markers from rice (part 2.1.4. STS markers design and analysis). Amplicons showing a clear single band were sent for sequencing (GATC, Germany) to detect SNPs using the forward primer. Polymorhic markers were then used to screen the entire population and the analysis was carried out by sequencing all the PCR products. The scoring data were added to the map generated with the DArT markers keeping the same setup for the Maximum Likelihood analysis and using the marker order of the map as fixed order.

# Chapter 3: Mapping of ryegrass specific SSR markers and rice STS markers

#### **3.1 Introduction**

The origin of linkage maps dates back to the beginning of the 20<sup>th</sup> century with the discovery of linkage between genetic factors by Morgan (Paterson et al. 1991). However due to the mathematical and practical complexity of generating genetic maps, this method was largely abandoned. Only in the 1980's, with the advent of advanced DNA molecular marker technologies, the construction of linkage maps became a basic tool in the characterization of important traits leading to the possibility for QTL mapping, marker assisted selection, gene isolation and cloning. Genetic maps are also powerful tools in the study of genome structure and of relation between genomes of different species. Linkage maps are a prerequisite for the detection of QTL and the construction of primary maps with few markers is straightforward. But the map density is often too low with a range of intervals between markers around 10-30 cM, which covers several hundred genes (Salvi and Tuberosa 2005, Mott 2006). Nowadays genetic maps have been developed for a wide range of species and further work is focusing on the development of detailed maps.

### 3.1.1 Fine mapping

Obtaining a 'fine map' with a distance of 1-2 cM between markers is necessary to better define QTL positions and thus to facilitate the use of the map-based cloning to isolate genes linked to the trait of interest (Ronin et al. 2003, Mott 2006). Different methods are available for fine mapping and include the selection of certain genotypes, the increase of the mapping population size or of the number of markers (Melchinger 1998, Ronin et al. 2003, Vales et al. 2005, Xu et al. 2005). A large population reduces the impact of incorrect genotyping scoring when screening the population with polymorphic markers and is a way to accumulate more recombinants in the interval of interest (Ronin et al 2003). Accumulation of recombinants can be done by the selection of recombinant genotypes in the population. However different studies indicated some limits to this method (Melchinger 1998, Vales et al. 2005). The development of highly polymorphic DNA molecular markers which are easy to use and transferable between populations and/or species has facilitated the creation of saturated maps. Molecular markers are specific fragments of DNA that show a variation of the DNA sequence among the different individuals of a population (Jones et al. 1997). Restriction Fragment Length Polymorphism (RFLP) markers were first developed for use in the production of genetic maps in humans (Botstein et al. 1980), and were quickly adapted to plants (Helentjaris et al. 1985). Over the last three decades new generations of markers have been introduced (Collard et al. 2005). Compared to morphological or isozyme markers, DNA molecular markers are abundant throughout the genome, completely independent of environmental conditions, can be detected at any stage of development of the plant and do not disturb the physiology of the organism (Mohan et al. 1997, Jones et al. 1997). They are tools in various fields and are nowadays widely used in crop improvement, for cultivar identification, parental analysis, synteny mapping, genome mapping and tagging of agronomically important genes (Joshi et al. 1999, Saha et al. 2005).

Molecular markers can be classified as dominant or codominant. RFLP or Simple Sequence Repeat (SSR) markers are codominant and can discriminate between homozygote and heterozygote genotypes allowing the determination of genotypes and allele frequencies at loci. With molecular markers, the fine mapping can be focussed on a specific region of interest and the varieties of markers available give the possibility to select the most useful for the study. The markers can be selected in function of the population, the species, the amount of DNA available and rely on the objectives of the study. Some markers, such as AFLP, are able to saturate maps, but they require a large amount of DNA and are time consuming. On the other hand, SSR markers are efficient tools widely used for their high degree of polymorphism, abundance, distribution along the genome and ease of use (Powell et al. 1996, Kuleung et al. 2004, Saha et al. 2006). But detailed knowledge of DNA sequences are required for their development and few SSRs are available for species with low genomic sequence resources. For these species another approach, which consists of the use of the synteny between species to transfer markers between species or to develop additional markers.

#### 3.1.2 Synteny

Plant genomes show considerable variation in chromosome number and genome size even within the same family. In the grass family alone genome sizes vary from around 450 Mbp in rice to around 16000 Mbp in wheat (Schmidt 2002). This phenomenon is largely due to repetitive DNA duplication or deletion (Bennetzen et al. 1998). Despite these length variations, the majority of plant genomes reveal a high conservation of some chromosome segments called synteny (Renwick 1972, Keller and Feuillet 2000, Eckardt 2001). With the development of molecular markers and genetic maps it became possible to compare genetic maps and to investigate the degree of conservation of markers within chromosome segments by using the same set of molecular markers in closely related species (Gale and Devos 1998). Thus, a considerable amount of conservation of chromosome segments has been demonstrated, initially within the Solanaceae (Bonierbale et al. 1988). Some generalisations on the

genome structure could be assumed and well studied plant species such as rice or *Arabidopsis* could be used as models. Genomic knowledge gained on important agronomic traits from these model crops could be transferred to related species with larger or more complex genome structures. A further expectation is to facilitate the isolation of agronomic traits using information on gene conservation (Schmidt and Heslop-Harrison 1998).

A remarkable conservation of gene content and gene order has been established in the grass family and diverse studies confirmed that all species within this family diverged from the same common ancestor (Ahn and Tanksley 1993, Kurata et al. 1994). Moore et al. (1995) showed that the genomes of major grass species can be aligned and combined in a single synthesis. The rice genome, one of the smallest among the grasses, has been completely sequenced and can be used as a model plant. The conservation of gene order in grass species can be described in terms of "rice linkage blocks" (Moore 1995). In this way, nine genomes differing in basic chromosome numbers from five to twelve chromosomes and nuclear DNA contents from 400 to 6000 Mb were described in only 25 "rice linkage blocks" (Gale and Devos 1998). A few years later, this comparison was applied to further species including forage grasses such as perennial ryegrass and meadow fescue (Devos 2005). A first comparative map published in 2002 between L. perenne and some cereals demonstrated a high degree of orthology and colinearity as already observed between different species within the Poaceae (Jones et al. 2002a). A presume good syntenic relationship has been observed for each ryegrass linkage group with the corresponding homoelogous chromosome of the Triticeae. A good conservation was shown between L. perenne and rice. Ryegrass LGs 3 and 6 share a perfect synteny with rice chromosomes 1 and 2, respectively. Large chromosome rearrangements have been identified for the other linkage groups which were syntenic with two or more rice chromosomes. LG 2 is syntenic to rice chromosomes 7 and 4 and LG7 to rice chromosome 6 and 8 (Sim et al. 2005).

#### **3.1.3** Context and objectives of this chapter

A partial linkage map was published in 1994 for an interspecific cross between L. perenne and L. multiflorum (Hayward et al. 1994). The first linkage map specific to L. perenne was constructed based on AFLP markers (Bert et al. 1999), and using the same population Jones et al. (2002b) developed the first map based on SSR markers. Aided by the development of molecular markers, additional maps were constructed and previous maps extended. To date more than 20 L. perenne maps have been developed based on different populations and marker types. Some maps were aligned with those of other grass species such as rice, barley or wheat (Jones et al. 2002a, Sim et al. 2005) to understand the synteny between species and the evolution within the grasses. But most maps were established for further QTL analysis to develop the MAS approach. However only one of the published maps was specially developed for analysis of the trait biomass. Biomass yield is one of the essential agronomic traits in forage crops, and increasing biomass yield is the most important objective of breeding programmes. In a precursor study, a genetic map of L. perenne was constructed based on an inbredderived F2 population segregating for that trait. A total of 75 SSR and AFLP markers were mapped on the seven linkage groups (Anhalt et al. 2008). With an average distance between markers of 7.5 cM the map was powerful for the detection of biomass yield QTL which were detected on LGs 2, 3 and 7. However the map resolution was too low for a fine definition of the QTL regions and thus not useful for MAS or for the identification of candidate genes linked to the trait biomass.

The work presented in this chapter focuses on the fine mapping of this preliminary map by the integration of more molecular markers on the selected chromosomal regions of LGs 2, 3 and 7. This study used a dual approach to add a maximum number of markers. One approach is the use of ryegrass specific SSR markers. Some were already successfully mapped and showed high polymorphism. Thus they were ideal for fine mapping. The second approach used synteny relationships between *L. perenne* and rice to develop STS markers. The rice STS markers were also used to reveal the synteny between the two species. Information from this genetic linkage map will be used further to better define the position of biomass yield QTL.

#### 3.2 Results

#### 3.2.1 Evaluation of the SSR and STS markers

A total of 130 *L. perenne* specific SSR primer pairs from different sources were tested for genetic polymorphism using the parents, the F1 and a subset of F2 genotypes from an inbred line derived population (Figure 3). Using information from published maps (Gill et al. 2006, Van Daele et al. 2008, King et al. 2008), the primers were selected to map to LGs 2, 3 and 7, regions where biomass yield QTL have been detected (Anhalt et al. 2008). An additional set of 23 EST-SSR markers developed for the construction of a consensus map (Studer et al. 2010) and non specific to the region of interest were also included in this study. An amplification signal was produced by 90% of the primer pairs. This result was expected since the primers were specific to *L. perenne*. However only 31 markers (22%) shown polymorphism and could be scored on the entire population. The polymorphism degree was very low and varied from 0% to 70.5% depending on the SSR sources (Table 6). To avoid any scoring errors, all primers producing weak or equivocal amplification were considered as monomorphic

which can partly explain the low level of polymorphism. Anhalt et al. (2008) also obtained a low polymorphism degree of 45% for the same population.

obtained after testing on the parental mes, the 11 and a subset of 12 genotypes							
Markers' source	Number of tested primers	Amplified primers Number %		Number of polymorphic primers	Number of monomorphic primers	Polymorphism degree in %	
	Pillions	1 (0111001	/0	primero	primers		
Studer <sup>c,1</sup>	23	17	74	12	5	70.5	
Van Daele <sup>2</sup>	13	13	100	0	13	0	
IBERS, UK <sup>*,3</sup>	23	20	87	2	18	10	
Vialactia, NZ <sup>*,4</sup>	94	87	95	17	72	19.5	
Total	153	139	91	31	108	22	

Table 6: Summary of the number of *Lolium perenne* specific SSR primers tested for the different sources with the amplification rate and the polymorphism degree obtained after testing on the parental lines, the F1 and a subset of F2 genotypes

Markers' origin via licence agreement (\*), from a project of a consensus map (<sup>c</sup>) or from public sources: <sup>1</sup>Studer et al. 2010, <sup>2</sup>Van Daele et al. 2008, <sup>3</sup>King et al. 2008, <sup>4</sup>Gill et al. 2006



Figure 3: An example of an ABI chromatogram that shows the allelic pattern of a *L. perenne* specific SSR marker (here rv 0433 from Vialactia (New Zealand)) on the maternal line (PM), the paternal line (PFa), the F1 and a selection of F2 genotypes (F2/number of the genotype)

In addition to the SSR primers, 126 STS primer pairs were designed from rice sequences to map on LGs 2, 3 and 7, according to the synteny existing between the species (Sim et al 2006, Annex I). Primer pairs were successfully designed with 78.5% producing a clear amplification (Table 7). As for the SSR markers, the polymorphism remained low and varied between 0% and 15.8% depending on the respective linkage group. Three markers showed a clear size polymorphism (Figure 4). Low polymorphism rates could be explained technically in part by the limited resolution of the agarose gel electrophoresis used for the screening.

Table 7: Summary of the number of rice STS primers designed for each linkage group with the amplification rate and the polymorphism degree obtained after testing on the parental lines, the F1 and a subset of F2 genotypes

		,			0 1	
Number		Amplified primers		Number of	Number of	Polymorphism
primers	Number	%	polymorphic primers	primers	degree in %	
2	24	19	79.2	3	16	15.8
3	77	70	90.9	2	68	2.85
7	25	10	40	0	10	0
Total	126	99	78.5	5	94	4





To overcome the lack of polymorphism, the amplicons of non-polymorphic STS primers showing a clear and simple amplification were sequenced and only 2 out of 35 contained SNPs indicating a population with little polymorphism. After digestion with the appropriate restriction enzymes, the markers LOC\_Os01g36890 and LOC\_Os01g16152 (Figure 5) were used for mapping.



Figure 5: Chromatogram (top) of the sequences obtained with the marker LOC\_Os01g16152 for the maternal line (M) and the paternal line (Fa). The black arrow showed the SNPs. Amplicons were digested with the enzyme Hinf I and the reactions were separated by electrophoresis on 3% agarose gel (bottom). The primer was tested on the mother (M), the father (F), the F1 generation (F1) and five genotypes from the F2 generation (F2) to confirm the polymorphism before scoring the entire population.

#### **3.2.2** Construction of the linkage map.

Molecular marker data were collected from a population of 360 F2 genotypes. The genetic mapping dataset was constructed using the data of the polymorphic SSR and STS markers in combination with data from the entire set of AFLPs and SSRs previously mapped (Anhalt et al. 2008). SSR and STS markers were scored as codominant, but the two SNPs were scored as dominant. Initial groups were formed using a LOD score range between 4 and 6 in JoinMap 3.0 (Van Ooijen and Voorrips 2001) and three SSR loci were automatically excluded from the grouping. The 21 remaining SSR and STS markers were assigned as expected to LGs 2, 3 and 7 and also two of the EST-SSR (Figure 6).



Figure 6: Molecular marker genetic map on the LGs 2, 3 and 7 of a F2 population of *Lolium perenne* including the 18 new SSR markers (in red) and the 5 rice STS markers (in green). The map was generated in Joinmap V3.0 using the Kosambi's mapping function. Loci labelled with *asterisks* indicates the segregation distortion (\* P<0.1, \*\*P<0.05, \*\*\*P<0.01, \*\*\*\*P<0.005, \*\*\*\*P<0.001, \*\*\*\*\*\*\*\*P<0.0001)

In total, 23 markers were integrated into the region which previously contained 38 markers (Anhalt et al. 2008), which increased the map length moderately from 289.2 cM to 313.5 cM. The additional markers led to few changes in the map order and the AFLP markers remained at the telomeric region while the SSRs tended to cluster at the centromere although some newly mapped SSRs clustered at the end of the chromosomes. Despite a reduction of the marker interval from 7.5 to 5 cM, seven large gaps remained. The largest gap was found on LG7 with an interval of 26 cM.

Markers with distorted segregation ratios were present in the three linkage groups (Figure 6). Sixty seven percent of the markers mapped showed significant segregation distortion (SD). The SD was not substantially influenced by marker types but the significance values differed; one locus showed distortion at P<0.1, nine at P<0.05, one at P<0.01, eight at P<0.005, two at P<0.001 and twenty at P<0.0001. The highest level of distortion occurred on LG3 were 75% of the markers were skewed with only one telomeric region not distorted. Similar distortion for this linkage group was previously observed (Bert et al. 1999, Jones et al. 2002a, Armstead et al. 2002, Anhalt et al.2008). For LGs 2 and 7 the SD is distributed along the chromosomes. Theoretically for codominant markers without SD the ratio is 25% for each class of homozygote alleles and 50% for the heterozygote alleles. For dominant markers segregation in a 3:1 ratio is expected corresponding to the presence vs. the absence of the allele. Figure 7 shows the ratio obtained for each marker. While for LG3 the distortion is unilateral favouring alleles from the maternal line, on LGs 2 and 7 the favoured genotype varies for each marker.



Figure 7: Frequency distribution of the genotypes for each marker in the F2 population. The X axis corresponds to the position along the chromosome.

#### 3.2.3 Synteny

The five polymorphic markers developed from rice chromosome 4 and 1 mapped as expected on *L. perenne* LGs 2 and 3 respectively. Twenty four sequences from amplicons of non-polymorphic STS primers were blasted against the rice genome (http://rice.plantbiology.msu.edu/) to compare the similarity between the two species. A cutoff was applied with E value of  $10^{-5}$  or less and a sequence identity of 60% or more. Sixteen sequences showed homology with the expected rice gene from which they were developed (Table 8, Table 9). The percentage of identity varied between 62 and 73.06% for LG2 and, 57.09 and 94% for LG3. For half of the sequences, there was no discontinuity in the alignment. No relation was found between the discontinuity and the product size. Gaps could be present for a sequence of 279 bp and none for a sequence of 923 bp. Also eight sequences presented homology with a second gene on another rice chromosome. For eight sequences it was not possible to find identity with any part of the rice genome.

Table 8: Alignment between *L. perenne* and rice. *L. perenne* (Lp) is represented by the sequences from non-polymorphic amplicons from rice STS designed from rice chromosome 4. Rice is represented by genes showing orthology with the sequences.

Sequences from the STS amplicons	Annotated orthologous genes in rice genome	Number of gaps	E value	Highest % Identity
Lp-LOC_Os04g55050	LOC_Os04g55050	1	6.2e-27	71
Lp-LOC_Os04g55150 Lp-LOC_Os04g55180	LOC_Os04g55150/55180	0	1.3e-11	73.06
	LOC_Os07g30810	0	7.4e-08	60.39
	LOC_Os08g43400	0	7.6e-07	59.4
Lp-LOC_Os04g55220	No identity found			

Table 9: Alignment between *L. perenne* and rice. *L. perenne* is represented by the sequences from non-polymorphic amplicons from rice STS designed from rice chromosome 1. Rice is represented by genes showing orthology with the sequences.

Sequences from the	Annotated orthologous	Number	E value	Highest %
STS amplicons	genes in rice genome	of gaps	L value	Identity
$I = I \cap C \cap O(1 = 0)$	LOC_Os01g09570	2	4.6e-20	82.
Lp-LOC_0s01g09570	LOC_Os05g10650	0	6.7e-05	66
Lp-LOC_Os01g11370	LOC_Os01g11370	0	6.6e-07	70
	LOC_Os01g11710	0	6.9e-11	91
	LOC_Os06g34120	0	1.6e-08	64
Lp-LOC_Os01g11710	LOC_Os11g31940	0	5.9e-06	66
	LOC_Os01g11620	0	1.3e-05	77
	LOC_Os06g34070	0	5.6e-05	77.
	LOC_Os01g13770	0	4.3e-12	78
Lp-LOC_Os01g13770	LOC_Os01g13760	0	4.8e-12	78
	LOC_Os05g15160	0	7.5e-05	65
	LOC_Os01g14550	2	1.5e-19	71
Lp-LOC_Os01g14550	LOC_Os01g14590	1	2.1e-17	74
	LOC_Os01g53090	3	1.5e-09	81
Lp-LOC_Os01g14580	LOC_Os01g14580	0	2.6e-07	71
Lp-LOC_Os01g15850	No identity found			
Lp-LOC_Os01g16100	No identity found			
Lp-LOC_Os01g16152	No identity found			
Lp-LOC_Os01g16540	LOC_Os01g16540	1	2.7e-49	70
Lp-LOC_Os01g36890	LOC_Os01g36890	2	9.3e-53	77
	LOC_Os01g36920	2	1.7e-30	85
Lp-LOC_Os01g36920	LOC_Os01g36920	2	4.3e-30	73
Lp-LOC_Os01g37480	LOC_Os01g37480	0	1.1e-31	77
Lp-LOC_Os01g52110	LOC_Os01g52110	0	2.9e-05	90
Lp-LOC_Os01g53520	No identity found			
Lp-LOC_Os01g54010	No identity found			
$I_{p-I} OC Os01g67054$	LOC_Os01g67054	0	5.8e-42	73
Lp-LOC_Os01g67054	LOC_Os15g43170	0	1.5e-33	80
Lp-LOC_Os01g67850	No identity found			
	LOC_Os01g68260	1	3.7e-24	80
Lp-LOC_Os01g68260	LOC_Os02g19150	1	6.0e-12	76
	LOC_Os01g73970	0	2.8e-05	73
	LOC_Os05g01760	0	7.9e-05	68
Lp-LOC_Os01g68324	LOC_Os01g68324	1	5.1e-20	72

#### 3.3 Discussion

The aim of this study was to fine map the genomic regions of *L. perenne* where biomass yield QTL have been previously detected. A total of 18 ryegrass specific SSR markers and five rice STS were added to LGs 2, 3 and 7 of the F2 map originally made using SSR and AFLP markers. Fine mapping is a necessary step to identify QTL positions and facilitate further work such as the search for candidate genes. The map resolution can be affected by the statistical analysis methods used and by the experimental design applied. The most straightforward method to improve the map resolution is increasing the population size (Ronin et al. 2003). A population size between 200 and 400 individuals facilitates the accurate estimation of QTL with good accuracy (Schön et al. 2004, Xu et al. 2005). In L. perenne QTL studies population sizes varied from 95 (Bert et al. 1999) to 360 genotypes (Anhalt et al. 2008). Thus the present population is the largest one used so far in L. perenne mapping analyses and is adapted for fine mapping. Another approach for increasing the map resolution would be the selection of particular genotypes as recombinant genotypes (Ronin et al. 2003). However this method can induce a bias in the detection of QTL effects (Ytournel 2008) and QTL with minor effects detected in a large population are not always detected in a small population (Melchinger 1998, Vales 2005). In consequence this approach was not appropriate to our study. The most efficient way to improve the precision of mapping is to add more molecular markers to saturate the map. Since we focused our work on three of the seven linkage groups of L. perenne and because only a few markers were already mapped on these linkage groups, this method was the most relevant. Moreover the F1 progeny from a cross between two inbred lines is heterozygous and produces a F2 population with a Mendelian segregation which is easy to analyse statistically and facilitates the mapping of markers (Doerge 2002).

The results presented show the fine mapping of LGs 2, 3 and 7 using a set of L. perenne specific SSR markers and rice STS markers. For MAS and QTL studies, a robust set of informative markers is required (Saha et al. 2006). Due to their codominance, SSR markers are more informative than AFLP or RAPD. In an F2 population with codominant markers three types of alleles can be distinguished while dominant markers would show only two. Codominant markers can distinguish heterozygote from homozygote genotypes resulting in complete information. This is not possible with dominant markers. In addition SSR markers are locus-specific and SSRs linked to biomass yield traits will facilitate L. perenne breeding using MAS in the future (Gupta et al. 2002). For this study, the choice of SSR markers was also determined by the transferability of these markers from a map to another, allowing the comparison of the results with those of other *Lolium* studies and the selection of markers mapping specifically on the regions of interest. SSR markers are widely used in many studies, but their development is time- and money- consuming and for minor crops, like L. perenne, the number of available SSRs is limited. Hence, a second strategy, based on the synteny between species, was applied. Using the synteny between *L. perenne* and rice (Sim et al. 2005, Moore et al. 2005), rice STS markers were developed. Rice has the advantage of being already fully sequenced and the synteny with perennial ryegrass has been already reported, thus it was possible to design STS specifically for our targeted region. The STS markers are highly transferable between species particularly when developed from consensus sequences. We designed rice STS from rice sequences aligned with expressed sequence homologues to two or three Poaceae species. An amplification rate of 78.5 % was acheived and confirmed the utility of STS markers. However as for the SSR, they had a poor level of polymorphism. The polymorphism degree in the F2 biomass population appears to be lower than in other L. perenne populations rendering mapping with SSR and STS markers less efficient. A low level of polymorphism of only 22% was reported for the SSR markers decreasing to 4% for the rice STS. Such results are surprising for SSR markers, widely used due to their high degree of polymorphism (Lem and Lallemand 2003, Saha et al. 2006, Jones et al. 2001). Low polymorphism is common in inbreeding populations but not in outbreeding ones. Of all published *L. perenne* maps, only Jones et al. (2002b) and Muylle et al. (2005) reported a similar problem. A lack of genetic variation between the two parental lines or a restricted use of the germplasm during the development of the two inbred lines can be an explanation (Jones et al. 2002b).

Despite the addition of new markers, the map length for the three linkage groups increased little from 289.2 cM to 313.5 cM. A comparable map of 290.4 cM was reported for these linkage groups by Gill et al. (2006) using principally SSR markers. However most of the previous maps reported are longer for this specific region. Using only EST-SSR markers, Studer et al. (2010) reported a map distance of 373.4 cM and larger distances are generated for maps with high proportions of AFLP or RFLP markers. Bert et al. (1999) reported 463.4 cM, Jones et al. (2002a) reported 386 cM whereas Barre et al. (2009) showed a map distance of 601 cM in one parent and 365 cM in the other. Different populations were used for the construction of these maps and the variation in length can result from divergence in the genetic background (Barth et al. 2001). Also AFLP and RFLP markers can enhance better genome coverage than SSR markers alone and have a role in expanding the map length (Muylle et al. 2005, Gill et al. 2006, Hearnden et al. 2007). This could be an explanation for the present map being smaller and why the addition of SSR markers did not affect the length. Moreover the SSR markers added, tended to map only within the centromeric region. Similar results

were reported in other studies in Lolium (Jones et al. 2002b, Gill et al. 2006) and for other species like barley (Ramsay et al. 2000, Wenzl et al. 2006) or by physical mapping for rye and wheat (Cuadrado and Schwarzacher 1999). These clustering phenomena around the centromeric region can be attributed to a frequent association of the SSR with repetitive elements (Ramsay et al. 1999) which are present in high numbers in centromeric regions (Heslop-Harrison 2000, Schwarzacher 2003). The distribution of the markers is also linked to the level of recombination frequencies along the chromosome. At the centromere the frequency of recombination is generally low (King et al. 2002) which can explain a cluster of SSR within this region (Ramsay et al. 2000, Gill et al. 2006). Due to the tendency of the SSR to map near the centromeric region, the genome coverage of our map remains incomplete with large gaps of around 20 cM (up to 26 cM for LG 7) at the telomeric region while the average distance between two markers is 5 cM. This lack of markers in some part of the chromosome could have a negative influence on the identification of the QTL in the future. Hence the use of different types of molecular markers will be more appropriate to improve the coverage of the genetic linkage map. AFLPs or Diversity Array Technology (DArT) (see chapter 4) markers that are supposed to be distributed more evenly over the genome could be a valuable supplementation.

The addition of markers did not cause changes in the map order. Only four inversions between closed markers were found for the three linkage groups. One exception concerns the marker rv0863 on LG3, located in this study at the upper end of the linkage group, while it was previously in the centromeric region. However this result is in accordance with the map of Gill et al. (2006). A total of 24 markers are shared between the map presented here and the map from Gill et al. (2006) giving the

opportunity to compare their positions. On LG3, there are small differences in the finescale order of loci, differences probably a consequence of the high segregation distortion present on this linkage group in both maps. There were more rearrangements on LGs 2 and 7. The rearrangements concerned principally the SSRs not influenced by segregation distortion and are the results of another biological or non-biological event. Similar populations were used and both maps were calculated using the Kosambi mapping function. In the present study JoinMap (Van Ooijen and Voorrips 2001) was used while Gill et al. (2006) employed Mapmaker (Lander et al. 1987). The Mapmaker software uses different strategies to calculate the marker order (Cheema and Dicks 2009). This variation in the ordering algorithm can induce some changes in the map order (Mollinari 2009). Moreover, the marker order is determined on the basis of recombination frequency between loci, thus an incorrect recombination value can result in an incorrect locus order. Due to the duplication of some part of the chromosomes during plant evolution (Bennetzen et al. 1998), some DNA sequences usable as markers are present twice or maybe even more. The two duplicate marker loci cannot be distinguished and a single marker locus is scored during the mapping construction (Frisch et al. 2004).

Significant deviations of the observed genotypic frequencies from their expected values were found on all linkage groups. This segregation distortion is a common phenomenon in genetic mapping and can result from statistical errors, genotyping and scoring errors or from biological incidents (Xian-Liang et al. 2006). In our study, LG3 had the highest amount of SD with only one end not distorted. For the two other linkage groups, distorted markers are more distributed along the chromosomes with even alternation of severely distorted markers with non distorted markers. Some distortions

might arise from genotyping or scoring errors considering the extreme  $\chi^2$  values obtained for some markers. Thus on LG7, rv 0134 has a  $\chi^2$  value of 202.0 and G01-010 of 121.1. Similar on LG2, G04-030 and LOC\_Os04g55260 show a  $\chi^2$  value of 242.0 and 49.7 respectively while the other markers have a value equal or inferior to 20.0. For LGs 2 and 7, SD favoured no genotypes specifically and perhaps SD at some loci was not linked to a biological event. Different segregation patterns were observable indicating that a biological reason is unlikely. Although the surrounding markers also show SD, the distance between markers, particularly on LG7, is high and only by adding more markers will it be possible to determine whether the SD was locus specific or due to a region subject to segregation distortion. When the SD concerns a complete region, the hypothesis of genotyping or scoring error is disproved (Xian-Liang et al. 2006). But for most of the distorted markers, the occurrence of a biological event seems the principal cause. All severely distorted markers were tested a second time on the entire population and similar results were obtained. On LG3 a high  $\chi^2$  value up to 60 was detected but all the markers were concerned and most markers on the other linkage groups do not have extreme values. SD has been frequently detected in L. perenne mapping populations. There is some concordance between the regions of SD detected in our study and the SD reported in previous maps. The most notable is the consistency of SD on LG3 across different maps (Bert et al. 1999, Armstead et al. 2002, Jones et al. 2002a, Faville et al. 2004, Gill et al. 2006, Anhalt et al. 2008). The SD phenomenon is poorly understood and could be an effect of a single or a combination of different mechanisms (Anhalt et al. 2008). In plants, the percentage, origin or genetic effect of the SD vary in relation to the species, the population type, specific crosses and molecular markers (Xian-Liang et al. 2006). A possible association of significant SD with self-incompatibility genes has been reported (Thorogood et al. 2002 and 2005).

In ryegrass, the two major S incompatibility loci have been reported on LG 1 and LG 2 and a particular locus linked to self-incompatibility on LG 3. Thus for LGs 2 and 3, the SD might be a consequence of self-incompatibility genes. However no incompatibility locus has been reported for LG7. Gill et al. (2006) observed SD on LG7 and explained it by the presence of lethal or semi-lethal loci remaining after generations of inbreeding.

Comparative mapping work has demonstrated that regions of the L. perenne genome show a potentially useful degree of conservation with rice (Jones et al. 2002a, Sim et al. 2005) and this can be used for the development of molecular markers or to pinpoint genes underlying QTL. Rice STS markers were designed to map specifically onto LGs 2, 3 and 7 of L. perenne. Five markers gave polymorphic amplification products and mapped to the expected linkage group in accordance with already reported results (Sim et al. 2005, Jones et al. 2002a). The other developed STS markers gave a high amplification rate of 78.5% confirming the possibility of increasing the number of molecular markers for a targeted region without the necessity to develop markers from perennial ryegrass. On LG2, the marker order is not conserved and the high level of SD could be an explanation. But different studies have shown disruption in the colinearity resulting from small rearrangements such as gene insertions or deletions, gene inversions, duplications or translocations (Bennetzen 2000, Alm et al. 2003). The lack of data available in our study does not allow confirmation of these conclusions. The amplicons of non-polymorphic rice STS primers were sequenced to get more information. The sequences were then blasted against the rice genome. They revealed a high identity with the expected orthologous genes from rice but half of them had gaps in the alignment indicating the presence of small rearrangements between the species without effect for the development of markers. However, this could have a limiting

effect on the identification of genes. Half of the sequences are also orthologous to one or more other genes. In general these genes were on the same chromosome and close to each other, sometimes even overlapping which explains why the sequences from L. perenne could cover more than one gene. However, identity was also found to genes on other rice chromosomes. This suggests some duplication events on some part of the rice genome during evolution. The duplication can be also present in L. perenne but in the absence of data it can be just deduced that the synteny approach has to be used carefully as already shown in previous studies in grass species (Dubcovsky et al. 2001, Sorrels et al. 2003). In rice, the trait grain yield is more important, biomass yield being just one of its components. But a number of different QTL have been associated with biomass yield in double haploid populations on rice chromosomes 1, 3 and 5 (Zhang et al. 2004), and on rice chromosomes 1, 4, 7 and 8 (Liu et al. 2006). Thus homoelogous rice chromosomes of LGs 2, 3 and 7 of L. perenne also present QTL for biomass yield. Data from rice could be useful to identify candidate genes for biomass yield in Lolium. However this would be only possible if more information were available for this trait in rice and with the conservation at the microstructure level within this region between the two species. Although rice worked so far well as a model plant for grasses, for forage grasses with biomass yield as major plant breeding trait, rice is not the ideal model plant. Moreover unlike L. perenne rice is not a crop of temperate areas. Brachypodium distachyon, with only 5 pairs of chromosomes, has a small genome of 355 Mbp fully sequenced since 2010 (International **Brachypodium** Initiative, http://www.brachypodium.org/node/8). This species is proposed as a more efficient model plant not only for forage grasses but for all grasses. Few markers were used but the results indicate synteny between Brachypodium chromosome 5 and L. perenne LG2, Brachypodium chromosome 4 to LG5 and Brachypodium chromosome 3 to LG6 (Hasterok et al. 2006). The LG1 seems to share syntenic regions with *Brachypodium* chromosomes 1, 5 and 2 (Hasterok et al. 2006, Garvin et al. 2008). No information is available yet about the synteny between *Brachypodium* and LGs 3 and 7 of *L. perenne*.

In this chapter, the addition of molecular markers to a restricted region of three linkage groups for fine mapping was described. The aim was to develop a map with a greater degree of locus saturation to facilitate the location of QTL. The mean distance between two consecutive loci is 5 cM which is an improvement compared to previous map distances but is still too large for QTL detection with the possibility to further search for candidate genes. Although the mapping of SSR markers will provide more information for the QTL calculation than dominant markers, the SSRs showed a limitation in the genome coverage and some gaps of around 20 cM remained. Overall more markers and more markers of different types need to be added on LGs 2, 3 and 7 of perennial ryegrass. In this chapter our results concerning the synteny between rice and *L. perenne* were consistent with published data and the study of the synteny need to be extended to species more closely related to *L. perenne* than rice, as for example *B. distachyion*.

# **Chapter 4: Fine mapping with DArT markers**

#### 4.1 Introduction

Saturation of a genetic linkage map is the first prerequisite to study agronomic traits aimed at marker assisted selection in breeding programs and/or the isolation of genes linked to these traits. The challenge consists of mapping a sufficient number of markers to cover the entire genome. Most published maps are based on RFLP, AFLP or SSR markers (Bert el al. 1999, Armstead et al. 2002, Jones et al. 2002a and b, Faville et al. 2004, Gill et al. 2006, Anhalt et al. 2008). However these markers are expensive and time consuming and hence not suitable for high throughput genotyping at the scale required in even a moderately size breeding program. SSR markers development is also dependant on DNA sequences and thus possible only for fully or partially sequenced species. To overcome these disadvantages, novel marker systems based on microarrays were developed. These markers and Diversity Array Technology (DArT) markers (Gupta et al. 2008). They will be particularly powerful for fine mapping of species with little publicly available genomic resources.

#### 4.1.1 Principle of Diversity Array Technology (DArT) and its application.

Diversity Array Technology is a microarray hybridization based technique detecting the presence vs. absence of individual fragments in genomic representations and allowing a simultaneous screening of thousands of anonymous polymorphic loci (Jaccoud et al. 2001, Wenzl et al. 2004). The polymorphism is generally detected from single base-pair (SNPs) changes or from insertions/deletions/rearrangements at restriction sites. The DArT procedure consists of several steps: (1) the digestion of the

DNA studied (2) the construction of a library (3) the microarraying of the library onto glass slides (4) the hybridization of the DNA previously labelled on the microarray and finally (5) the scanning of the slides to analyse the presence of hybridization (Diversity Arrays Technology Pty Ltd <u>http://www.diversityarrays.com/</u>). The DArT method offers several advantages: DArT markers are reproducible, require a small amount of DNA (around 50 to 100 ng) and are cost effective. A large number of markers is provided, allowing the construction of highly saturated maps within a short time and with large genome coverage. And due to the independence of DNA sequences, DArT is applicable to all species especially those with little genomic sequence resources such as *L. perenne* (Jaccoud et al. 2001, Gupta et al. 2008).

The application of the DArT marker system varies from construction of highdensity genetic linkage maps (Peleg et al. 2008, Hearnden et al. 2007) to identification of QTL (Parth et al. 2008, Grewal et al. 2008), genome profiling (Wenzl et al. 2004) or estimation of genetic diversity (Jaccoud et al. 2001). After an initial development in rice in 2001 (Jaccoud et al. 2001), DArT technology was also developed for barley (Wenzl et al. 2004). It was subsequently applied to the model species *Arabidopsis thaliana* (Wittenberg et al. 2005) and to agronomically important species such as wheat (Semagn et al. 2006), sorghum (Parth et al. 2008) and cassava (Xia el al. 2005) among other species. Recently a DArT array was developed within the *Festuca-Lolium* complex to estimate intra- and inter- specific genetic diversity (Kopecký et al. 2009).

#### 4.1.2 Mapping of DArT

A large amount of markers are generated with a DArT array and the construction of an accurate genetic map is not straightforward. The main difficulties concern the map order of so many markers and the dominant nature of the markers. The number of possible orders is equivalent to n!/2 with n the number of markers. When n is getting large the problem is to select the best order among a wide range of possibilities (Mester et al. 2003, Mollinari et al. 2009). Also for dominant markers two linkage phases exist: repulsion phase when two dominant alleles reside on different chromosomes and coupling phase when they are on the same. These two phases cannot be distinguished and complicate the estimation of recombination which could affect the ordering of the markers (Mester et al. 2003, Tan and Fu 2007). The Kosambi and Haldane functions are not adapted to deal with a high amount of dominant markers, thus several specific algorithms have been developed. The most common and powerful is the maximum likelihood (ML) approach based on expectation maximization algorithms (Dempster et al. 1977). Initially the calculation was based on a two-point analysis considering two markers at a time then on a three-point analysis (Tan and Fu 2007). The ideal would be multi-point analysis that could consider the whole group of markers for each marker data. However for a large set of data, this method takes time and Jansen (2009) proposed simplifying it by minimizing the number of recombinations between markers using hidden inheritance vectors. Although most markers are in general in coupling phase two groups can be distinguished in function of the parent line from where the markers are segregated from (Tan and Fu 2007). The map calculation can be done on two separate maps before combining them. One represents the markers derived from the paternal line and the second the markers derived from the maternal line (Tan and Fu 2007). For maps consisting of dominant and codominant markers, the codominant markers can be present in both sets and used to establish the junction between the two separate maps.

#### 4.1.3 Objectives

The work presented in the previous chapter showed the difficulty of achieving high resolution mapping even when focusing on a restricted region. The integration of 23 markers to the 38 markers already mapped did not provide a sufficient genome coverage with several gaps remaining and a distance of 5 cM between markers. Different molecular markers (SSR, STS and AFLP) were used and it was possible to distinguish the tendency of each type of marker to map on some region of the chromosomes. AFLP generally mapped at the telomeres while SSR and STS markers clustered around the centromere. The inclusion of another marker type will probably enhance the genome coverage. In addition the F2 biomass population is little polymorphic. Markers detecting SNPs like DArT are powerful enough to detect more polymorphism. Work presented in this chapter describes the saturation of LGs 2, 3 and 7 using DArT markers. The aim was to fill large gaps and generate a fine map to allow a precise positioning of biomass yield QTL.

# 4.2 Results

#### 4.2.1 Construction of the linkage map

The construction of the linkage map is based on a combination of DArT markers, an EST marker (EST1) mapped for a study on rust and the markers previously mapped (Anhalt et al. 2008 and Chapter 3 of this thesis). Data for 1205 DArT markers were provided by Diversity Arrays Technology Pty Ltd (Australia) after hybridization of the parents and the 360 F2 genotypes on the DarTFest array (Kopecky et al. 2009). Only the 675 markers with a quality criteria score above 90% were selected. For some genotypes the hybridization did not work and several genotypes had a lot of missing values. In order to limit the number of missing data only 325 genotypes were retained

for the mapping calculation. In the first place the two sets of markers (a/c + a/h/b andb/d + a/h/b) were analyzed separately using the ML algorithm from JoinMap 4.0. When doing the grouping, the markers were distributed within only six linkage groups at a high LOD of 25 indicating the difficulties of assigning some markers to a linkage group. JoinMap automatically excluded half of the AFLP markers and the others were assigned to different groups than previously. The AFLP were removed from the data sets for the subsequent rounds of calculation. As for the previous map, many codominant markers showed considerable SD with high  $\gamma^2$  values but none of the co-localizing DArT markers were skewed. Markers with a  $\chi^2$  value above 25 were removed which led to a reduction of map inflation and stability in the marker order. With the highly skewed markers the addition or deletion of a marker results in important changes in the map order. During the grouping with the two simplified subsets of markers, the markers fell into seven linkage groups at a low LOD of 5 maximum. The seven linkage groups were identified by the codominant markers. Once the markers were assigned to each of the seven linkage groups, the longest map estimated from the markers segregating from the maternal line was selected as fixed order when the two sets of marker were combined. For closely linked markers the confidence in the marker order was low and markers having a high level of similarity were excluded. Re-calculations were performed until the map displayed a high confidence in marker order with redundant information removed. Finally this brought a total of 297 DArT with 29 co-dominant markers mapped on the entire genome (Figure 8, Tomaszewski et al. 2012); and 18 SSR, 1 STS and 134 DArT markers mapped on LGs 2, 3 and 7 (Figure 9).

LG1	LG2	LG3	LG4	LG5	LG6	LG7
0.0 1.6 2.7 3.7 1.6 1.6 1.6 1.6 1.6 1.6 1.6 1.6	0.0         LeP136512           1.7         LeP1355783           1.9         LeP1558025           1.4         LeP1558025           1.9         LeP1558025           1.4         LeP1558025           1.9         LeP1556319           2.6         LeP1556733           2.4         LeP1556733           2.5         LeP1556733           2.6         LeP1556739           3.3         LeP1556739           3.4         LeP1556739           3.5         LeP1556739           3.6         LeP1556739           3.7         LeP1556739           3.8         LeP1556739           3.4         LeP1556739           3.5         LeP1556739           3.6         LeP1556739           3.7         LeP155740           LeP1558397         LeP155739           2.8         LeP15557116           3.7         LeP15557116           13.7         LeP1557116           13.7         LeP1557116           13.7         LeP1561324           152.6         LeP1561324           152.6         LeP1561324           152.6         LeP1561324	0.0         LoPt566224           11.3         LoPt566224           11.3         LoPt56700           16.1         LoPt561085           23.1         LoPt567209           24.7         LoPt567529           32.6         LoPt567529           33.9         LoPt56760           1.04755629         LoPt56763           3.9         LoPt56760           1.04755729         LoPt56763           3.9         LoPt56760           1.04755729         LoPt56760           5.8         LoPt56760           1.04755667         LoPt56289           3.9         LoPt55736           3.7.5         LoPt55736           3.8         LoPt557475           3.9         LoPt555339           3.1.9         LoPt555339           3.1.9         LoPt555776           3.1.9         LoPt55	0.0 NFFa142 13.6 DLF025 27.5 JL5 14.2 JLP1355956 LP155123 14.2 LP1355956 LP155123 14.5 LP1355956 LP155123 14.5 LP155123 14.5 LP155123 14.5 LP155123 14.5 LP155123 14.5 LP155123 14.5 LP155123 14.5 LP155123 14.5 LP155123 14.5 LP155123 14.5 LP155225 14.5 LP155225 14.5 LP155225 14.5 LP1552215 14.5 LP155225 15.5 LP15525 15.5 LP	0.0 7.0 8.1 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1	0.0 2.0 LoPt56708 4.3 LoPt55708 1.0Pt55708 1.0Pt55708 1.0Pt555315 1.0Pt555315 1.0Pt555315 1.0Pt555315 1.0Pt555315 1.0Pt555355 2.7, LoPt556784 3.1 LoPt56784 3.1 LoPt56784 3.1 LoPt56784 3.1 LoPt56778 3.1 LoPt5678 3.1 Lo	0,0 15,8 16,5 10,5 1

Figure 8: Mapping of DArT markers onto the genetic map developed by Anhalt et al. (2008) to generate a high resolution genetic map of the entire genome of *L. perenne*.


LG3



Figure 9: High resolution genetic map of the LGs 2, 3 and 7 of *L. perenne* including 18 SSR and 1 STS markers (in red) and 134 DArT markers (in black). The map was generated using the ML mapping function in JoinMap 4.0. Loci labelled with *asterisks* indicates the segregation distortion (\*P<0.1, \*\*P<0.05, \*\*\*P<0.01, \*\*\*\*P<0.005, \*\*\*\*\*P<0.001, \*\*\*\*\*P<0.0005, \*\*\*\*\*P<0.0001)

## 4.2.2 Characteristics of the map

The composite map spanned a length of 1052.6 cM and a density of 1.54 markers per cM. For the region of interest the length of the map was 544.4 cM which represents a 74% increment compared with the map without the DArT markers (313.5 cM). In spite of the inflation, the marker density had increased with an average of one marker per 3.5 cM (5 cM previously). Not many SSR and STS markers were conserved but they mapped more along the chromosomes with some being at the telomeric regions. However they tended to cluster together in groups of two or three. In general the DArT markers were evenly distributed and able to fill the gaps on LG3 and to reduce one on LG7 and one on LG2. But some regions remained poorly covered with large gaps between 10 to 20 cM and even a gap up to 30 cM on LG 2 remaining. As mentioned before, this phenomenon concerned principally the telomeric region but for LGs 2 and 7 it extended all along the long arm of the chromosome. The gaps were also more frequent around the codominant markers.

In accordance with the previous results, SD was detected on LGs 2 and 3, but not anymore on LG7 where with the exception of rv0459 none of the markers were distorted. SD was not influenced by marker type and the DArT markers were as distorted as the other markers. LG3 is still characterized by a high percentage of skewed markers (62%) and only the telomeric regions remain not distorted. The level of significance on this chromosome was important and most markers present a distortion at P<0.0005 and P<0.0001. A lower level of SD (35%) was detected on LG2 and concerned markers distributed randomly along the chromosome although markers around the centromere tend to be more distorted.

The DArTFest array has been used in the creation of a *L. multiflorum* genetic map (Bartoš et al. 2011) making the alignment with the *L. perenne* map possible (Figure

10). Of the positively scored probes only 6 % were mapped to *L. multiflorum* linkage groups 2, 3 and 7 thus few markers are in common between the two maps. In general there is conservation of the marker order which tends to confirm the robustness of the map. But some markers mapped at different position, and the DArT markers LoPT 556224 and LoPt 561957 mapped on *L. multiflorum* linkage groups 7 and 6 respectively.



Figure 10: Alignment of LGs 2, 3 and 7 of the *L. perenne* DArT genetic map to linkage groups Lm1, Lm2 and Lm7 of the *L. multiflorum* genetic map (Bartoš et al. 2011). The dashed line indicates the conservation of the co-dominant marker M15185 and the DarT markers in red were mapped on another *L. multiflorum* linkage group.

## 4.3 Discussion

The screening of the studied population with the DArTFest array was appropriate for fine mapping as proven by the large number of markers it provided. A total of 134 DArT were added to LGs 2, 3 and 7 but some of the markers already mapped had to be removed and only one rice STS and 18 SSR were maintained. While the main difficulty encountered before was to identify sufficient markers to cover the entire genome, with the DArT markers the difficulty was the large amount of data generated. Furthermore the DArT markers are dominant which complicates the calculation of the genetic map. Although the construction of high-density linkage maps with dominant markers is increasingly common; increasing marker numbers makes heavy computational demands. Algorithms including multipoint Maximum Likelihood have been developed for such applications. The accuracy of ML mapping has been demonstrated (Hackett and Broadfoot 2003, Tan and Fu 2007, Jansen 2009, Keyser et al. 2010) and the algorithm seemed the most relevant for our study. But errors in the map order may arise due to missing values, segregation distortion and more importantly genotyping errors (Hackett and Broadfoot 2003, Kearsey and Farquhar 1998). The impact of these factors tends to increase when the distance between markers is reduced (Cheema and Dicks 2009) and a single percentage of typing errors can result in an incorrect map order (Hackett and Broadfoot 2003). Genotypes without data were removed to avoid any errors linked to missing values. However some missing values remained particularly for the SSR and STS markers that never gave data for the entire population. Also after the first round of ML mapping was performed on the separate subsets, some markers showed extreme levels of SD with a  $\chi^2$  value up to 65, while the surrounding markers did not have such a level of SD. Similar results were already observed on the map without the DArT markers and the possibility of genotyping errors was suggested. Markers with important incoherence in the segregation ratio, as only 20 out of 325 genotypes segregating from one parent, were removed. A low level of recombination was observed between some DArT markers which hampers the positioning on the map. Highly similar DArT markers were removed but some inversions between DarT markers mapping close to each other can not be excluded and have to be taken into account for the OTL positioning. The number of SSR and STS markers in common between this map and the previous one is too small for an accurate comparison of marker order particularly on LG7 where only three SSR remained. However for LG3, with the exception of one inversion, the order stayed the same indicating a good integration of the DArT results to the map. There were more rearrangements on LG2, perhaps a consequence of SD with alternation of skewed markers with non-distorted markers. The SD has less effect on LG3 where all markers from the same region were skewed. A comparison to the L. multiflorum map generated with the same DArTFest array reveals a relatively identical marker order but some major rearrangements were also noticed (Figure 10). Due to a lack of a larger number of common markers these results are only an indication to confirm the robustness of the map, however the presence of errors cannot be completely excluded.

The current map spanned a length of 544.4 cM (313.5 cM previously). The longest map published so far in *L. perenne* for the restricted region LGs 2, 3 and 7 was based on 58 dominant and codominant markers and had a length of 602.2 cM (Barre et al. 2009). The map with the highest amount of markers was based on 191 AFLP markers and covered a genetic distance of 569.7 cM (Bert et al. 1999). The addition of a large number of markers explains the inflation. But with a large set of markers the probability of typing errors increases. It was demonstrated previously that scoring errors

affect the marker order and also inflate the map length (Kearsey and Farquhar 1998, Hackett and Broadfoot 2003). The map generated with the entire positively scored DArT was twice as long as the final map. This could in part explain why despite a great genome coverage some large gaps remained. A physical map will be useful to evaluate the real distance between the markers.

The distribution of the markers is homogenous along the chromosomes although the SSR moderately cluster in small groups but they are better distributed than on the previous map. In general the DarT markers enabled a good genome coverage and three of the six telomeric large gaps were filled. Gaps remained at both telomeres of LG2 and at one end of LG7 but were reduced indicating that the gaps on the previous map were associated to some limitation of the markers used and not only specific to our population. For most of the published maps in L. perenne, the telomeres are generally less covered. The DArT markers seem to partially circumvent the problem. With the previous map, we noticed the tendency for the AFLP markers to map at the telomere. Simultaneously mapping the DarT markers and AFLPs would have been interesting to see their capacity to cover this specific region. Unfortunately, probably due to some genotyping errors, JoinMap could not assign the AFLPs to any linkage groups after addition of DArT markers. But gaps were not restricted to the chromosome ends and large gaps were also flanking all the codominant markers. Gaps on the map can be also explained by a large number of dominant markers being added to a genetic map based on a F2 population. The ML mapping algorithm is adapted to the construction of mixed maps but dominant and codominant markers have a different type of segregation that could affect the calculation of recombination frequencies.

The map presented in the previous chapter was characterized by an omnipresent level of segregation distortion. The addition of DArT markers confirmed SD on LG3 with 62% of the markers significantly distorted and only the markers at the telomeres being not affected. Similar segregation was highlighted in different L. perenne maps (Bert et al. 1999, Armstead et al. 2002, Jones et al. 2002a and b, Faville et al. 2004, Gill et al. 2006, Anhalt et al. 2008) and was associated to the gene for self-incompatibility (Thorogood et al. 2002). For the same reason, LG2 generally shows SD. However for this linkage group, the addition of dominant markers showed clearly that the distortion was unilateral, with only markers segregating from the paternal line distorted. SD can be a consequence of deletion and duplication events or inversion and duplication events occurring at the inter- and intra-chromosomal level or at the intrachromatid level. This could explain why alleles from one parental line are favoured to the alleles of the other parental line (Anhalt et al. 2008). Concerning LG7, all the skewed markers of the previous map were removed due to suspicious  $\chi^2$  values and to the absence of SD within the DArT markers. No SD remained for this linkage group in accordance with results reported for L. perenne (Faville et al. 2004, Muylle et al. 2005). It confirms the presence of some genotyping errors and that the strategy of removing these markers was adapted to generate a robust map.

In this chapter the first use of DArT markers for mapping in *L. perenne* was described. The markers were selected to overcome the lack of polymorphism of the population and for the large number of markers provided. From a set of 675 DArT markers only 297 were mapped on the seven linkage groups with 134 of them mapping on LGs 2, 3 and 7. The objective of this chapter was to provide a greater genome coverage to define later, precise biomass QTL positions. With an average of 3.5 cM

between markers, the map density is not as high as expected. By keeping all the DArT markers a coverage of one marker every 1 or 2 cM was possible but with a too high risk of errors in the marker order. However the density of the map generated will be sufficient enough to get a precise indication of the biomass QTL positions.

# **Chapter 5: Fine mapping of biomass QTL**

## 5.1 Introduction

To date improving biomass yield is a major objective in *L. perenne* breeding programs to increase the profitability for the farmers. But due to the complex nature of this trait, it is difficult to determine a relation between the trait biomass and the genetic background. Biomass yield is a quantitative trait controlled by multiple genes with most likely several genomic regions involved and influenced by environmental interactions (Yamada et al. 2005). Therefore the study of this trait relies on Quantitative Trait Loci (QTL) mapping. Several studies on QTL mapping have been carried out in a range of plant species, but to date no QTL for biomass yield has been successfully cloned.

#### 5.1.1 Methods for QTL detection

QTL mapping consists of the detection in a population of a possible association between phenotypes and the genotypes of molecular markers. The markers are used to split the population in three classes according to the marker genotype. The difference between groups is evaluated by mean and variance analyses and a significant difference suggests a link between the marker loci and a QTL controlling the trait (Young 1996, Collard et al. 2005). Dense genetic maps have been generated in plant species with hundreds of markers. Different strategies have been suggested for mapping QTL with large sets of genotypic data and nowadays the analyses are performed using specific statistical tests available in software packages dedicated to QTL analysis. The easiest way to perform a QTL analysis is by using the statistical test of variance analysis at marker loci. However this method is not applicable if data is missing since it underestimates the phenotypic effects, and does not estimate the QTL location and the QTL effect (Broman 2001, Lander and Botstein 1989). To overcome these problems, Lander and Botstein (1989) developed the Interval Mapping (IM) test. This approach detects putative QTL within a region flanked by two markers via the elaboration of a LOD profile against the marker positions. The LOD score corresponds to the assumption that there is a QTL divided by the null hypothesis (no QTL) (Kearsey and Farquhar 1998, Doerge 2002). A OTL is declared present when the LOD score exceeds a predetermined threshold value (Lander and Botstein 1989). The IM method provides a better estimation of the QTL effects and allows the assumption of QTL between two markers. In addition, this approach is reliable in the presence of missing marker genotype data (Broman 2001). But this test is based on the hypothesis that there is only one QTL on the chromosome and does not allow the consideration of interaction between multiple QTL (Doerge 2002). This might lead to the non-detection of some QTL. Therefore multiple QTL models were developed. The most common models are the Composite Interval Mapping test (CIM (Zeng 1993)) or Multiple QTL Model mapping test (MQM (Jansen 1993)). CIM is a combination of IM with a linear regression using a pair of markers to locate the tested position and the other markers to control the background. MQM is an extension of IM with multiple linear regressions using one or more cofactors to eliminate the effects of possible QTL in other intervals. The entire genome is considered simultaneously which allows the detection of multiple QTL even within the same chromosome and the differentiation between linked and/or interacting QTL. For both algorithms, errors might arise from the markers selected as covariates (Broman 2001)

## 5.1.2 Factors to consider for QTL mapping

In addition to the statistical methods, the accuracy of QTL mapping depends on the experimental design, the size of the population, the heritability of the trait, the number and distance between markers and the genetic properties of the QTL (Asins 2002). The initial step for QTL mapping consists of the cross between two parental lines differing for the quantitative trait studied. The most common populations used are F2 and backcross populations that are easy to produce within a short time. The major advantage of the F2 over the backcross population is the possibility to estimate the presence of dominant effects while for backcross, additive and dominant effects are not differentiated, and some QTL might not be detected due to similarity between the genetic values of the homozygote and the heterozygote individuals (Carbonell et al.1993). An F2 deriving from two inbred lines appears more powerful for the detection of QTL because it increases the linkage disequilibrium between the QTL and the genetic markers used (Erickson et al. 2004). Furthermore the individuals are more informative for QTL segregation (Mackay 2001). Additionally it is recommended to use a large population size with 300 to 500 individuals to increase the chance of detecting the genes with small effects (Tanksley 1993, Erickson et al. 2004). In a small population, two closely neighboured QTL will appear as one single QTL (Tanksley 1993). For the construction of the genetic map, the type and the quantity of markers are determinant for the QTL detection. Codominant markers have the advantage of being more informative than dominant markers. In consequence, the genotype of a putative QTL flanked by dominant markers will depend on the genotype or phenotype of the next codominant markers (Jiang and Zeng 1997, Xie and Xu 1999). However to conduct a robust QTL analysis a sufficient map density with a distance between markers around 10-15 cM is required (Erickson et al. 2004). The presence of gaps within the linkage

map can result in the non-detection of some QTL. When a high precision of the QTL position is needed for further work, the map density has to be increased. Compared to the codominant markers, the dominant markers such as DArT or AFLP generally provide a large amount of data and are distributed over the genome permitting a better saturation of the map and a great genome coverage.

The detection of QTL also depends on the genetic properties of the QTL principally the number and effect of each QTL. Often QTL with small phenotypic effects are not detected because they are under the level of significance, and QTL closely linked may be detected as a single QTL. In addition the effects of the genes controlling the quantitative traits are sensitive to external environment and the genotype x environment (G x E) interaction has to be taken into account in QTL mapping. Only a subset of QTL is common to all environments. For *L. perenne* presents in different ecoclimatic regions it is important to understand the influence of the environment. The detection of this effect requires the collection of phenotypic data using an experimental design with replications in various environments for each genotype (Mackay 2001, Asins 2002).

## 5.1.3 Context and Objectives

In *L. perenne* several QTL studies have been reported for large range of traits such as heading date (Armstead et al. 2004, Yamada et al. 2004, King et al. 2008, Byrne et al. 2009) or crown rust resistance (Muylle et al. 2005). A summary of all the QTL analyses performed in *L. perenne* is given in Table 10. Traits related to biomass yield were analysed. Thus a QTL for leaf length was detected on LG5 (Yamada et al. 2004), on LG7 (Armstead et al. 2008) and on LG4 (Barre et al. 2009). In parallel, QTL for leaf width were detected on LG3 (Yamada et al. 2004), on LG7 (Armstead et al. 2008) and

on LGs 3 and 4 (Anhalt et al. 2009) and for the leaf extension rate on LGs 2 and 3 (Turner et al. 2008).

Traits	Population type	Analysis method	Authors
Heading date	F2	IM, MQM	Armstead et al. 2004
Morphological development and winter hardiness-associated traits.	DH*F1	IM, CIM	Yamada et al. 2004
Herbage quality traits	DH*F1	IM, CIM	Cogan et al. 2005
Crown rust resistance		IM, MQM	Muylle et al. 2005
Vernalization response	F2	IM, MQM	Jensen et al. 2005
Water soluble carbohydrates	F2	IM, Kruskal- Wallis, MQM	Turner et al. 2006
Components controlling fertility	DH*F1 F2	IM, MQM, CIM	Armstead et al. 2008
Heading date	BC <sub>2</sub>	IM, MQM	King et al. 2008
Fructan and growth	F2	IM, Kruskal- Wallis, MQM	Turner et al. 2008
Heading date, spike length and spikelets per spike	F1	IM, MQM	Byrne et al. 2009
Biomass yield	F2	IM, MQM	Anhalt et al. 2009
Leaf length	F1	CIM	Barre et al. 2009
Morphological traits influencing waterlogging tolerance	F1	IM, CIM	Pearson et al. 2010

Table 10: Review of Quantitative trait loci analyses reported in *Lolium perenne*.

But there were few QTL analyses of the trait biomass yield itself. The first QTL for fresh weight was described by Yamada et al (2004) in a population generated for a pair-cross between a multiple heterozygous plant and a doubled haploid plant where a QTL was found on LG5. An autumn dry matter QTL was reported on LG3 by Turner et al. (2008). Furthermore a biomass QTL analysis was performed on a F2 inbred line derived population and the three traits fresh weight, dry weight and dry matter were measured over a two years experiment in the field and in the greenhouse (Anhalt et al.

2009). The traits fresh weight and dry weight showed consistent positions on LGs 2, 3 and 7 in both environments with one additional QTL for both traits on LG5 in the field. The explained variance for fresh weight was up to 28% in the greenhouse and 50% in the field and 29% and 53%, respectively, for dry weight. The trait dry matter had different positions in the greenhouse (LG3) and in the field (LG4) and results varied also in function of the statistic test applied (IM or MQM). The trait leaf width was also measured in the greenhouse and QTL were identified on LGs 3 and 4 accounting for 15% of the phenotypic variation. The map used was not saturated enough to define precisely the QTL position, to see if what appeared as a single QTL was in fact two or more QTL and to identify molecular markers linked to biomass traits. The flanking markers were generally far apart from the QTL intervals. Knowing the number of genes present within a few centimorgans, the markers could be linked to other traits resulting in the selection of undesirable traits when used for MAS.

The work presented in the two previous chapters consisted of the fine mapping of this map using different types of molecular markers to achieve a high genome coverage. Using the new map generated and the phenotypic data previously collected over a two years experiment (Anhalt et al. 2009), the objective of this chapter was to recalculate the position of the biomass yield QTL, to confirm the QTL positions by using different markers and to reduce the QTL intervals. A small QTL interval will facilitate future work of identification and description of *L. perenne* genomic sequences underlying biomass QTL. The results will be also useful to determine molecular markers linked to biomass yield to facilitate selection in early stages of plant breeding programs in the long term.

## 5.2 Results

QTL analysis was carried out using the two maps presented in chapters 3 and 4 with the phenotypic data collected in a greenhouse and field experiment between 2005 and 2007 by Anhalt et al. (2009). Anhalt et al. (2009) showed that all traits were normally distributed after a log transformation and no significant variation between blocks was detected. Generally the traits were significantly correlated. The variation of data is higher for the field than for the greenhouse probably due to genotype x environment interactions: two years were characterized by 80 mm of rainfall in May followed by 120 mm of rainfall from June to August while for the third year May was dry (40 mm) but the three months of summer particularly wet with 350 mm of rainfall.

Interval mapping was initially applied to detect the QTL positions and to identify some potential cofactors for the MQM mapping test. In parallel CIM was performed to confirm the positions of the QTL. For CIM, the cofactors were automatically selected which reduces potential mistakes in the cofactor selection. QTL were calculated and displayed by LOD scores and percentage of explained variance of the QTL.

## 5.2.1 QTL positions after addition of SSR and STS markers

QTL identified with the less saturated map are shown in Figure 11 and Figure 12. QTL for the traits fresh weight and dry weight were detected on LGs 2, 3 and 7 in accordance with the results previously reported by Anhalt et al. (2009). Generally the two traits were located at the same chromosomal regions and a large proportion of QTL locations identified by IM were also detected by CIM while MQM mapping detected less QTL. Most of the QTL positions were similar in both environments. In the field a total of 14 QTL was detected with IM and CIM and in the greenhouse CIM revealed

three QTL less than IM (18 in total for IM and 15 for CIM). These QTL were detected but did not reach the significant LOD score value. The CIM method did not allow a more precise positioning of the QTL. Looking at the LOD profile generated, MQM mapping gave sharper and clearly distinguishable peaks compared to the other methods (annex III). For MQM, the markers shown in Table 11 were identified as cofactors. The total number of QTL was narrowed down to 8 with one QTL on LG2, two on LG3 and one on LG7 for each trait. But an important difference was found out between the greenhouse and the field experiments. On LG2 the QTL was close to one telomere for the field and close to the opposite telomere for the greenhouse experiment.

	Traits	LG2	LG3	LG7	LG4	
Field	Fresh weight	M15185	rv1133	rv1411	-	
		LpHCA17C11	rv1316			
	Drv weight	M15185	rv1133	rv1411	_	
		LpHCA17C11	rv1316			
	Dry matter	M15185		ry1175	_	
	Dry matter	W113185	LOC_0301g30890	1111/5	-	
Greenhouse	Fresh weight	LOC_Os04g55260	rv1133	rv1411	-	
			LOC_Os01g16152			
	Dry weight	LOC 0s04g55260	rv1133	rv1411	_	
	Dry weight	100_030+g35200	LOC_Os01g16152	1 1 1 1 1 1		
	Dry matter	-	EacaMctc-100	-	-	
	Leaf width	rv0959	LOC_Os01g16152 G04-054	-	rv0380	

Table 11: List of the markers identified as cofactors for MQM mapping for the map with the SSR and STS markers.

Small differences in the LOD values and percentage of explained variance were observed between the two traits and between IM and CIM mapping. The major QTL was detected on LG3 (LOD score CIM: 14.44 and % explained variance CIM: 20.5, Table 12). However with MQM, the QTL with the largest effect was found on LG2 with a LOD score of 9.22 and explaining only 8.4% of the phenotypic variance. The LOD

score values and percentage of explained variance between the two experiments were quite equivalent except for LG3 where the values were lower for the greenhouse experiment with a maximum LOD score of 8.32 in the greenhouse while it was 13.61 in the field.

Less QTL were identified for the trait dry matter and the number and position varied. However one QTL on LG2, one on LG7 and the QTL on LG3 were consistently detected in the field. But in the greenhouse data set, dry matter QTL were only localised on LG3. For both environments, Anhalt et al. (2009) identified a dry matter QTL only on LG3. After addition of markers, this QTL showed the highest LOD score (LOD score CIM: 7.26) and explained up to 18.6 % of the phenotypic variation.

In the greenhouse, an additional trait, leaf width, was measured. A single QTL was detected on LGs 2 and 4. On LG3 the number of QTL varied between the methods with four QTL for IM, two QTL for CIM and only one for MQM. Leaf width QTLs were already found by Anhalt et al. (2009) for these three LGs.



Figure 11: Location of the QTL related to the trait biomass yield on the LGs 2, 3 and 7 of *L. perenne* after addition of SSR and STS markers. QTL were detected in the field experiment using IM, MQM mapping and CIM. The traits fresh weight (FW, blue), dry weight (DW, green) and dry matter (DM, black) were measured. QTL bars represent the 1 LOD fall-interval



Figure 12: Location of the QTL related to the trait biomass yield on the LGs 2, 3, 4 and 7 of *L. perenne* after addition of SSR and STS markers. QTL were detected in the greenhouse experiment using IM, MQM mapping and CIM. The traits fresh weight (FW, blue), dry weight (DW, green), dry matter (DM, black) and leaf width (LW, red) were measured. QTL bars represent the 1 LOD fall-interval.

<b>.</b>			IM		· ·		MQM			CIM		
		LG	LOD threshold	LOD score	%expl var.	Flanking markers	LOD score	%expl var.	Flanking markers	LOD score	%expl var.	Flanking markers
Fresh weight Field		2	2.8	3.09	3.8	LOC_Os04g54940-rv0062	9.22	8.4	rv1068-LpHCA17C11	3.98	8.4	EacaMcac415-rv0062
				6.9	8.5	rv0347-rv0188	-	-	-	6.92	9.6	rv0347-rv0188
		3	2.7	9.8	11.9	rv0863-EagaMctc230	4.91	4.3	EacaMctc100-EagaMctc230	13.17	20.7	rv0863-LpSSRK14F12
				12.95	15.5	EagaMctc230-LpSSRK14F12	6.76	6.1	rv1144-LOC_Os01g16152	2.94	11.5	LpSSRK14F12-rv1131
				9.72	11.8	LpSSRK14F12-rv0433	-	-	-	4.65	8.8	rv0029-rv1046
		7	2.5	5.58	7	rv0459-LpSSR020	4.69	4.1	G04-002 - rv1175	5.51	6.4	rv0459-rv1175
				-	-	I	-	-		3.05	4.2	rv1175-B1C8
Greenho	reenhouse	2	2.8	3.94	5	rv0062-NFFa136	4.34	4.4	rv0062-NFFa136	3.21	6.6	rv1269-G01-039
				3.82	4.8	rv1269-G01-039	-	-	-	3.33	5.8	rv1068-LpHCA17C11
				3.84	4.9	G04-059-G04-053	-	-	-	-	-	-
		3	2.6	7.46	9.2	rv0863-EagaMcta230	4.61	4.7	EacaMctc100-EagaMcta230	8.09	16.2	rv0863-EagaMcta230
				8.32	10.3	EagaMcta230-LpSSRK14F12	4.89	5	rv1316-B1A2	8.6	15.8	EagaMcta230-LpSSRK14F12
				5.88	7.4	LpSSRK14F12-rv1131	-	-	-	6.77	10.5	LpSSRK14F12-rv1131
				3.59	4.6	B3B8-rv1046	-	-	-	3.53	7	rv0029-rv1046
		7	2.3	4.15	5.3	EagcMcta051-rv0459	4.29	4.4	G04-002 - rv1175	5.27	7	rv0459-rv1175
				5.02	6.3	rv0459-rv1175	-	-	-	-	-	-
Dry weight Field		2	2.9	3.08	3.9	LOC_Os04g54940-rv0062	8.67	7.7	rv1068-LpHCA17C11	3.99	8.9	LOC_Os04g54940-rv0062
		2	2.8	6.34	7.9	rv0347-rv0188	-	-	-	6.56	8.9	rv1282-rv0188
		2	2.6	10.59	12.8	rv0863-EagaMctc230	5.47	4.7	EacaMctc100-EagaMctc230	14.44	20.5	rv0863-LpSSRK14F12
		3	2.6	13.61	16.1	EagaMctc230-LpSSRK14F12	6.68	6.0	rv1144-LOC_Os01g16152	9.81	12.5	LpSSRK14F12-rv0433
				10.13	12.3	LpSSRK14F12-rv0433	-	-	-	-	-	-
		_		6.58	8.2	rv0134-LpSSR020	5.63	4.8	B3A3 – rv1175	6.55	7.6	rv0459-rv1175
		7	2.7	-	-	-	-	-	-	4.01	5.5	rv1175-B1C8
	_											
Greenh	ouse	2	2.6	3.44	4.4	rv0062-NFFa136	3.47	3.5	rv0062-NFFa136	3.05	6.1	rv1269-G01-039
				3.39	4.3	rv1269-G01-039	-	-	-	3.06	5.2	rv1068-LpHCA17C11
				3.45	4.4	G04-059-G04-053	-	-	-	-	-	-

Table 12: QTL positions identified by IM, MQM and CIM, for the traits fresh weight, dry weight, dry matter and leaf width in the field and in the greenhouse experiment. For each test the QTL are displayed by LOD scores and percentage of explained variance.

		3	2.7	7.06	8.8	rv0863-EagaMcta230	4.53	4.6	EacaMctc100-EagaMcta230	8.09	16.1	rv0863-EagaMcta230
				7.52	9.3	EagaMcta230-LpSSRK14F12	4.35	4.4	rv1316-B1A2	8.9	15.7	EagaMcta230-LpSSRK14F12
				5.64	7.1	LpSSRK14F12-rv1131	-	-	-	7	10.7	LpSSRK14F12-rv1131
				3.11	4.0	B3B8-rv1046	-	-	-	3.76	7.2	B3B8-rv1046
		7	2.4	4.78	6.0	EagcMcta051-rv0459	6.11	6.3	G04-002 - rv1175	6.99	8.9	rv0459-rv1175
				6.82	8.5	rv0459-LpSSR020	-	-	-	3.66	5.7	rv1175-B1C8
Dry matter	Field	C	27	27	47	Log Oc04a54040 m0062	1 69	5.4	C04.050 L pHCA17C11	2 5 1	4.4	EacaMcac415-
		2	2.7	5.7	4./	C04.050 L :: UCA 17C11	4.08	5.4	004-039 – LPHCAT/CTT	5.51	4.4	LocOs04g55060
				4.44	5.0	G04-059 – LphCA1/C11	-	-	-	4.51	8.4	G04-059 - rv0188
				3.2	4.1	G04-030 – LpSSRK12E06	-	-	-	-	-	-
		3	2.7	-	-	-	4.51	5.2	rv1316-LocOs01g16152	7.26	18.6	m/11/4 B1A2
												IVII44-DIA2
		7	2.4	2.89	3.7	EagcMcta051-EacaMca213	4.8	5.5	rv1411-B1C8	4.12	5.9	rv0134-B1C8
				3.46	4.4	rv0459-B1C8	-	-	-	-	-	-
	Greenhouse		2.6	4.48	5.7	rv0863-EacaMctc100	4.48	5.57	rv0863-EacaMctc100	3.46	9.2	rv0863-EacaMctc100
		3	2.6	3.90	5	EacaMctc100-LocOs01g36890	-	-	-	4.69	6.9	EacaMctc100-LocOs01g36890
Leaf width	Greenhouse	2	2.7	4.07	5.1	rv1282-G01-039	4.54	4.8	rv1117-G01-039	4.69	8.2	rv1282-rv0037
		3	2.6	5.81	7.2	rv0863-EagaMcta230	7.99	8.8	rv1316-B1A2	7.14	9.6	rv0863-LpSSRK14F12
				8.35	10.1	EagaMcta230-LpSSRK14F12	-	-	_	4.05	54	LpSSRK14F12-rv1131
				64	79	LpSSRK14F12-rv1131	_	_	_	-	-	-
				3.61	4.5	B3B8-rv1046	_	_	_		_	_
				5.01	т.5	2320 1110-0						
		4	2.1	5.06	6.3	NFFa142-rv1412	5.38	5.8	NFFa017-rv0262	4.4	5.4	NFFa142-DLF025
				-	-	-	-	-	-	3.34	6.6	rv0262 - G04-099

## 5.2.2 QTL positions after addition of the DArT markers

#### Field experiment

DArT markers were added to the seven linkage groups and the QTL analysis was performed on the entire genome to check whether additional QTL could be detected after saturation of the genetic map. In accordance to the previous results, an association between markers and biomass yield traits was established for LGs 2, 3 and 7 (Figure 13). An additional QTL was detected on LG2. The test IM was not adapted to a map saturated with dominant markers. QTL were detected but the interval was very large covering half of the chromosome. A better precision was obtained with MQM and CIM which gave quite similar results. For MQM mapping the following markers were selected as cofactors: LG2-rv1269 and LoPt 355957, LG3- rv0674, LG7-rv1411 and LoPt 356133 only for the dry matter. As previously observed, QTL for fresh weight and dry weight were identified within the same chromosomal regions and remained linked to the same codominant markers. For each trait, two QTL were found on LG2, one on LG3 and one on LG7. A second QTL was detected on LG3 with CIM and has the particularity to be linked only to the DArT markers. IM and MQM mapping revealed dry matter QTL on LGs 2 and 7. The dry matter QTL on LG2 co-localised with the QTL for the two other traits.

All tests revealed the most significant QTL on LG3 but the LOD values and percentage of explained variance were considerably different between the tests (Table 13). In general the values were higher with CIM. Thus the contribution of the trait fresh weight was 76.3% of the total phenotypic variation with CIM while it was 23.9% with MQM.



Figure 13: Location of the QTL related to the trait biomass yield on the LGs 2, 3 and 7 of *L. perenne* using the map saturated with DArT markers. QTL were detected in the field experiment using IM, MQM mapping and CIM. The traits fresh weight (FW, blue), dry weight (DW, green) and dry matter (DM, black) were measured. QTL bars represent the 1 LOD fall-interval



Figure 13 continued

#### Greenhouse experiment

Compared to the field experiment no additional QTL were detected in the greenhouse experiment and the QTL positions were similar (Figure 14). The results obtained with CIM were the same with two QTL on LG 2, two on LG3 and one on LG7 for fresh weight and for dry weight with only one exception since no fresh weight QTL were found on LG7. As for the field experiment, no dry matter QTL were detected with the CIM test. The markers used as cofactors for MQM mapping were: LG2-M15185, LG3-rv0674 and LG7-rv1411. With IM and MQM mapping, QTL for the three analysed traits were found at the same position on LG7 than for the field experiment. However no biomass yield QTL were identified on LG2. Potential QTL could be detected on the LOD profile but the LOD score did not reach the LOD threshold. A common QTL was detected on LG3 and the second QTL found with CIM was also detected with IM. In

general the QTL effects in the greenhouse were lower than in the field experiment particularly for the trait fresh weight for which the LOD value and the percentage of explained variance were nearly divided by two. The QTL on LG3 detected with the three tests remains the most relevant QTL which could explain why this QTL was consistently identified.

For the trait leaf width, a QTL was detected on LG2 with IM and MQM within the same chromosomal region that the other traits. IM and CIM also identified QTL linked to that trait on LG3 and LG4 in accordance to the results prior to the addition of the DArT markers.



Figure 14: Location of the QTL related to the trait biomass yield on the LGs 2, 3, 4 and 7 of *L. perenne* using the map saturated with DArT markers. QTL were detected in the greenhouse experiment using IM, MQM mapping and CIM. The traits fresh weight (FW, blue), dry weight (DW, green), dry matter (DM, black) and leaf width (LW, red) were measured. QTL bars represent the 1 LOD fall-interval

(MQM)

LG2

LG7





**Figure 14 continued** 

			IM				MQM			CIM		
		LG	LOD threshold	LOD score	%expl var.	Flanking markers	LOD score	%expl var.	Flanking markers	LOD score	%expl var.	Flanking markers
Fresh weight	Field	2	2.4	6.96	9.5	LoPt356352- LoPt561324	3.04	3.1	rv1212-rv1269	3.8	5.1	LoPt558654-LoPt561989
-							6.6	7	LoPt355957-LoPt556019	3.81	5.8	rv0959-LoPt556019
		3	2.5	7.82	10.6	LoPt556224-LoPt355674	7.82	10.6	LoPt557365-G04-054	6.09	28.1	LoPt562096-LoPt561975
					-	-	-	-	-	14.02	26.8	LoPt556100-LoPt562281
		7	2.2	2.83	3	LoPt560802-LoPt555730	3.12	3.2	LoPt356133-LoPt557688	2.81	10.5	LoPt356133-rv1411
				3.04	4.2	LoPt555591-rv1411	-	-	-	-	-	-
	Greenhouse	2	2.5	-	-	-	-	-	-	6.42	8.6	LoPt558654-LoPt561989
				-	-	-	-	-	-	6.26	8.7	LoPt561153-LoPt557216
		3	2.6	3.77	5.3	LoPt556269-LoPt562096	3.56	5	LoPt557365-G04-054	3.81	17.8	LoPt562096-LoPt561975
				3.56	5	LoPt556100-G04-054	-	-	-	5.14	14.04	LoPt556100-LoPt562281
		7	2.2	2.55	3.6	LoPt560802-rv0474	3.07	4.3	LoPt561957-LoPt557688	-	-	-
				3.09	4.3	LoPt555591-LoPt561895	-	-	-	-	-	-
Dry weight	Field	2	2.5	5.86	8	LoPt356352- LoPt561324	2.79	3.6	rv1212-rv1117	3.97	5	LoPt558654-LoPt561989
							5.86	8	M15185-LoPt556019	6.04	8.7	rv0959-LoPt556019
		3	2.5	8.22	11.1	LoPt556224-LoPt355674	8.22	11.1	LoPt557365-G04-054	6.53	29.9	LoPt562096-LoPt561975
				-	-	-	-	-	-	14.94	27.2	LoPt556100-LoPt562281
		7	2.3	3.27	4.6	LoPt560802-rv0474	3.65	5.1	LoPt555591-rv1411	3.56	12.8	LoPt561957-rv1411
				2.81	3.9	LoPt555717-LoPt560176	-	-	-	-	-	-
	Greenhouse	2	2.4	-	-	-	-	-	-	5.62	7.7	LoPt558654-LoPt561989
				-	-	-	-	-	-	5.65	8.6	LoPt561153-LoPt557216
		3	2.6	3.65	5.1	LoPt556269-LoPt562096	3.23	4.5	LoPt557365-G04-054	5.85	27.6	LoPt562096-LoPt561975
				3.24	4.5	LoPt556100-G04-054	-	-	-	12.11	25.7	LoPt556100-LoPt562281
		7	2.2	3.26	4.6	LoPt560802-rv0459	4.49	6.2	LoPt561957-LoPt557688	2.97	7.8	LoPt356133-LoPt557688
				4.49	6.2	LoPt555717-LoPt560364	-	-	-	-	-	-

Table 13: QTL positions after addition of the DArT markers. QTL were identified by IM, MQM and CIM, for the traits fresh weight, dry weight, dry matter and leaf width in the field and in the greenhouse experiment. For each test the QTL are displayed by LOD scores and percentage of explained variance.

Dry matter	Field	2	2.5	3.41	4.8	rv1212-LoPt561989	7.71	10.5	M15185-LoPt556019	-	-	-
				7.66	10.4	rv0959-LoPt561324	-	-	-	-	-	-
		7	2.4	6.36	8.7	LoPt555303-LoPt557688	6.36	8.7	LoPt555303-LoPt356343	-	-	-
	Greenhouse	7	2.4	2.95	4.1	LoPt561957-LoPt557688	3.18	4.3	LoPt561957-LoPt557688	-	-	-
Leaf width	Greenhouse	2	2.4	2.6	3.6	M15185-LoPt556019	2.62	3.6	M15185-LoPt557688	-	-	-
		3	2.5	6.95	9.4	LoPt556224-G04-054	-	-	-	6.17	19.24	LoPt562096-LoPt561975
		4	2.5	4.96	6.8	NFFa142-LoPt560251	-	-	-	5.99	7.5	LoPt561704-LoPt555270

## 5.3 Discussion

The phenotypic analysis of the F2 mapping population in a field and in a greenhouse experiment leads to the detection of biomass yield QTL on LGs 2, 3 and 7 and of OTL for the morphological trait leaf width on LGs 2, 3 and 4. The fine mapping work presented in the two previous chapters focussed on LGs 2, 3 and 7 but DArT markers were mapped to cover the entire genome (Annex II). After saturation of the map no additional QTL were found on the other linkage groups. DArT markers were used to identify QTL for crown rust resistance using the same population (Tomaszewski et al. 2012) but to our knowledge no other QTL studies using DArT marker have been reported so far and little is known about the efficiency of this type of markers to detect QTL. Thus the QTL analysis was performed using the two "fine" maps to allow a comparison, using the common markers, of the QTL positions with or without the DArT markers. In addition, to overcome a lack of information about the effectiveness of the statistical models to analyze the DArT markers, three statistical tests were applied: IM, MQM mapping and CIM. The tests were selected due to their known power to detect QTL in maps saturated with dominant markers. The selection of the statistical model applied is important to detect all QTL. Thus the CIM test seems to be better adapted for QTL with a small effect and in the greenhouse experiment, the biomass QTL on LG2 could only be detected with that test. On the other hand non existant QTL can be identified. The values for the LOD score and the percentage of explained variance were higher with CIM than with the two other tests. A QTL analysis on barley was also performed using CIM and MQM and as for this study, the percentage of variation explained was higher with CIM (Laido et al. 2009). CIM might overestimate the effect of the QTL which could explain why some QTL were identified only with CIM. Independent of the efficiency of the test, a combination of different analyses is essential for identifying definitive QTL but also QTL that need to be treated with caution. A lack of regularity in the detection of some QTL may suggest the presence of false positives. But these QTL cannot be completely rejected since they might be real QTL.

Despite some variation in the QTL numbers and positions in function of the test applied and of the map, a majority of the OTL was detected in both maps. Thus on LG2, QTL linked to the markers M15185 and LoPt 355957 were always detected. The same on LG3 for the QTL linked to G04-054 and rv0674 and for the QTL on LG7 linked to rv1411. All these QTL had been previously identified by Anhalt et al. (2009) and represent a starting point for the dissection of the chromosomal regions affecting the biomass yield trait. Additional QTL regions were found but not systematically. On LG2, a QTL linked to rv1269 was detected and prior to the addition of the DArT markers, QTL linked to rv1269 were identified with IM and CIM. But the marker rv0959 was also within the QTL interval and after the addition of the DArT markers, rv0959 and M15185 mapped together. Perhaps it is the same QTL or two QTL were narrowed down to a single one. With both maps a second QTL region was found on LG7, flanked with rv0474, but only with IM. Similarly a second QTL region was detected on LG3 only with CIM. Before the addition of the DArT markers, between two to four QTL per trait were identified on LG3 but due to a lack of common markers it is not possible to compare the QTL. These last three QTL regions need to be treated carefully and have to be considered as indicative rather than unequivocal. This is also the case for the leaf width QTL analysed in only one environment.

The three highly statistically confident QTL regions were characterized by overlapping QTL for the traits fresh weight, dry weight and dry matter. Some variations

in the detected QTL were observed within those regions. In general the QTL in the greenhouse had less effect and some on LG2 or LG7 could not be detected by all tests. These are indications that the traits are under environmental influence. In the greenhouse, with IM and MQM, none of the biomass QTL on LG 2 were detected which could result from a LOD score below the significant level. A total of four harvests were performed in the field and three in the greenhouse. Field data were collected over a period of two years characterized by two distinct climatic conditions during summer and it is likely that there are less phenotyping errors in the field which could explain the unequal QTL detection and effects. But for some of the LOD peak was observed (Annex III, Annex IV). This suggests that different genes are involved depending on the environment. Environmental effects have been reported in many studies of *L. perenne* (Barre et al. 2009, Turner et al. 2006) and for the trait fresh weight in *Arabidopsis thaliana* (Rauh et al. 2002) and in rice (Liu et al. 2006). This highlights the complexity of the regulation of the biomass yield traits.

One of the objectives of the study was to identify molecular markers linked to biomass yield to aid the breeding selection process in early cycles. The success of MAS in a complex trait like biomass yield is challenging. But the persistency of the QTL within the three targeted regions over the environments and the tests suggests a possible application in a *L. perenne* breeding program. Selection based on one marker per QTL is not as reliable as selection using the flanking markers, or a suite of markers, (Collard et al. 2005) but due to some light shift of the QTL positions between the different tests or environments, consistent pairs of flanking markers could not be determined. Three markers, G04-054, M15185 and rv1411, were underlying the QTL in the different maps and are potential candidates for MAS. They are codominant, which is advantageous in plant breeding. For the test of homozygosity often required in selection, with codominant markers less material needs to be tested than with dominant markers (Bonnett et al. 2005). The markers G04-054 and M15185 were mapped respectively in a consensus map (Studer et al. 2010) and in an integrated map from a F2 and BC1 population (Armstead et al. 2002). The transferability of markers between populations is fundamental, allowing the identification of similar chromosomal regions within populations with different genetic backgrounds. Additionally G04-054 is linked to a QTL region particularly interesting for its large effect. Indeed on LG3, the fresh weight and dry weight QTL explain with CIM respectively 26.8% (LOD score 14.02) and 27.2% (LOD score 14.94) in the field and 17.8% (LOD score 5.14) and 25.7% (LOD score 12.11) in the greenhouse. The values were not that high with the other tests and tend to be around 10% in the field and 5% in the greenhouse. On LG7 the effect of the QTL is above 10% when using CIM but it falls down to 3-5% with IM and MQM. It is commonly accepted that QTL with an effect in excess of 10% are considered major QTL (Collard 2005). Major QTL are in general highly heritable and stable across environments (Collard 2005), and might be controlled by a few major genes with strong effects and are appropriated to underline the genes involved in the trait. In consequence, markers associated to a major QTL are more effective for MAS. In addition to the major QTL, several QTL with small effects were detected, as expected for a large population and for a complex trait like biomass. Among them the QTL associated to the markers M15185 on LG2 is particularly interesting since QTL from different kind of traits were overlapping. The three biomass traits and the morphological trait leaf width were analogous. Another trait affecting the biomass is the heading date and heading date QTL were identified in two separated studies on LG2, approximately at 20 cM from the marker M15185 (Byrne et al. 2009, Armstead et al. 2004). It is uncertain whether these QTL are at the same position as the biomass and leaf width QTL, but considering that our QTL interval covers a large region of around 35 cM it is an hypothesis that needs to be considered. On LG2, a QTL associated to leaf length was also identified (Barre et al. 2009) but in the absence of shared markers the QTL position cannot be compared. It was observed in several studies (Studer et al. 2008, Yamada et al. 2004, Hittalmani et al. 2002) that correlated traits tend to co-localise in the same genomic regions. Since leaf width and heading date both affect biomass yield the co-localisation of their QTL could be used to simultaneously identify plants with one or more alleles for several interesting traits. But overlapping QTL can result from pleiotropy or tight linkage of the genes or even overlapping genes and are a problem for selection of specific traits. And allele favoured by one trait may be removed by selection of another trait. And in the case of negative correlation between the traits, undesirable traits can be selected.

The benefit of QTL analyses for an application of MAS in breeding programs is limited, due to the specificity of the identified QTL to the population studied, and information collected may not be valuable in populations with different backgrounds (Collard and Mackill 2008, Yamada et al. 2004). The markers linked to a QTL in a population can be quite distant from the responsible polymorphism in another population. Few studies have been carried out on the four traits analysed in this study. Armstead et al. (2008) reported a leaf width QTL on LG7 but their map has no markers in common with the current study and it is speculative if the QTL are identical. Turner et al. (2008) reported autumn dry matter QTL on LG 3 but none on LGs 2 and 7. Interestingly, Yamada et al. (2004) found a fresh weight major QTL on LG 4 (LOD score 5.6, % explained variance 22.8 using CIM). It is surprising that this QTL with a strong effect was not detected in our population, and underlines the influence of the genetic background. In addition, the biomass QTL detected were highly affected by the environment. Due to a lack of sufficient resolution, all the markers under biomass QTL will have to be tested to confirm the linkage between the markers and the QTL. This is particularly important for a map with segregation distortion and having a high proportion of dominant markers. Segregation distortion and dominant markers may cause problems in ordering the markers on the map (Asins 2002) and then for the detection of the QTL. To avoid the effect of the SD, highly distorted markers were excluded from the mapping analysis. But it was clear that the SD in the F2 population was the result of a biological phenomenon and the SD had to be taken into account for the QTL analysis. Therefore, the QTL mapping work needs to be confirmed in several segregating populations and under different environmental conditions.

The testing and validation of the QTL and the markers in several populations and environments is time- and money- consuming. The reproducibility of the biomass QTL over environments and over the statistical tests indicates a robustness of the QTL regions identified. This is not sufficient for a direct application in MAS but it is promising for the identification of candidate genes in the future. The second objective of this study was to reduce the QTL intervals to a size suitable for the search for potential candidate genes. Despite the fine mapping work with the addition of a high number of markers, the QTL were assigned to regions of about 20 cM, which is large at the molecular level. Because of the map expansion due to the necessary use of the Maximum Likelihood algorithm, it is difficult to estimate how much the QTL intervals were reduced. The distance on the genetic map is not necessarily representative of the distance on the physical map and additional analyses are required. Usually to achieve the construction of a physical map, AFLP fingerprinting or Fluorescent *In Situ* Hybridization methods are applied. However despite the presence of large gaps, the map can be saturated. A QTL analysis was performed as a test by removing around a quarter of the DArT markers. The results were identical to the results presented in this chapter and seem to indicate that the map was saturated. Thus the addition of more markers would not lead to a reduction of the QTL intervals. The saturation of the QTL intervals only, an increase of the population to 500-1000 genotypes or the construction of near-isogenic lines (NILs) population may overcome the problem (Asins 2002, Kearsey and Farquhar 1998). NILs are obtained by backcrossing to establish lines containing introgression of one parent's alleles into the genetic background of another parent (Buerstmayr et al. 2009). This allows association of the phenotypic variation to the introgressed region (Brouwer and St Clair 2004) and thus reduces the QTL interval to that specific region. NIL populations have been successfully used for fine mapping and cloning (Liu et al. 2009, Li et al. 2004).

In this chapter the positions of QTL for the traits fresh weight, dry weight, dry matter and leaf width were recalculated using a map saturated with DArT markers. This was the first application of the DArT markers in a QTL analysis. QTL were found on LGs 2, 3 and 7 in accordance with the previous results (Anhalt et al. 2009) and mostly remained associated to the same codominant markers. It is difficult to know whether there is a strong linkage between the codominant markers and the QTL or if the dominant markers cannot detect QTL. The markers underlying the QTL will have to be tested to confirm the linkage and will thus be powerful tools in breeding programs.
The main purpose of the work presented in this chapter was to reduce the QTL interval to facilitate the identification of candidate genes. Despite the saturation of the map, the size of the QTL intervals did not decrease. However compared to the first biomass yield QTL detected (Anhalt et al. 2009), the intervals were better covered with markers. These markers will be used to screen a *L. perenne* BAC library in order to analyze the QTL region. This will give more information about the physical distance of the QTL intervals and about the genomic sequences underlying the biomass QTL.

# Chapter 6: Identification and description of *Lolium perenne* genomic sequences underlying biomass QTL

# 6.1 Introduction

Although QTL mapping is a necessary step to dissect complex traits, the identification of QTL regions is not sufficient. Due to a lack of resolution and mostly to the variation of the genetic background between populations, markers linked to quantitative traits are not directly suitable for an application in breeding programs. Thus marker assisted selection (MAS) will only become a powerful tool when the relation between the phenotypic variation in a trait and the genetic polymorphism underlying this trait has been established (Morgante and Salamini 2003). Therefore the isolation of the genes responsible underlying the trait is crucial. Markers developed from the candidate genes will permit the systematic prediction of phenotypes (Varshney et al. 2006). For complex traits, gene cloning remains difficult due to the small effects of some genes on the phenotypic variation and to the poor precision in the localisation of the genes on the genetic map. Several gene cloning approaches were developed with among them map-based QTL cloning. The development of Bacterial Artificial Chromosome (BAC) libraries has facilitated map-based cloning with or without prior knowledge of the genes associated to the trait studied.

# 6.1.1 BAC libraries as tool for identifying genomic sequences

The ability to clone large fragments of DNA is vital for the structural and functional analysis of complex genomes. Cosmid vectors were introduced by Collins and Hohn (1978) and consist of a combination of a plasmid vector with the cohesive end sites of the  $\lambda$  bacteriophage, allowing the insertion of the targeted DNA into the  $\lambda$ 

head. The DNA fragment can be easily extracted and purified. However the cosmid cannot carry DNA fragments larger than 50 kb (Burke et al. 1987). To circumvent this limit, Yeast Artificial Chromosome (YAC) libraries were developed (Murray and Szostak 1983) allowing the cloning of DNA fragments larger than 100 kb. However a high level of chimerism, low stability and difficulties in purification of the cloned DNA inserts have restricted their application (Peterson et al. 2000, Shizuya et al. 1992). On the contrary, the BAC libraries are a new alternative for cloning large fragments of DNA and have the advantages of being more stable, easy to manipulate and the sequencing of DNA from BAC clones is simple (Farrar and Donnison 2007). The general procedure for the production of a BAC library consists of the isolation of nuclei for High-Molecular-Weight DNA extraction. The DNA is then partially digested, sizeselected and inserted into a vector before the cloning into bacteria (Zhang and Wu 2001). The first BAC library was developed for the human genome in 1992 (Shizuya et al. 1992) followed two years later by the first BAC library for the grass species Sorghum bicolor (Woo et al. 1994). From then on the number of BAC libraries has been steadily growing and BAC libraries exist for most grasses such as rice (Zhang et al. 1996), wheat (Nilmalgoda et al. 2003), maize (O'Sullivan et al. 2001), Brachypodium (Farrar and Donnison 2007), meadow fescue (Donnison et al. 2005) and perennial ryegrass (Farrar et al. 2007). The construction is a long and difficult process but once developed, it is a powerful tool for genome analysis, physical mapping, cloning of genes and the comparison of specific regions between species (Farrar and Donnison 2007, Chalhoub et al. 2004). A BAC library is useful to identify candidate genes within QTL regions and for MAS by facilitating the detection of valuable markers linked to the genes of interest. Using BAC sequences further molecular markers can be developed within the critical regions. It also enables the identification of candidate orthologous

sequences within a species for a gene already identified in another closely related species.

# 6.1.2 Amplified Fragment Length Polymorphism (AFLP) fingerprinting

On a chromosome, the recombination frequency varies from location to location, resulting in inconsistency between the genetic and physical distances along the chromosome (Peters et al. 2003). In addition to the genetic map, construction of a physical map for the region of interest will inform about the distance between markers or about the size of a QTL interval. A physical map consists of the ordering of DNA fragments, with the distance between them expressed in base pairs (Meyers et al. 2004). For its construction, clones are isolated by screening a BAC library with mapped markers. Afterwards the isolated BAC clones are fingerprinted using AFLP. The BAC DNA is digested with two restriction enzymes. Generally for BAC DNA, EcoRI or HindIII are combined with MseI or TaqI. The digested DNA is then ligated to adapters homologous to one 5'- or 3'- end generated during the digestion (Blears and al. 1998, Vos et al. 1995). The sequences of the adapters serve as primers for a PCR amplification. Overlapping clones produce shared restriction fragments and the number of shared bands is counted to determine the degree of overlap (Meyers et al. 2004). Clones are considered to overlap when the probability that the shared bands are not a coincidence is above a predefined cutoff (Soderlung and al. 2000). The fingerprints of the clones also allow the arrangement of genomic fragments to build contiguous sequences. Physical maps using AFLP fingerprinting have been reported in potato (Visser et al. 2009) and sorghum (Klein et al. 2000) among others. This method is straightforward and independent of genomic information, and thus suitable for large genomes.

#### 6.1.3 Context and Objectives

Several genes that code for various traits such as the vernalization response (Ciannamea et al. 2006) or sucrose 1-fructosyltransferase (Chalmers et al. 2003) have already been successfully identified in L. perenne. By screening a L. perenne BAC library, with primers involved in candidate genes for disease resistance and forage quality, a full length sequence of those candidate genes was obtained (Farrar et al. 2007). In another study, the Festuca pratensis BAC library was screened to identify the orthologous region to the heading date *Hd1* gene from rice with a marker close to this gene. The BAC sequence obtained was then used to identify the same region on L. perenne (Armstead et al. 2005). But due to the complexity of the trait biomass yield and perhaps because there is less interest in that trait in major or model crops, there is only limited knowledge of potential genes linked to that trait. Three highly confident QTL regions associated to the trait biomass yield were identified on LGs 2, 3 and 7 in an F2 biomass population and are valuable for the search of potential candidate genes. The screening of the BAC library with the markers spanning the QTL regions may allow the isolation of the genomic sequences containing genes affecting the trait biomass yield. To facilitate the identification of candidate genes, the construction of a contig with the clones underlying the QTL is required. But despite the saturation of the genetic map with SSR and DArT markers, the QTL intervals remained large. The addition of many dominant markers induced a map expansion which perhaps does not reflect the real distance between markers.

A *L. perenne* BAC library has been developed by the University of Arizona. The work presented in this chapter consists of the screening of this BAC library with the markers underlying the QTL to isolate clones covering the QTL regions. The clones will be analysed using the technique of AFLP fingerprinting to build contigs and see if

some clones are overlapping. The objectives were (1) to analyse the large distance between markers observed with the genetic map, (2) to detect overlapping clones to construct a single contig for each QTL interval, and (3) to select BAC clones for full BAC clone sequencing to enable the prediction and annotation of genes. The BAC sequences will provide information to facilitate the cloning of genes linked to the trait biomass yield such as the presence of transposable elements, the gene density, and the presence of already known genes that could affect the trait found within the three QTL regions.

# 6.2 Results

#### 6.2.1 Screening of a L. perenne BAC library

The *L. perenne* BAC library provided by the University of Arizona was screened with markers underlying and flanking the three highly confident biomass QTL found on LGs 2, 3 and 7 (QTL linked to the markers M15185 on LG2, G04-054 on LG3 and rv1411 on LG7) (see chapter 5). Since there is a possibility of inversions between markers mapping close to each other, markers were selected to cover a region larger than the QTL interval. On LG7, all the markers until the end of the chromosome were used. Often QTL localised at the telomeric region do not map at the extreme end although it is there real position (Kearsey and Farquhar 1998). For the DArT markers, primer pairs were developed from their sequences (Table 14). The sequences were provided by David Kopecky (Institute of Experimental Botany, Olomouc, Czech Republic). The absence of primers for some DArT present around the QTL intervals was due to the impossibility of sequencing the DArT.

Name	Forward primer seq. 5' to 3'	Reverse primer seq. 5' to 3'
LoPt 556100	TGAAGCACGGAGCACTCTTG	GGAAATCCCACACAGGCAGT
LoPt 557216	GCCATTTCTCTGCCAATTCC	TAAACCCACGAGCACACTGC
LoPt 355957	CGTTCCACTGCAACTCTCCA	GCCGACAACCAAAGATCCTC
LoPt 356232	CCGTTCACCTAACTCCCTGTG	TCAGAGCATAAGATGGGAGCA
LoPt 557688	GGCCGAGCTAAACCGATACA	TACCTCTCCGCCACTGAACA
LoPt 557724	TTTGGACGAAAGCATGAACG	AGGGTTTGCCAAATGCAGTG
LoPt 562127	GGCCTGCTCCTATTGTCCAC	GAGACATGCGGTGAGACAGC
LoPt 560176	CAACCTGGAGCAAGAAAATGC	TGCAGCTGGTGCTAATCCTG
LoPt 556019	CCGCGCAGAAGATTCAACTA	GGTGACAGAAAATGGCAGCA
LoPt 557116	ATCACGCGCAGAAGATTCAA	CTGGCCTGCTCACTGTCTTG
LoPt 562722	TGTTCATTACCGCAGGCAAC	CGCAGACCCCTTTGAAGGTA
LoPt 561895	GGCTCATTGCTGAGGGTGTT	CTGACCCAACTCAGGCATTG
LoPt 561690	CTCAGAAACATGAGCGGAATG	GACAGTAGCGATGCGAAAACC
LoPt 355674	GGTATTCTCCACGACGCAGA	TCATCCAGGTGCGTTTATGC
LoPt 555591	AGTGCTCCGAAACATCTTGG	TATCGACGCAAGGCATGTTA
LoPt 561957	GACCAGCTCGGCTGTTCTCT	GCCAATGCTTCTGAGCACAC
LoPt 356133	ACATTGGAAACCAAGGCATC	TTGAAGCATCTGACCCCTTC
LoPt 562281	GAAGTTGCACAGCTTGGACA	GGACTGCTTGAAGTTGCTC
LoPt 557365	TGCGATCATTCTGCTGACCT	CGCATTTTGAGAGCTCTTGC
LoPt 556516	CTGAGGCGACCAGTCTTCAG	AGGGGGAATTTGGAAGTTGG

**Table 14: Primers developed from the DArT sequences** 

Two types of pooling were used for the screening: (1) a modified 6 dimensional pooling combining the screening of library plates followed by the screening of a two-dimensional row-column superpool of the positive plates (see 2.4.3 Pooling of the BAC DNA) and (2) a multi-dimensional pooling, provided by Amplicon express (US) that is quicker for the screening with many markers since BAC DNA isolations are not required. The screening of the SSR markers was performed with the modified step-wise pooling and clones were identified for all the markers except rv0674 mapped on LG3. The Amplicon express pooling was initially used for the DArT markers but positive clones were identified for less than half of the DArT markers. After several repetitions under various conditions, the step-wise pooling was used and allowed identification of clones for three more DArT markers but not for all. The DArT markers gave too much

amplification with the Amplicon express pooling but too little with the step-wise pooling.

The main difficulty was to adapt the PCR conditions for each marker to get clear bands of amplification as presented in Figure 15. The PCR conditions described in the materials and methods are the ones systematically used and giving the best results. But for many markers, using these conditions, it was not possible to get hits or on the contrary all the pools were amplified. The PCR had to be redone with different annealing temperatures or DNA concentrations.



Figure 15: BAC library screening showing the analysis of the Amplicon express pooling with the marker M15185. A first round of PCR was performed with the superpool (A). Then a second PCR was performed on a matrixed plate, row, column pools (B) to identify the position of the clones in the BAC library. DNA from the *L. perenne* variety Cashel was used as positive control (+).

On average 3.6 clones per marker were identified with a variable number from 1 to 7 clones per marker (Table 14, Table 15). The BAC library has a 5 genome equivalents coverage and an average of 5 hits per marker was expected. The difficulty of getting amplification with some of the DArT markers probably induced the non detection of some clones. For many markers it was not possible to identify clones inspite of the application of different PCR conditions. For these markers, it was often not possible to

amplify the genomic DNA used as positive control, indicating that perhaps nonamplification was not necessarily the result of improper PCR conditions.

LG	Markers	Hits	Clones
LG2	LOC-Os04g55060	4	279G11, 198L10, 159A21, 305C24
	rv0959	6	41J23, 81H2, 173M6, 160J13, 203H2, 231I13
	M15185	4	30C14, 180D4, 113A19, 232B24
	LoPt 355957	2	153A21, 157L6
	LoPt 556019	0	
	LoPt 557116	0	
	LoPt 557216	0	
	LoPt 561690	0	
	LoPt 556516	2	157G15, 162O6
LG3	LoPt 562722	4	63P16, 76L17, 205K12, 232C22
	LoPt 562127	3	279M7, 167A14, 266J13
	LoPt 557365	7	193O23, 18A15, 53N22, 251B9, 61J2, 12P7,
			219A19
	LoPt 556100	6	39L3, 63P18, 76L17, 143A22, 152C15, 205G12
	rv0674	0	
	G04-054	2	20119, 222J11
	LoPt 562281	3	176L10, 142O21, 218P11
_	LoPt 355674	0	
LG7	LoPt 356232	2	33K7, 239I19
	LoPt 555591	0	
	LoPt 561957	0	
	LoPt 356133	0	
	rv1411	4	12D24, 113M13, 165L14, 59E15
	LoPt 557688	0	
	LoPt 557724	0	
	LoPt 560176	3	29C16, 64G18, 80B12
	LoPt 561895	2	203P18, 220O22

 Table 15: List of the clones isolated for each marker used to screen the BAC library.

# 6.2.2 AFLP fingerprinting

For the construction of contigs encompassing biomass QTL intervals, a fingerprint of each BAC clone was produced by non-selective AFLP EcoRI and MseI

primer amplification. A total of 45 fingerprints out of 54 reactions were obtained and on average the BAC clones displayed 22 bands. Fingerprinting failures resulted from a low efficiency of the enzymes to cut the BAC DNA. Another enzyme, HindIII, was tested but gave similar results. The assembly was performed with the software FPC version 9.3 (Soderlung et al. 1997) and 26 clones were classified into eight contigs consisting of overlapping clones linked to the same molecular marker (Figure 16). This confirms a large distance between markers as previously observed with the genetic linkage map. An exception concerns the two markers LoPt 562722 and LoPt 556100. It was possible to form one contig with clones from both markers. This result was expected since the clone 76L17 showed amplification with both markers. However, these two markers do not map together on the genetic map, and three markers were mapped between them (Figure 16). For two of the three markers the fingerprint failed however fingerprint for the clones isolated by LoPt 557365 were available but could not be assigned to the contig formed with the clones from LoPt 562722 and LoPt 556100 even when the cutoff was reduced. This indicates that some errors in the order of markers remained. Another specificity concerns the clones from the marker M15185, separated into two contigs, suggesting that the clones were in repulsion phase.



LG3

Figure 16: Result of the assembly of the BAC clones with the program FPC version 9.3 (Soderlund et al. 1997). A total of 26 BAC clones out of 54 were assembled into eight contigs. Each yellow square corresponds to a contig and displays the overlapping clones.

# 6.2.3 Gene prediction

Five clones were selected for sequencing. The clones were chosen based on markers positioned below the LOD peak of the biomass QTL to increase the probability of studying a genomic sequence containing genes affecting the biomass yield. The map generated with the DArT was not fine enough to reduce the QTL intervals and perhaps markers mapping far away from the LOD peak are also far away from the genes. The selection of clones underlying the QTL on LG2 was justified by the identification of crown rust QTL overlapping with the biomass QTL using the same genetic map (Tomaszewski et al. 2012). Thus the collected information will be valuable for different studies. Clones from LG3 were selected because the LG3 biomass QTL was a major QTL and the strongest detected in this study. Identification of genes is usually more successful when based on a major QTL that is generally characterized by few genes with major effects (Collard 2005). The clones were sequenced to a minimum 100 fold coverage using a Roche GS FLX by GATC (Germany) and the assembly was performed with the program GS De Novo Assembler (Table 16). The BAC clone sequences have an average size of 144Kb.

Address	Marker	Number of contigs	Total size	% GC content
20I19	G04-054	36	167109	42.62
222J11	G04-054	31	129905	42.07
180D4	M15185	46	108282	42.77
30C14	M15185	54	169419	42.17
153A21	LoPt 355957	27	149283	43.46

Table 16: Characteristics of the five BAC clone sequence assemblies.

The presence of repetitive sequences throughout the BAC sequences was observed with a dot plot of the sequences against themselves (Figure 17) using PipMaker (Schwartz et al. 2000). The numbering of the contigs was done in function of the contig size, from the largest to the smallest. Duplications and translocations were mainly observed within small contigs although the dot plot revealed widespread repetitive sequences along the sequences. Repetitive sequences often lead to false positive results. Contig 1 of the clones 222J11 was characterized by a region of low complexity while little repetitions were observed for the clones 153A21.



Figure 17: Dot plot of the five BAC sequences against themselves performed with Pipmaker (Schwartz et al. 2000). The contig numbers were assigned in function of the contig size, from the largest to the smallest





Figure 17: continued





Figure 17: continued

Considering the synteny observed previously with the rice STS markers (see chapter 3), a great microcolinearity between L. perenne and rice was expected. A strategy based on the synteny was applied to order the contigs from a clone and each contig was blasted against the rice genome. Little microsynteny was observed with less than a third of the contigs showing homology to the rice genome. Regarding the similarity, less than 10% of the contig sequences were covered but for those regions covered the similarity level itself was high. The contigs were also blasted against the Brachypodium genome without better results. In consequence, the gene prediction and annotation was performed with BAC clone sequences in phase I, i.e. with unordered contigs (Zhao 2000). Only the contigs with a length above 1000 bp were further analysed since small contigs were mainly composed of repetitive elements. Sequences containing repetitive regions tend to be problematic when performing sequence analysis and often lead to false positive results. Using the *ab initio* method several predicted genes were identified (Table 17). The Rice Genome Automated Annotation System (RiceGAAS http://ricegaas.dna.affrc.go.jp/) program was used to optimize the gene prediction. This program has the advantage of combining several gene prediction softwares. Notable variations between the different softwares were observed with Fgenesh and GenScan giving the highest number of predicted genes. The gene prediction revealed that genes on LG3 were smaller than genes on LG2 (Table 17).

BAC clone	20I19	222J11	180D4	30C14	153A21
Total number of genes	44	32	23	35	39
Gene density (1gene/x bp)	x=3797	x=4059	x=4707	x=4840	x=3827
Average gene size	1580 bp	1417 bp	1962 bp	1763 bp	1888 bp
Exon number	125	86	78	109	151
Average number of exons per gene	2.8	2.7	3.4	3.1	3.9
Average exon size	305 bp	305 bp	260 bp	283 bp	228 bp
Average number of introns per gene	1.8	1.65	2.35	2.1	2.87
Average intron size	391 bp	361 bp	461 bp	417 bp	349 bp

Table 17: Features of the five genomic sequences studied.

From a total of 173 gene structures predicted, it was possible to assign a function to 59 (34%) genes using BLASTn and BLASTp (Table 18). The low percentage is explained by the lack of alignment between L. perenne and species with annotated genomes such as rice, to the low amount of annotation available or to the presence of false positives resulting from the repetitive sequences. The annotation revealed numerous gene structures showing homology to proteins of unknown function. However the two clones linked to the markers M15185 showed homology to Hordeum vulgare partial lg1 gene for liguleless-like protein (Rossini et al. 2006). And the five clones showed homology to regions flanking four interesting genes: L. perenne heading date (Hd1) gene (Skøt et al. 2007), Avena strigosa beta-amyrin synthase (Sad1) and cytochrome P450 CYP51H10 (Sad2) genes (Qi et al. 2006) and Lolium multiflorum gene for cold responsive protein (Oishi al. 2010) et

Clones	Accession	Description	Distance to	E value	% of identity
180D4	AM489608	Lolium perenne hd1 gene	14051 bn	e-118	99
1000	AB533342	Lolium multiflorum gene for cold responsive protein	3402 bp	3e-64	96
	DO680849 1	Avena strigosa cytochrome P450 CYP51H10 (Sad2) genes	21479 bp	8e-41	86
	BAH79979.1	putative unclassified retrotransposon protein	<b>_</b> 1.//> op	9e-88	59
	XP 002467226.1	hypothetical protein SORBIDRAFT 01g021625		0	59
	AM117950.1	Hordeum vulgare partial lgl gene for liguleless-like protein, exons 1-3	() hn	1e-123	90
	NP 001105543.1	protein LIGULELESS 1	0 bp	2e-73	66
	CAJ41454.1	liguleless-like protein [Hordeum vulgare]	0 Up	2e-50	95
	EAY95872.1	hypothetical protein OsI 17738		1e-75	65
	ABA98614.1	transposon protein, putative, CACTA, En/Spm sub-class		e-120	31
30C14	AM489608	Lolium perenne hd1 gene	14051bp	e-140	99
	ABA98838.1	transposon protein, putative, Mutator sub-class		e-121	37
	ADB85429.1	putative retrotransposon protein		7e-46	37
	AB533342	Lolium multiflorum gene for cold responsive protein	3237 bp	9e-78	96
	DQ680849	Avena strigosa cytochrome P450 CYP51H10 (Sad2) genes	21954 bp	0	85
	BAH79979.1	putative unclassified retrotransposon protein	<b>_</b>	1e-90	59
	XP_002467226.1	hypothetical protein SORBIDRAFT_01g021625		0	87
	AM117950.1	Hordeum vulgare partial lgl gene for liguleless-like protein, exons 1-3	() hn	4e-123	90
	ABA98614.1	transposon protein, putative, CACTA, En/Spm sub-class	0 op	e-120	31
222J11	BAI39843.1	hypothetical protein		2e-37	38
	AB533342	Lolium multiflorum gene for cold responsive protein	2210 bp	3e-66	95
	ABA95102.1	retrotransposon protein, putative, unclassified	Ĩ	0	57
	BAK08192.1	predicted protein		1e-20	50
	AAM11916.1	unknown cold induced protein		5e-51	65
	EEC71449.1	hypothetical protein OsI_03671		e-173	68
	BAD88086.1	ribosomal protein-like		6e-81	54
	BAJ98145.1	predicted protein		e-163	85
	AM489608	Lolium perenne hd1 gene	11553 bp	4e-35	95

 Table 18: Annotation of the predicted genes using BLASTn and BLASTp for the five sequenced BAC clones. The five interesting genes are highlighted in blue.

20I19	AM489608	Lolium perenne hd1 gene	11553 bp	4e-35	95
	BAJ85556.1	predicted protein	-	0	89
	AAV25047.1	putative polyprotein		3e-94	38
	BAD88086.1	ribosomal protein-like	2368 bp	6e-81	54
	AB533342	Lolium multiflorum gene for cold responsive protein	<b>- I</b>	4e-66	95
	BAD19994.1	cytokinin inducibl protein-like		2e-21	42
	ABA98574.1	retrotransposon protein, putative, Ty3-gypsy subclass		0	45
	EAZ13478.1	hypothetical protein OsJ_03394		1e-66	69
	BAJ98072.1	predicted protein		4e-66	63
	AAM11916.1	unknown cold induced protein		5e-51	65
	ABA95102.1	retrotransposon protein, putative, unclassified		0	60
153A21	BAJ93991.1	predicted protein [Hordeum vulgare subsp. vulgare]		e-145	87
	EEE53594.1	hypothetical protein OsJ_36841		3.00e-18	73
	BAJ99959.1	predicted protein [Hordeum vulgare subsp. Vulgare		2e-20	70
	ABF96975.1	retrotransposon protein, putative, Ty1-copia subclass		1.00e-19	44
	AAO37490.1	putative gag-pol polyprotein [Oryza sativa Japonica Group]		1.00e-19	44
	XP_002449887.1	hypothetical protein SORBIDRAFT_05g024920		4.00e-68	32
	AB533342	Lolium multiflorum gene for cold responsive protein	2275 hr	1e-26	96
	CAH65761.1	H0215A08.3 [Oryza sativa Indica Group]	5275 Up	9e-81	50
	XP_002465880.1	hypothetical protein SORBIDRAFT_01g047515		1e-73	46
	XP_002451167.1	hypothetical protein SORBIDRAFT_05g025250		2e-65	42
	XP_002465250.1	hypothetical protein SORBIDRAFT_01g034970		6e-48	34
	EAY95642.1	hypothetical protein OsI_17507		4e76	47
	ABA93762.1	transposon protein, putative, CACTA, En/Spm sub-class		0	46
	XP_002438099.1	hypothetical protein SORBIDRAFT_10g007985		4e-48	44
	NP_001169638.1	hypothetical protein LOC100383519 [Zea mays]		2e-37	30
	EEC78180.1	hypothetical protein OsI_17777		0	81
	BAJ88077.1	predicted protein [Hordeum vulgare subsp. vulgare]		3e-81	80
	NP_001157806.1	topoisomerase I [Zea mays]		0	62
	EEC78185.1	hypothetical protein OsI_17782		0	48

# 6.2.4 Mapping of the four genes

Ab initio gene prediction is based on statistic analysis and a validation of the results is necessary. In order to confirm the presence of the lg1-like gene within the biomass QTL region on LG2 in *L. perenne*, primer pairs were designed from the gene sequences. In addition, the fact that the five clones showed homology to regions closed to genes suggested that these genes may underlie the QTL. In the absence of sequences covering the entire genomic region under the QTL, the only possibility to confirm or not the possible presence of the genes was to map markers designed from the genes sequences. Thus, seven primers were designed for the gene heading date, eight for cold response, five for the gene *Sad2* and one for liguleless. An amplification signal was produced by 85% of the primers and SNPs were detected for two primers. These two primers, presented in Table 19, were added to the genetic linkage map (Figure 18).

Table 19: Primers designed from gene sequences that were mapped onto the *L*. *perenne* linkage map.

Genes	Sequences
Liguleless (Lg1)	F: TGCTTGATGAGTTCGACGAC R: GAAGGATGTTGCTGTGCTGA
Cytochrome P450 CYP51H10 (Sad 2)	F: TGATTGTACTGCGGAAGCTG R: GTCCGTTTGCATTCGTAGGT

The markers were designed from genes identified on both LGs 2 and 3 but they both mapped within the QTL interval on LG2 and not on LG3. While identified in a single clone indicating a small physical distance between them, the distance between the two genes on the genetic map was 23.2 cM confirming a lack of precision of the map.



Figure 18: Mapping of three markers that could amplify a part of the genomic sequences of the genes *Lg1*-like and *Sad2*-like (Sad).

# 6.2.5 Heading date QTL

From the gene annotation, four clones showed homology to a region closed to the heading date gene suggesting a possible presence of the gene in the biomass QTL regions. A strong correlation exists between the trait biomass and the trait heading date indicated by the fact that late flowering increases the biomass yield. Heading date QTL were already detected on LGs 2, 7, and 4 in *L. perenne* (Armstead et al. 2004). In consequence the phenotypic variation observed for the biomass yield could also result from variation in the time of flowering. To localise the heading date QTL in the studied

LG2

population, heading date data collected in 2006 in the field experiment (Anhalt 2008) were combined to the information from the genetic map presented in chapter 4 using the software MapQTL 6.0 (Van Ooijen 2009). The statistic tests IM and MQM mapping were applied. Heading date QTL were detected on LGs 2, 3 and 5 (Table 20). No heading date QTL was detected on LG7 and the one on LG2 does not co-localise with the biomass QTL. However on LG3 the heading date QTL and biomass QTL overlapped. The difference in the number of days to heading seems to partially explain phenotypic variation of the biomass yield for this QTL.

Table 20: QTL positions identified by IM and MQM mapping for the trait heading date in the field experiment. The QTL are displayed by LOD score and percentage of explained variance

or explained variance							
IM	LG	LOD threshold	LOD score	%expl var.	Flanking markers		
	2	2,2	10	13,8	LoPt356352-sad		
	3	2	9,19	12,8	LoPt556224-rv1131		
	5	2,3	4,02	5,8	LoPt561630- LoPt555245		
MQM	LG	Cofactor	LOD score	%expl var.	Flanking markers		
	2	rv1269	11	12,5	LoPt558654-rv1117		
	3	G04-054	9,66	11	LoPt556100-rv1131		
	5	LoPt 560308	4,7	5,2	LoPt562538- LoPt556854		

# **6.3 Discussion**

Despite of its economic importance for forage species, the trait biomass yield is poorly understood. In the absence of prior knowledge about the genes involved in that trait, the identification of potential candidate genes consists on map-based QTL cloning. To isolate genomic sequences a *L. perenne* BAC library was screened with markers underlying the three highly confident biomass yield QTL. Based on the five genomes equivalent coverage, a total of 130 positive BAC clones was expected. The low level of

positive BAC clones might result from an under representation of clones from the QTL region and a poor hybridization efficiency of the markers used for the screening (Hohmann et al. 2003). The difficulty to detect positive BAC clones concerned essentially the DArT markers. This is probably linked to some technical problems such as DNA concentration or quality inducing a lower hybridization efficiency of this type of markers. On the contrary, several DArT markers induced too much amplification perhaps resulting from the amplification of repetitive elements distributed throughout the L. perenne genome. Large genomes are characterized by many duplications and repetitions and L. perenne is concerned by this phenomenon. The choice of the BAC library is also important to get the best results. It was demonstrated that two BAC libraries constructed from the same genotypes but with different enzymes were not perfectly similar and few clones identified with one were not with the other (Nam et al. 2005). In L. perenne, two BAC libraries were developed from different genotypes and little variations in the results were observed (Farrar et al. 2007). The use of a BAC library developed from a different L. perenne population can explain the non identification of positive clones for all markers because the BAC library might be incomplete or biased. This is especially a limit when markers used as probes for the screening are specific to a population. For example, for the marker rv0674 no clones were identified even after many repetitions under various PCR conditions. The marker did not amplify the positive control consisting on genomic DNA from a third population, while it amplified a positive control consisting of DNA from the F1 of the biomass population. This suggests that rv0674 is specific to some populations. A BAC library from the studied population would have been more representative but after several trials the number of clones and the insert size remained too low to provide sufficient genome coverage. Thus it was decided to use an existing L. perenne BAC

library developed by the University of Arizona. With an insert size above 125 kb, the BAC library was appropriated for the search of candidate genes since fewer clones are required to cover the interval containing the genes of interest (Whisson et al. 2001).

To identify genes responsible for the observed QTL effect, the construction of a BAC contig spanning the entire QTL interval is required. AFLP fingerprinting was successfully used for building physical maps covering entire genomes (Visser et al. 2009, Whisson et al. 2001, Klein et al. 2000). In this study, only half of the digested clones were assigned to a contig. This can be explained by the presence of singletons but also by the varying number of fragments between clones resulting in a too low frequency of common bands to build contigs. But these results tend to confirm the large distance observed on the genetic map between markers, an obstruction to the search of candidate genes. Unfortunately for non-overlapping clones, due to a lack of microsynteny with related species it was not possible to know the physical distance between them. Chromosome walking will be a solution to fill gaps between clones (Kubat 2007, Marra et al. 1997) by screening the BAC library with markers developed from the BAC end sequences until the complete QTL interval is covered by overlapping clones.

The AFLP fingerprinting also revealed errors in the order of markers leading to a possible mispositioning of the markers under the QTL. This was suggested by the construction of one contig for LG2 with clones from two markers that were not neighboured on the genetic map. The analysis of the full BAC sequences for that linkage group indicated the presence of a duplication within the QTL interval. A large contig of 6567 bp from the clone 153A21 showed perfect homology with a contig of the clones 30C14 and 180D4. Difficulty to order markers can arise from inter-chromosomal duplication events.

Furthermore information collected from the AFLP fingerprinting was useful to select appropriate clones for full sequencing. Ideally for the search of potential candidate genes linked to the trait biomass, the entire genomic region covering the QTL should be sequenced. However due to the presence of large gaps this approach was not directly applicable in this study. To complete the analysis two complementary strategies were possible: (1) the development of additional markers specific to the QTL regions to fill gaps between clones or (2) the sequencing of selected clones for gene prediction and annotation. In the previous chapters of this thesis, the difficulty to increase the marker density within the QTL intervals was demonstrated. The mapping of the genes *Sad* and *Lg1* confirms the problem since the genetic map showed a large distance between them while both genes were identified within a single clone indicating a low physical distance between the two genes. The approach consisting of the addition of molecular markers to the three specific regions was not relevant enough and the second approach was preferred to obtain prior information about genes within those regions.

With a haploid genome size of 2034 Mb (Farrar et al. 2007) and a total genetic length of 966 cM observed for this population (Figure 8), for the whole genome, 1cM is equivalent to 2.1 Mb of DNA. If we focus on the QTL detected on LG3, the interval size is on average 20.6 cM or 43.4 Mb. Considering that 38 genes were identified within a region of 148 Kb, a potential of 293 genes might be identified only for that QTL. Several studies showed that the ratio between genetic and physical distance differs across the genome and tend to be smaller around genes (Hohmann et al. 2003, Ballvora et al. 2002). And the program RiceGAAS revealed considerable variations in the number of genes identified in function of the predicted gene program used suggesting a lack of robustness of the *ab initio* method with a possible identification of false

positives. In addition, transposable element sequences can be annotated as genes (Bossolini et al. 2007). But despite the hypothesis of an overestimation of the gene content, the number of genes remained too high for an analysis of each gene and a better fine map will be required. The screening of a larger population of thousands of genotypes or the construction of a Near Isogenic Lines (NILs) population will eventually induce the elaboration of the fine map.

The search of candidate genes was focussed on a restricted region and only few clones were sequenced. The clones were selected to be exactly underneath the LOD peak of the biomass QTL suggesting a strong link between the genomic region and the phenotypic variation. The rice genome has been fully sequenced and annotated. Assuming the macrosynteny observed previously in this study between rice and L. perenne, a microcollinearity conservation was expected and the BAC sequences were firstly blasted to the rice genome to identify genes among the sequences. But only a maximum of three contigs per BAC sequence showed synteny. At the macrolevel, a good conservation among the genomes was observed within the grass family, however significant rearrangements were also noticed (Bancroft 2001). A comparison has been extended at the individual gene level and the results indicate that the microstructure is not as well conserved as the macrostructure (Bossolini et al. 2007, Liu et al. 2006, Bancroft 2001). Genes can be conserved between grass species or present in only one. Frequent interruptions in the microcollinearity might arise from the conservation of the coding sequences but not of the introns of the genes or from the presence of many transposable elements (Bossolini et al. 2007). Thus between barley and rice, the orthologous Fr1/Hd6 region is conserved but the distance between genes was higher in barley due to the presence of different transposable elements within the intergenic

region (Dubcovsky et al. 2001). The gene annotation revealed that the five *L. perenne* BAC sequences are rich in transposable elements. Hence at least one predicted gene per BAC showed homology to a transposable element. During grass evolution many deletions, insertions, duplications or translocations have occurred limiting the use of synteny to identify genes although the synteny with rice was successfully used to assist *in silico* identification of the candidate gene for heading date in *L. perenne* (Armstead et al. 2005). In the absence of microcollinearity conservation, the gene prediction had to be performed with the *ab initio* approach followed by a BLASTn and a BLASTp. This method was relevant and permitted the identification of five interesting genes affecting the trait biomass. A function was assigned to only a part of the predicted genes due to a lack of annotated genome resources and it is likely that other genes have an effect on the phenotypic variation within the three studied regions.

Of all the predicted genes, five were of particular interest since they were homologs to a region closed to known genes. Although they did not show homology to the genes themselves, the fact that clones from different genomic regions are homolog to the same regions suggests that the genes may underlie the biomass QTL. Maybe the sequences flanking the studied clones would show homology to the genes. The five genes are interesting since their functions are known to induce variation in the biomass yield of a plant. The gene *lg1* controls the formation of the ligule, a component of the leaf development (Rossini et al. 2006), the genes *Sad1* and *Sad2* regulate the disease resistance compounds (Wegel et al. 2009, Qi et al. 2006) and the cold-resistance gene regulates the freezing resistance, particularly important during early spring and autumn since a low temperature stress inhibits the plant development (Wilkins and Humphreys 2003). Finally the gene *Hd1* controls the number of days to heading and a late flowering

is usually associated to a higher vegetative production (Armstead et al. 2005). While the gene lg1 was identified only on LG2, the three others were detected under the QTL on LGs 2 and 3, which indicates their active role in the biomass production. An analysis of BAC clones under the QTL on LG7 would permit to see whether these genes are systematically associated to the biomass yield QTL. The fact that similar genes were observed in two chromosomal regions associated to the same traits suggested a possible duplication event. The difference of exons number and genes size are an indication that the region was probably not duplicated.

The *Hd1* gene was identified on rice chromosome 6 (Yano et al. 2000), homologous to *L. perenne* chromosome 7, and *Hd1* in *L. perenne* was identified on LG7 (Armstead et al. 2005). The gene *Hd1* was associated to a major heading date QTL in rice (Yano et al. 2000) and in *L. perenne*, the gene was found closed to a heading date QTL (Armstead et al. 2005). In the current study, the putative presence of the gene within at least two of the three QTL regions suggested that the biomass phenotypic variation observed might result directly from variation in heading date. A QTL analysis of this last trait indicated that the heterosis observed on LG3 is mainly due to the heading date effect since QTL for both traits co-localised. This is not surprising considering the strong correlation existing between the two traits. The five identified genes are valuable information for future analyses but remain only potential candidate genes. Several tests will need to be applied to evaluate the effect of these genes and see if they explain the observed phenotypic variation between the genotypes of the F2 biomass population.

The work presented in this chapter consisted on the study of the genomic sequences underlying the three biomass yield QTL previously detected. In the absence

of knowledge about genes involved in the trait biomass yield, a map-based QTL cloning was applied. The large size of the QTL intervals was confirmed by the AFLP fingerprinting analysis and further work had to be focussed on a restricted region below the LOD peak of the QTL. The genomic region underneath the QTL revealed a high density of potential candidate genes but their function remained mainly unknown. A function could be assigned to four genes that might partially explain the heterosis observed at the QTL chromosomal region but there effect on the trait biomass yield need to be confirm. But the number of genes within the QTL interval remained too high for an analysis of each genes and a better fine map of the QTL intervals is required to enable the isolation and cloning of a biomass yield candidate gene.

# **Chapter 7: General discussion**

A genetic linkage map developed from an inbred derived F2 population (Anhalt et al. 2008) was refined using several types of molecular markers. The objective was to improve the precision of the position of biomass yield QTL previously identified (Anhalt et al. 2009) in this population and enable a search for potential candidate genes linked to the trait biomass yield.

#### 7.1 Fine mapping

The fine mapping work focused on the chromosomal regions where the QTL were previously localised (Anhalt et al. 2009). Several strategies for saturating a genetic map exist. For the present thesis, it was decided to add more molecular markers to the map to provide sufficient information to refine the QTL intervals and for a further application in breeding programs. SSR markers are codominant inherited and thus provide robust informative data for the QTL analysis. Ryegrass specific SSR markers were available from different sources (Gill et al. 2006, Van Daele et al. 2008, King et al. 2008, Studer et al. 2010). In addition STS markers were developed from rice to provide more molecular markers and to reveal the synteny between ryegrass and rice. Synteny being a powerful approach to pinpoint genes underlying QTL. The first part of the project was to improve the genome coverage using these two types of markers. However, due to a lack of polymorphism only few SSR and STS markers could be added to the map.

Hence a second strategy based on the mapping of a new type of markers, called DArT markers, was applied. DArT markers have the disadvantage of being dominant complicating the estimation of recombination necessary to order the markers on the linkage map. And dominant markers present some limitation in the detection of QTL. However the DArT technology is straightforward to generate a large amount of data and 134 DArT markers were added to the map in contrast to only 18 SSR and one rice STS markers. The DArT markers allowed homogeneous genome coverage and allowed the large gaps to be partially filled. Remaining gaps flanked all the codominant markers demonstrating the difficulty to generate a map combining a high amount of dominant markers with few codominant markers. Furthermore this study showed that an increase in the marker number results in a higher risk of errors in the order of markers and requires heavy computational algorithms. The maximum likelihood algorithm applied in this study was the most accurate to generate a map in a F2 population with so many dominant markers. However it seemed that it induced a map expansion with the current map spanning a length of 544.4 cM compared to 289.2 cM previously (Anhalt et al. 2008). This could explain partially the gaps surrounding the codominant markers. Information on the order of markers provided by the map may be useful in the case of an application of the DArT markers to the genotyping of L. perenne association mapping population.

#### 7.2 QTL analysis

The map after addition of the SSR and STS markers and the map after addition of the DArT markers were used to perform the QTL analysis. It allowed a comparison of the number and position of the detected QTL to evaluate the potential of the DArT markers to identify QTL. Although the efficiency of the DArT markers to saturate a map was already demonstrated (Wenzl et al. 2006, Peleg et al. 2008, Hearnden et al. 2007), little QTL analyses using the DArT markers have been reported so far. Our study showed that the detection of QTL using DArT markers is limited. All QTL identified were systematically located in a chromosomal region where SSR markers were mapped. Removing these SSR markers from the map resulted in a loss of the QTL.

In the present study several QTL mapping models were applied to increase the accuracy of the analysis. Overall the fine mapping work confirmed the biomass yield QTL previously observed on LGs 2, 3 and 7 (Anhalt et al. 2009) which represent a starting point for the study of the trait biomass yield. Three additional QTL were detected on the same linkage groups but due to inconsistency in their detection, they were not further analyzed. QTL for the four traits measured (fresh weight, dry weight, dry matter and leaf width) overlapped and showed a higher effect in the field than in the greenhouse. QTL detected in the field were not systematically identified in the greenhouse confirming the interaction between the genotypes and the environment previously observed (Anhalt et al. 2009).

Despite the addition of many markers, the QTL intervals remained large with a distance between the flanking markers of around 20-30 cM. Thus each QTL region included hundred of genes (Salvi and Tuberose 2005, Mott 2006) which complicated later the use of the map-based cloning approach to isolate candidate genes linked to the trait biomass yield. After a confirmation of the linkage between these markers and the QTL, the markers may be used in a marker-assisted backcrossing strategy to transfer the QTL to elite breeding material.

#### 7.3 Description of the genomic sequences underlying the biomass QTL

The ultimate goal of a QTL analysis is to isolate the genes linked to the quantitative traits. The trait biomass yield is poorly studied so far and the genes or biochemical and signalling pathways involved are not known yet. A map-based QTL cloning approach was the most appropriate to dissect this complex trait in the absence

of prior knowledge. However, despite the fine mapping approach the QTL intervals remained too large to enable the isolation of candidate genes. But a description of the genomic sequences underlying the QTL intervals is useful to understand the genome structure and to organize the future work leading to the cloning of the genes of interest. The screening of a L. perenne BAC library permitted the isolation of genomic sequences underlying the three OTL. The number of positive clones identified was less than expected resulting perhaps from an under representation of the QTL regions in the BAC library, but more likely due to the use of a BAC library developed from a different population. Indeed the screening revealed the specificity of few markers to the studied population. Thus despite of their strategic position within the QTL regions, these markers are not transferable to another population and thus cannot be used in marker assisted selection. The BAC clones were initially analyzed using the AFLP fingerprinting method to study the physical distances between the markers within the QTL regions. The impossibility of forming contigs with the BAC clones confirmed the large distances observed on the genetic map between SSR and DArT markers was not just a consequence of the algorithms used. The AFLP fingerprinting also revealed that errors in the order of markers on the genetic linkage map remained.

Two regions underneath the QTL on LGs 2 and 3, of around 320 Kb and 170 Kb respectively, were completely sequenced to enable a gene prediction and annotation. The mapping of rice STS markers onto the *L. perenne* map indicated a high degree of conservation of large chromosome segments between the two species and suggested a good synteny at the DNA level. But the BAC clone sequences revealed a poor microsynteny between the two species. Although the species are quite similar at the macrolevel, there are too many rearrangements at the microlevel to enable a systematic identification of a gene in *L. perenne* using the sequence of the gene identified in rice.

Thus the best approach to identify genes present in the studied genomic sequences was the *ab initio* approach. It revealed a high density of potential candidate genes within a small genomic region confirming a too large amount of genes within the entire QTL interval to allow an identification and characterization of each gene. Of all the predicted genes identified one showed homology to the *Hordeum vulgare* partial *lg1* gene for liguleless-like protein (Rossini et al. 2006) and four to regions closed to the *L. perenne* heading date (*Hd1*) gene (Skøt et al. 2007), the gene cluster *Avena strigosa* beta-amyrin synthase (*Sad1*) and cytochrome P450 CYP51H10 (*Sad2*) (Qi et al. 2006) and the *Lolium multiflorum* gene for cold responsive protein (Oishi et al. 2010) and. These five genes are of special interest to this study due to their possible implication in the variation of the biomass yield observed between the genotypes. These results indicated that the strategy of map-based QTL cloning followed by an *ab initio* approach was adapted to identify potential candidate genes involved in the trait biomass yield. However it also showed the importance of fine mapping to reduce the size of the region to be analyzed.

#### 7.4 Further work

Despite a lack of common markers, the alignment of several *L. perenne* genetic maps allowed a comparison of the positioning of QTL regions for different traits. The Figure 19 to 23 show that some QTL are overlapping the biomass yield QTL and confirms that the three regions identified in this study are genes rich regions controlling many traits. The use in breeding programs of the markers flanking the biomass QTL may induce the selection of undesirable traits and the cloning of the genes linked to the biomass is necessary to develop markers for an application in MAS. But the presence of many genes will be a problem to isolate the genes of interest and further work is needed

to reach a sufficient saturation of the linkage map to obtain a QTL interval size around 1 cM or less. The construction of another type of mapping population will be time consuming and difficult to carry on. However since seeds from the F2 population are available the size of the population can easily be increased to a thousands of individuals. Also the map needs to be saturated with more codominant markers. Many SSR and rice STS markers were removed due to a too high level of segregation distortion but it seemed that the distortion resulted from scoring errors. Perhaps it would be useful to screen the entire population again with the removed markers because they were polymorphic. By sequencing the PCR products the segregation for each genotype could be determined. Using the sequencing technique, scoring errors should be avoided. The BAC clone sequences can also be used to develop SSR markers specific to the QTL and thus induce a better precision of the QTL intervals. In addition the gene annotation revealed the presence of similar genes under the QTL on LGs 2 and 3 and it will be interesting to sequence clones under the QTL on LG7 to see whether they are also present under that QTL. Finally the gene prediction and annotation strategy used was adapted to this study and once the QTL intervals are reduced to a few cM, a similar approach should be applied to reveal how chromosomal rearrangements contribute to the expression of strong phenotypes like heterosis for biomass yield.



Figure 19: Alignment of LG2 of several *L. perenne* maps developed to study QTL. In blue are the markers mapped in the F2 biomass population and in red the markers in common between other populations. A) F2 biomass map with the DArT markers with the biomass QTL identified with MQM mapping in the field experiment, B) F2 biomass map without the DArT markers with the biomass QTL identified with MQM mapping in the field experiment C) map from Turner et al. (2008) and D) map from Cogan et al. (2005)


Figure 20: Alignment of LG2 of several *L. perenne* maps developed to study QTL. In blue are the markers mapped in the F2 biomass population and in green the markers mapped by Barre et al. (2010). A) F2 biomass map with the DArT markers with the biomass QTL identified with MQM mapping in the field experiment, C) map from Turner et al. (2008) and D) map from Barre et al. (2010)



Figure 21: Alignment of LG3 of several *L. perenne* maps developed to study QTL. In blue are the markers mapped in the F2 biomass population and in red the markers in common between other populations. A) F2 biomass map with the DArT markers with the biomass QTL identified with MQM mapping in the field experiment, B) F2 biomass map without the DArT markers with the biomass QTL identified with MQM mapping in the field experiment C) map from Turner et al. (2008) and D) map from Cogan et al. (2005)



Figure 22: Alignment of LG7 of several *L. perenne* maps developed to study QTL. In blue are the markers mapped in the F2 biomass population and in red the markers in common between other populations. A) F2 biomass map with the DArT markers with the biomass QTL identified with MQM mapping in the field experiment, B) F2 biomass map without the DArT markers with the biomass QTL identified with MQM mapping in the field experiment C) map from Turner et al. (2008), D) map from Cogan et al. (2005) and E) map from Barre et al. (2010)

## ANNEXES

## Annex I: List of the STS markers designed from rice sequences obtained from the rice genome annotation project webpage.

Name	LG	Forward primer seq. 5' to 3'	Reverse primer seq. 5' to 3'
LOC_Os04g54870	2	GAGTTCTTCTCATTCTCTGG	GCATAGGGTTCCTTGAAAGTC
LOC_Os04g54940	2	ATCCTCAGCTGGGTCAATG	GAATCTTGTAGTTCTGGATCATGTC
LOC_Os04g55030	2	GACGCCATCCGCGAGATTTA	CATCTCACCCATCCTCTCTC
LOC_Os04g55050	2	CTGGATGTTGTGGGATGCTT	CCCTCTCCAAGAGTTTCTTG
LOC_Os04g55060	2	GTGGATTGGACAGACAGAGG	CTAGCAAGCCATCTTCATCAG
LOC_Os04g55150	2	ACTGTTTCTTCTCCAGTGCC	GCTTCAATGAATCTCTCAGCC
LOC_Os04g55180	2	CTGTTTCTTCTCCAGTGCCT	GCTTCAATGAATCTCTCAGCC
LOC_Os04g55220	2	GAAGACTCAATAACATGGCC	GCTTGACTTCCAATGTACC
LOC_Os04g55260	2	CAGACTTGGGCATCAGATGA	CCAAACATGTCACCTAGAAG
LoC-Os07g12490	2	AGAGGGAAGGGTTCTATC	GCCAGGTTTTCCTTTCAG
LoC-Os07g12520	2	AAGGCGTCCACCAACAAT	CAGAGCCACCCTTTCTAT
LoC-Os07g12530	2	TTTTTGAGCGCATGGGTG	TGGTGCAGAAGCTACGAA
LoC-Os07g12610	2	TATTTTACGGAAGCCCGC	TCCACCAAGAGCTTATCG
LoC-Os07g12640	2	CCGCCAACCGTTTATTGT	TTGGATGCCACCACCATA
LoC-Os07g25004	2	AATGGACTCCTGACGTCT	GCTGAAGCAGCAACTTTC
LoC-Os07g25390	2	CTGGCAACCATCTATCCA	CTATCTTCTCTGCATGGC
LoC-Os07g25270	2	CCAAAGAAGGGTTTCCAA	TGATGTTCTGCCCTTGAC
LoC-Os07g25016	2	AATGGGAAATGCCCTACC	TGTTATGTATGGCCGTGC
LoC-Os07g31850	2	CGATGCGGTTTTGGAACT	CTCCACGATCTAGAGCAT
LoC-Os07g35260	2	AGACTCCTTGGTGTTTGC	CCAATCCAGCACATTGCT
LoC-Os07g38880	2	TGTGGGTATTTGGCAATG	CAGTTTGTATCGTCGCTG
LoC-Os07g43330	2	AAGGGAACCCTTGAGGAA	ACGTAATGTCATTTCCCG
LoC-Os07g43360	2	AGGTAAAGTAGGGACACC	AGCTTGCTAACGTCGACC
LoC-Os07g43420	2	CCGAACAGATTGATGTGG	TGTCATCTGCGTACGGCA
LoC-Os07g43470	2	AACAAGACACCCCCCAAT	CTGTTGCAGAAGTCTTCC
LOC_Os01g09560	3	GCTAGCACAAGTGTTGCACA	TCCTTGACTTCCCAATCAAG
LOC_Os01g09570a	3	AGTGGGCGATGAAGAAGAAG	GCACTGGTCTTATGAACAAC
LOC_Os01g09570b	3	AGTGCAGAGAATGGCATTGG	CAAAGGCCAACATCCAGAAG
LOC_Os01g10400	3	CATAGTTGAAGCCGAGAAGAG	TCACTTGCCGCACTTGCAGT
LOC_Os01g11370	3	GGTCTAACCTTGTTTCCTGC	GGAGAAGCTGGCCAATCAAA
LOC_Os01g11650	3	TGCACACCCAGATCAAGTG	AGCGGCGAGTTGTAGTCGTT
LOC_Os01g11710	3	GACAAGCTCTGGACCAACCT	TTGCCCTTGAAGAAGGCGTA
LOC_Os01g11940	3	GACATGCGCACCTTCTACA	TTGTAGAGCTCGGCGAAGTC
LOC_Os01g12470	3	ATTGAGTTGGCAAGAACACG	TTCCACATTCCAGGGGGTTT
LOC_Os01g13160	3	CTCGAGGAGATCTGCGAGCT	GCAAAGTTGCTCAAACCCTG
LOC_Os01g13770	3	CATGCTCTTGGTCATGTCAC	GCTGCATTGAAGAAGGGCTC
LOC_Os01g14550	3	GACAAGTACCGGTCGTTC	CTCCATCTCCCAGGTCTT
LOC_Os01g14580	3	GGTGATCAATACCGAGCAAC	GCTACTCCACCAGCACCA
LOC_Os01g15600	3	TTCCTCGACCTCTTCGACTA	AATGCTTCCCAACCTTCTTC
LOC_Os01g15850	3	GCATATGCTTGGAAGCGTCA	AGTGAAGGCCTTGAGAGCTA

LOC\_Os01g16100 3 LOC\_OS01g16152 3 3 LOC\_Os01g16540 LOC\_Os01g16900 3 LOC\_Os01g17020 3 3 LOC\_Os01g18070 LOC\_Os01g18320 3 LOC\_Os01g24750 3 3 LOC\_Os01g27520 LOC\_Os01g34330 3 LOC\_Os01g34840 3 LOC\_Os01g36890 3 LOC\_Os01g36920 3 LOC\_Os01g37480a 3 LOC\_Os01g37480b 3 LOC\_Os01g37770 3 LOC\_Os01g40840 3 3 LOC\_Os01g43160a LOC\_Os01g43160b 3 LOC\_Os01g43360 3 3 LOC\_Os01g45750 LOC\_Os01g46980 3 LOC\_Os01g51620 3 3 LOC\_Os01g67054 LOC\_Os01g67850 3 LOC\_Os01g68260 3 LOC\_Os01g68324 3 LOC\_Os01g53000 3 LOC\_Os01g52500 3 LOC\_Os01g53520 3 LOC\_Os01g52770 3 LOC\_Os01g53150 3 LOC\_Os01g52110 3 LOC\_Os01g53810 3 LOC\_Os01g53990 3 LOC\_Os01g54010 3 LOC\_Os01g54300 3 LOC\_Os01g54510 3 LOC\_Os01g54860 3 LOC\_Os01g54540 3 LOC\_Os01g54550 3 LOC\_Os01g54810 3 LOC\_Os01g54920 3 LOC\_Os01g55030 3 LOC\_Os01g55360 3 LOC\_Os01g55540 3

CACGCCGAGTACCTCAAC CTGAATGGATGGGCAGAAAA AAGAGGGTTATGGAGCGAGT CGAGTCGGTCAAGCAGGTA GCATTATCACCCTGTGCAAC CATCGCACACTTATGAACATC CATATGTTGCTAGTGATATC CCCGTCAAGATCTGCGGTGA GGTGGAAGCCAAAGGGAGGA CTGATGAACAAGGAGCTTGA ATTGTGATGATTCCTGG CAAGTGATGATGTTTTCAGC CAAGTGATGATGTTTTCAGC CGGATATCGTTCTACGCCTC TGTAGAGCAGGCACATGCAC GAAAAGCTTGAGCAAGAGG TAGCAATGATGGAGAGCCTT TGGTGGGACATGTGGAAGAA AATGTGGTCTTCCAGGATTC TGCAAGGGCGACTACATCGG GTGGGATTTCTTGCGCAGTA CAACTTGTTGAAGCTGAG AACAATACACACCCTTTCC GAGCAGGATATTGAATGTGG TACAGCTCCATGGCTCCAAT AGGTCTGACATGGCAAATGA AAGGCAGAAATCGGAAT GGACTAGGATTTGGATTTAGA ATCTTCCCAGGGTTTGGCCT AAGGCCAAGATCTACCGCCT GACAACGGCAACAACAAC TCCAACGTCGGCAAGTCGT TCTGTGAAGTGTGCAAGCTC ACAGCAGCTATGCTCGAG GCAGGACACGCTGTACGA CATAGCTTGTATGATGACTT CTCCAACGATTCCTTCTTCT AAAGCAGCATTCCTCATCCC GATGCCAAGCTGGTCATCTT AAGCAGAAGAGGCAGCCGCT CCCCAAGTACTTCAAGCACA TACGACTTCCACGTGTCCGG TATGAAGCTTCAGGAGCAAGG AAGGTGCTCAGCTCCAAGAC TATGCCTGGACGCACATGAT TGGTGAATTGCTCATGGCTC

CGACTTGCTTTCACAAGCCT CCAAATCCAAGCTTTTGTGG AGCCATATGTTGCATTCA GAACGGTGGCCTTTTCCAAT ATCACAGTGATGACATCC GTATATTGGTCCCAGATCC ACAGTTACAGCAGCAACTGG TCCAAGCTTTGTTTGCCTCT ACATCTCTGATGTTCTCCA CGATCTCCATGACCAGCATG TCACCTTCAGAGCAGTCAAT CCATGAAGGGTCAGTTTAGC TGAAGGGTCAGTTTAGCCTC TGCCTGCTCTACAAGGGA AGACCCCAGATAACCAAAAG TAGACCATAGCCATGCC AACAGCACCTTCAGGCCAAT CCACTCTTGATGGCGATCAG TCCCAGCCACTCTTGATG CGCCTTTTCATCTTTCCTGA AGACCAGTTGCAAGAGGAGC TCAGCTGCATAGAGTACTCA GAACAAATGCAGCTTCTCTC GATGTCCCAGTCAGTGTACA GGAAGGACCATCTTGGTGCA GCATCCCTAGAGAACACAGT ACCTGGTGAGGTTTAAATGT TCTGAGCACTGGTTATT CTGATGTTTGAGAAGGGAGG GAGATCTTCTTGTTCCTCCA GGGATGAAGTTGTCGAA CTAGCATAGTCAAGATCAATC GCTCCTTCAACACATGCATG AACTCTTCTGCATCCCAACT CCCGTCACCCGGCAGTTGA TGCTGTAGTTGAAGGGTAGG AGCTCCCACGCGAGGATGGT GTCATGCCTATGACATCTGG CCTTCATAGAACATCTTTAG CTGTTGCATCAACAACATACC GATGAATGGCTGAATATAGG GCAGGCCATTACCATCCTTT CGTTTCTTTGCAGCATC ACAGCAAGCCTCATTGA GCACCAATAAAACCTTCCCA CATAGGCCTAATTACAAGCT

LOC\_Os01g55700 3 LOC\_Os01g55780 3 3 LOC\_Os01g55940 LOC\_Os01g16540a 3 3 LOC\_Os01g16100a 3 LOC\_Os01g16100b LOC\_Os01g16900a 3 3 LOC\_Os01g16900b 3 LOC\_Os01g16900c 3 LOC\_Os01g16900d LOC\_Os01g18070a 3 LOC\_Os01g18070b 3 3 LOC\_Os01g18320a LOC\_Os01g18320b 3 LOC\_Os01g18320c 3 7 LOC-Os06g01170 LOC\_Os06g01390 7 7 LOC\_Os06g09280 LOC\_Os06g08390 7 LOC\_Os06g08720 7 7 LOC\_Os06g08840 LOC\_Os06g01360 7 LOC\_Os06g08850 7 7 LOC\_Os06g16350 LOC\_Os06g16390 7 LOC\_Os06g16420 7 7 LOC\_Os06g17285 LOC\_Os06g28740 7 LOC\_Os06g29020 7 7 LOC\_Os06g29120 LOC\_Os06g29180 7 7 LOC\_Os06g15990 7 LOC\_Os06g30380 LOC\_Os06g47550 7 7 LOC\_Os06g48160 LOC\_Os06g48750 7 LOC\_Os06g49440 7 7 LOC\_Os06g49460 7 LOC\_Os06g49530 LOC\_Os06g49670 7

AACTTCTTCCGGATCTAC TATAACATTGCTCACCGTGG AGAAGCTGGAGTTCATCGAG TGCTAGACTACACTTTCATG ACTCGACTATGTTTTAGGGG GAAGGGCCTATACCTTTG TAGACCTTGCAACAGCCT TGCGATACACAAGGCCAA GAGGAGGTCATCATTGAC GATTGGAGGCTTGGGCTT TTGGGGATAACGTCACAG ACAGAACACAAGGATCCC TGCTCAGGTGTGTATGCT AAACAATCCAGGAGAGGG GTAGTCAGGGAGTTGAGA TGCTTGATGGTACTGAAGTA AGCTTCCACTCATGCAGT ACTTATCAGCCTGAGTC GACCAAAGCTTCTGTCAA GCCAGGTTGCTCATCAAT TTGATGACAGGAGGGATGCT GACGGCGATTTCCTCATTGT CTTCTATTGCCTTTGGTG AGCCTGGACCTAGCAAACA AGACCAATCTGTACTTGGTC TGGGAACTTTTCACAGCAGT TCCGGGATCCCAAAATACCA GAAGCTGCTGCTACATTT AGGAAGGCAAAGAGCAGGC AGCTCTTGTCTACCTACACT AGAAGATGAGCTCATGCGCA TGGAAAGTTGGCCCTGCTT CATTTTGGCATCAGACTTC TGTTGCTGCAGCTATGGAGA TCGACATGCAGATCAAGCTC GTCCTTGATGAAGCTGATGA GTTGCATGCATTGATTACAG GACGAGTTCAAGCCCTTCAT TATACTCAGATGCAGGCT GGCCATTTTCGTGATGCC

AAAAAGGCATCCACTCCTGG CCCATCTTTGCTAGCAAGGA ATCGTCGGCATTAGCTTCCT TCGTACTAATCTCTGGTC ATCAGGCTCTCCTGACAAGT GCCAGTCTTGCAAAATCT AAGCTTTCCACAACACCC TTCTCAGCCACTTCACGA AGACCAGGATTCTTCACA CCTTCACCAATGTTGCA GTCCTCAACAGTACATCC CAGGAATTTGGCAGGGTT TGCAGCCTTCATACTGAG GCAACTGTCTGCCCCTTT CTGGAGCACGATTTGGAA CAGTAGATTCTTTGATCC TTGGGATGCGCACATGGTCA AGCACGCTCCTGAATAAG CAGCCATATCATTGACACC GGTGAATCAAGGAACAAG ACGGCATTCACGAGCAAAGT CCAAGATCAGGGAGCTGAAA CCAATCAAATGCGCAAAC ATGGCGCTGATGTTGTCGT GAATCCAATGCTGCATCCA GTGATATAGACAGTTGAGCA TTGCCTTGAATCACTCGCGA AGGATACTTTGGGCATTC CTGTAGTCAGTTTCTGGTCG ACTCCAAAGTCTCCTAGCTT CACCTTTCTTCATCTTTGCA ACATCCATGTGACTAGCAAG CAAAATCCAGGCCTTTCGAG ACAACTCCTCTACCAGGTA CAGGAACTCGAAGTCGAT TCTTCCTTCTCCACATTGAC AGGATCACCTCCATAACT TGCCGCTTCATTGTCAG AGCATCACCAATCTGGGTGT CTCAATTCCTTCTGCCTCCA

Annex II: List of annealing temperatures for the ryegrass specific SSR markers mapped on the F2 biomass population.

Source	Name	Linkage	Annealing
ABERS, UK (King et al. 2008)	LpHCA17C11	<u>group</u> 2	60
	rv0347	2	58
	rv0122	2	60
	rv1282	2	58
	rv0959	2	60
	rv0037	2	57
	rv1212	2	58
ViaLactia, New Zealand (Gill et	rv1068	2	60
al. 2006)	rv1144	3	62
	rv0433	3	60
	rv0474	7	58
	rv0459	7	58
	rv1175	7	60
	rv1060	3	62
	rv1316	3	64
Studen et al. (2006)	G01-039	2	58
Studer et al. (2006)	G01-010	7	58

Annex III: LOD profiles of the biomass QTL obtained on LGs 2, 3 and 7 with the position of the markers along the chromosome. The linkage map after addition of the SSR and the STS markers was used. QTL were calculated for the field experiment (F) and the greenhouse experiment (G) with IM (a), MQM (b) and CIM (c). The horizontal dashed line indicates the LOD threshold value at 95%. The beige line for IM and MQM corresponds to the % of explained variance.



Figure 1a: LOD profiles of the fresh weight biomass QTL detected on LG 2.



Figure 1b: LOD profiles of the fresh weight biomass QTL detected on LG 3



Figure 1c: LOD profiles of the fresh weight biomass QTL detected on LG7



Figure 2a: LOD profiles of the dry weight biomass QTL detected on LG 2



Figure 2b: LOD profiles of the dry weight biomass QTL detected on LG 3



Figure 2c: LOD profiles of the dry weight biomass QTL detected on LG 7



Figure 3a: LOD profiles of the dry matter biomass QTL detected on LG 2 in the field experiment



Figure 3b: LOD profiles of the dry matter biomass QTL detected on LG 3



Figure 3c: LOD profiles of the dry matter biomass QTL detected on LG 7 in the field experiment



Figure 4a: LOD profiles of the leaf width biomass QTL obtained on LG 2



Figure 4b: LOD profiles of the leaf width biomass QTL obtained on LG 3



Figure 4c: LOD profiles of the leaf width biomass QTL obtained on LG 4

Annex IV: LOD profiles of the biomass QTL obtained on LGs 2, 3 and 7 for the map generated with the DArT markers with the position of the markers along the chromosome. The names of some markers are omitted for clarity. QTL were calculated for the field experiment (F) and the greenhouse experiment (G) with IM (a), MQM (b) and CIM (c). The horizontal dashed line indicates the LOD threshold value at 95%. The beige line for IM and MQM corresponds to the % of explained variance.



Figure 5a: LOD profiles of the fresh weight biomass QTL detected on LG 2



Figure 5b: LOD profiles of the fresh weight biomass QTL detected on LG 3



Figure 5c: LOD profiles of the fresh weight biomass QTL detected on LG 7



Figure 6a: LOD profiles of the dry weight biomass QTL detected on LG 2



Figure 6b: LOD profiles of the dry weight biomass QTL detected on LG 3



Figure 6c: LOD profiles of the dry weight biomass QTL detected on LG 7



Figure 7a: LOD profiles of the dry matter biomass QTL detected on LG 2



Figure 7b: LOD profiles of the dry matter biomass QTL detected on LG 7



Figure 8a: LOD profiles of the leaf width QTL detected on LGs 2 and 3



Figure 8b: LOD profiles of the leaf width QTL detected on LG 4

## REFERENCES

- Ahn S and Tanksley D (1993). Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA. 90: 7980-7984
- Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z and Rognli OA (2003). A linkage map of meadow fescue (*Festuca pratensis* Huds.) and comparative mapping with other Poaceae species. Theor Appl Genet 108: 25-40
- Anhalt UCM (2008). Characterisation of the initial generations of recombinant inbred lines in perennial ryegrass (*Lolium perenne* L.) using molecular markers and cytogenetics. PhD thesis. University of Leicester, England.
- Anhalt UCM, Heslop-Harisson P (JS), Byrne S, Guillard A, Barth S (2008). Segregation distortion in *Lolium*: evidence for genetic effects. Theor Appl Genet 117: 297-306
- Anhalt UCM, Heslop-Harisson P (JS), Piepho HP, Byrne SL, Barth S (2009). Quantitative trait loci mapping for biomass yield traits in *Lolium* inbred line derived F2 population. Euphytica 170: 99-107
- Armstead IP, Turner LB, King IP, Cairns AJ and Humphreys MO (2002). Comparison and integration of genetic maps generated from F2 and BC1-type mapping populations in perennial ryegrass. Plant breeding 121: 501-507
- Armstead IP, Turner LB, Farrell M, Skøt L, Gomez P, Montoya T, Donnison IS, King IP and Humphreys MO (2004). Synteny between heading-date QTL in perennial ryegrass (*Lolium perenne* L.) and the *Hd3* heading-date locus in rice. Theor Appl Genet 108: 822-828
- Armstead IP, Skøt L, Turner LB, Skøt K, Donnison IS, Humphreys MO and King IP (2005). Identification of perennial ryegrass (*Lolium perenne* (L.)) and meadow fescue (*Festuca pratensis* (Huds.)) candidate orthologous sequences to the rice *Hd1* (*Se1*) and barley *HvCO1 CONSTANS*-like genes through comparative mapping and microsynteny. New Phytologist (2005) 167: 239-247
- Armstead IP, Turner LB, Marshall AH, Humphreys MO, King IP and Thorogood D (2008). Identifying genetic components controlling fertility in the outcrossing grass species perennial ryegrass (*Lolium perenne*) by quantitative trait loci analysis and comparative genetics. New Phytologist 178: 559-571
- Asins MJ (2002). Present and future of quantitative trait locus analysis in plant breeding. Plant Breeding 121 (4): 281-291
- Ballvora A, Ercolano MR, Weiss J, Meksem K, Bormann CA, Oberhagemann P, Salamini F, Gebhardt C (2002) The R1 gene for potato resistance to late blight (*Phytophtora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. Plant J 30:361–371

- Bancroft I (2001). Duplicate and diverge: the evolution of plant genome microstructure. Trends in Genetics 17: 89-93
- Barre P, Moreau L, Mi F, Turner L, Gastal F, Julier B Ghesquière M (2009). Quantitative trait loci for leaf length in perennial ryegrass (*Lolium perenne* L.). Grass and Forage Science, 64, 310-321
- Barth S, Melchinger AE, Devezi-Savula B and Lübberstedt T (2001). Influence of genetic background and heterozygosity on meiotic recombination in *Arabidopsis thaliana*. Genome 44: 971-978
- Bartoš J, Sandve SR, Kölliker R, Kopecký D, Christelová P, Stočes S, Ostrem L, Larsen A, Kilian A, Rognli OA and Doležel J (2011). Genetic mapping of DArT markers in the *Festuca-Lolium* complex and their use in freezing tolerance association analysis. Theor Appl Genet 122, 1133-1147
- Bennetzen JL and Freeling M (2008) The unified grass genome: synergy in synteny. Genome research 7: 301-306
- Bennetzen JL (2000). Comparative Sequence Analysis of Plant Nuclear Genomes: Microcolinearity and Its Many Exceptions. The plant cell 12: 1021-1029
- Bert PF, Charmet G, Sourdille P, Hayward MD, Balfourier (1999). A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers. Theor Appl Genet 99: 445-452
- Blears MJ, De Grandis SA, Lee H and Trevors JT (1998). Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. Journal of Industrial Microbiology & Biotechnology 21: 99–114
- Bonierbale MW, Plaisted RL and Tanksley SD (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120: 1095–1103.
- Bonnett DG, Rebetzke GJ and Spielmeyer W (2005). Strategies for efficient implementation of molecular markers in wheat breeding. Molecular Breeding 15: 75–85
- Bossolini E, Wicker T, Knobel PA and Keller B (2007). Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. The Plant Journal 49: 704–717
- Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. Am J Hum Genet 32:314-331
- Broman KW (2001). Review of statistical methods for QTL mapping in experimental crosses. Lab Anim 30(7):44–52

- Broman KW (2005). The Genomes of Recombinant Inbred Lines. Genetics 169: 1133-1146
- Brouwer DJ and <u>St. Clair</u> DA (2004). Fine mapping of three quantitative trait loci for late blight resistance in tomato using near isogenic lines (NILs) and sub-NILs. Theor Appl Genet 108: 628-638
- Buerstmayr H, Ban T and Anderson JA (2009). QTL mapping and marker-assisted selection for *Fusarium* head blight resistance in wheat: a review. Plant Breeding 128: 1-26
- Burke DT, Carle GF and Olson MV (1987). Cloning of large segments of exogeneous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-812
- Byrne S, Guiney E, Barth S, Donnison I, Mur LAJ and Milbourne D (2009). Identification of coincident QTL for days to heading, spike length and spikelets per spike in *Lolium perenne* L. Euphytica 166: 61–70
- Carbonell EA, Asins MJ, Baselga M, Balansard E and Gerig TM (1993). Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid populations. Theor Appl Genet 86 : 411-416
- Chalhoub B, Belcram H and Caboche M (2004). Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. Plant Biotechnology Journal 2: 181–188
- Chalmers J, Johnson X, Lidgett A and Spangenberg G (2003). Isolation and characterisation of a sucrose: sucrose 1-fructosyltransferase gene from perennial ryegrass (*Lolium perenne*). J. Plant Physiol. 160. 1385–1391
- Cheema J and Dicks J (2009). Computational approaches and software tools for genetic linkage map estimation in plants. Briefings in Bioinformatics 10 (6): 595-608
- Churchill GA and Doerge RW (1994). Empirical Threshold Values for Quantitative Trait Mapping. Genetics 138: 963-971
- Ciannamea S, Busscher-Lange J, de Folter S, Angenent GC and Immink RGH (2006). Characterization of the Vernalization Response in Lolium perenne by a cDNA Microarray Approach. Plant Cell Physiol. 47(4): 481–492
- Cogan NOI, Smith KF, Yamada T, Francki MG, Vecchies AC, jones ES, Spangenberg GC and Foster JW (2005). QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass (*Lolium perenne* L.) Theor Appl Genet 110: 364-380
- Cogan NOI, Ponting RC, Vecchies AC, Drayton MC, George J, Dracatos PM, Dobrowolski MP, Sawbridge TI, Smith KF, Spangenberg GC and Forster JW (2006). Gene-associated single nucleotide polymorphism discovery in perennial ryegrass (*Lolium perenne* L.). Mol Gen Genomics 276: 101–112

- Collard BCY, Jahufer MZZ, Brouwer JB and Pang ECK (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142: 169–196
- Collard BCY and Mackill DJ (2008). Marker-assistied selection: an approach for precision plant breeding in the twenty-first century. Phil. Trans. R. Soc. B 363 : 557–572
- Collins J and Hohn B (1978). Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. Proc. Natl. Acad. Sci. USA 75 (9) : 4242-4246
- Conaghan P and Casler MD (2010). A theoretical and practical analysis of the optimum breeding system for perennial ryegrass. In Grasses for the future. Perennial ryegrass: current and future genetic potential. Ireland. pp 69-97
- Connolly V and Wright-Turner R (1984). Induction of cytoplasmic male-sterility into ryegrass (*Lolium perenne*). Theor Appl Genet 68: 449-453
- Cuadrado A and Schwarzacher T (1999). The chromosomal organization of simple sequence repeats in wheat and rye genomes. Chromosoma 107: 587-594
- Dempster AP, Laird NM, Rubin DB (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society 39 (1): 1-38.
- Devos KM (2005). Updating the "crop circle". Current Opinion in Plant Biology 8:155-162
- Doerge RW (2002). Mapping and analysis of quantitative trait loci in experimental populations. Nature reviews 3: 43-52
- Donnison IS, O'Sullivan DM, Thomas A, Canter P, Moore B, Armstead I, Thomas H, Edwark KJ, Ian P, King IP (2005). Construction of a Festuca pratensis BAC library for map-based cloning in Festulolium substitution lines. Theor. Appl. Genet 110: 846-851
- Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11-15.
- Dubcovsky J, Ramakrishna W, SanMiguek PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL (2001). Comparative sequence analysis of collinear barley rice bacterial artificial chromosomes. Plant physiology 125: 1342-1353
- Eckard NA (2001). Everything in its place: Conservation of gene order among distantly related plant species. The Plant Cell 13: 723-725
- Erickson DL, Fenster CB, Stenøien HK and Price D (2004). Quantitative trait locus analyses and the study of evolutionary process. Molecular ecology 13: 2505-2522

- Faville MJ, Vecchies AC, Schreiber M, Drayton MJ, Hughes J, Jones ES, Guthridge KM, Smith KF, Sawbridge T, Spangenberg GC, Bryan GT and Forster JW (2004). Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 110: 12–32
- Farrar K and Donnison IS (2007). Construction and screening of BAC libraries made from Brachypodium genomic DNA. Nature protocols 2: 1661-1674
- Farrar K, Asp T, Lübberstedt T, Xu M, Thomas AM, Christiansen C, Humphreys MO, Donnison IS (2007). construction of two Lolium perenne BAC libraries and identification of BACs containing genes for disease resistance and forage quality. Mol breeding 19: 15-23
- Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, and Tanksley SD (2000). *fw2.2*: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. Science 289: 85-88
- Frisch M, Quint M, Lübberstedt T, Melchinger AE (2004). Duplicate marker loci can result in incorrect locus orders on linkage maps. Theor Appl Genet 109:305-316
- Gale MD and Devos KM (1998). Comparative genetics in the grasses. Proc.Natl.Acad.Sci. 95: 1971-1974
- Garvin DF, Gu YQ, Hasterok R, Hazen SP, Jenkins G, Mockler TC, Mur LAJ and Vogel JP (2008). Development of genetic and genomic research resources for *Brachypodium distachyon*, a new model system for grass crop research. The Plant Genome 1: 69-84
- Gill GP, Wilcox PL, Whittaker DJ, Winz RA, Bickerstaff P, Echt CE, Kent J, Humphreys MO, Elborough KM, and Gardner RC (2006). A framework linkage map of perennial ryegrass based on SSR markers. Genome 49: 354–364
- Grewal TS, Rossnagel BG, Pozniak CJ, Scoles GJ (2008). Mapping quantitative trait loci associated with barley net blotch resistance. Theor Appl Genet 116: 529-539
- Gupta PK, Balyan HS, Edwards KJ, Isaac P, Korzun V, Röder M, Gautier M-F, Joudrier P, Schlatter AR, Dubcovsky J, De la Pena RC, Khairallah M, Penner G, Hayden MJ, Sharp P, Keller B, Wang RCC, Hardouin JP, Jack P, Leroy P (2002). Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. Theor Appl Genet 105:413-422
- Gupta PK, Rustgi S, Mir RR (2008). Array-based high-throughput DNA markers for crop improvement. Heredity 101: 5–18
- Hackett CA and Broadfoot LB (2003). Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity 90: 33-38
- Haldane JBS (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. J Genet 8: 299-309
- Hasterok R, Masarek A, Donnison IS, Thomas A, King IP, Wolny E, Idziak D, Draper J and Jenkins G (2006). Alignment of the genomes of Brachypodium distachyon and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence *in situ* hybridization. Genetics 173: 349-362
- Hayward MD, Mcadam NJ, Jones JG, Evans C et al. (1994). Genetic markers and the selection of quantitative traits in forage grasses. Euphytica 77 : 269-275
- Hearnden PR, Eckermann PJ, McMichael GL, Hayden MJ, Eglinton JK and Chalmers KJ (2007). A genetic map of 1000 SSR and DArT markers in a wide barley cross. Theor Appl Genet 115: 383-391
- Helentjaris T, King G, Slocum M, Siedenstrang C and Wegman S (1985). Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Molecular Biology 5:109-118
- Heslop-Harrison JS (2000). Comparative genome organization in plants: from sequences and markers to chromatin and chromosomes. The plant cell 12: 617-635
- Hittalmani S, Shashidhar HE, Bagali PG, Huang N, Sidhu JS, Singh VP and Khush GS (2002). Molecular mapping of quantitative trait loci for plant growth, yield and yield related traits across three diverse locations in a doubled haploid rice population. Euphytica 125: 207-214
- Hohmann U, Jacobs G, Telgmann A, Gaafar RM, Alam S, and Jung C (2003). A bacterial artificial chromosome (BAC) library of sugar beet and a physical map of the region encompassing the bolting gene B. Mol Gen Genomics 269: 126–136
- Humphreys MO (2003). Utilization of plant genetic resources in breeding for sustainability. Plant genetic resources 1 (1): 11-18
- Humphreys MO, Yadav RS, Cairns AJ, Turner LB, Humphreys J and Skøt L (2005). A changing climate for grassland research. New Phytol 169 (1): 9-26
- Humphreys M, Feuerstein U, Vandewalle M and Baert J (2010). Ryegrasses. In B. Boller et al. (eds) Fodder Crops and Amenity Grasses. pp: 211-251
- Jaccoud D, Peng K, Feinstein D, Kilian A (2001). Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Research 29: e25.
- Jansen RC (1993). Interval Mapping of Multiple Quantitative Trait Loci. Genetics 135: 205-21 1
- Jansen J (2009). Ordering dominant markers in F2 populations. Euphytica 165:401-417
- Jensen LB, Andersen JR, Frei U, Xing Y, taylor C, Holm PB and Lübberstedt T (2005). QTL mapping of vernalization response in perennial ryegrass (*Lolium perenne* L.) reveals co-location with an orthologue of wheat *VRN1*. Theor Appl Genet 110: 527–536

- Jiang C and Zeng ZB (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. Genetica 101: 47–58
- Jones BM and Lazenby A (1988). The grass crop, the physiological basis of production. Chapman and Hall, London. pp. 226-240.
- Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW (2001). Development and characterisation of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102: 405-415
- Jones ES, Dupal MP, Dumsday JL, Hughes LJ, Forster JW (2002a). An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne L.*). Theor Appl Genet 105: 577-584
- Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphreys MO, King IP, Kishida T, Yamada T, Balfourier F et al. (2002b) An enhanced molecular marker based genetic map of perennial ryegrass (Lolium perenne) reveals comparative relationships with other Poaceae genomes. Genome 45: 282-295.
- Jones N, Ougham H and Thomas H (1997). Markers and mapping: we are all geneticists now. New Phytol. 137: 165-177
- Joshi SP, Ranjekar PK and Gupta VS (1999). Molecular markers in plant genome analysis. Curr. Sci. 77: 230-240
- Kearsey MJ and Farquhar AGL (1998). QTL analysis in plants; where are we now? Heredity 80: 137-142
- Keller B and Feuillet C (2000). Colinearity and gene density in grass genome. Trends in Plant Science. 5: 246-251
- De Keyser E, Yan Shu Q, Van Bockstaele E, De Riek J (2010). Multipoint-likelihood maximization mapping on 4 segregating populations to achieve an integrated framework map for QTL analysis in pot azalea (*Rhododendron simsii* hybrids). BMC Molecular Biology 11:1
- Klein EP, Klein RR, Cartinhour SW, et al. (2000). A High-throughput AFLP-based Method for Constructing Integrated Genetic and Physical Maps: Progress Toward a Sorghum Genome Map. Genome Res. 10: 789-807
- Kopecky D, Bartos J, Lukaszewski AJ, Baird JH, Cernoch V, Kolliker R, Rognli OA, Blois H, Caig V, Lubberstedt T, Studer B, Shaw P, Dolezel J, Kilian A (2009).
  Development and mapping of DArT markers within the Festuca - Lolium complex. BMC Genomics 10: 473
- King J, Armstead IP, Donnison IS, Thomas HM, Jones RN, Kearsey MJ, Roberts LA, Thomas A, Morgan WG, King IP (2002). Physical and Genetic Mapping in the Grasses *Lolium perenne* and *Festuca pratensis*. Genetics 161: 315-324

- King J, Thorogood D, Edward KJ, Armstead IP, Roberts L, Skøt K, Hanley Z and King IP (2008). Development of a Genomic Microsatellite Library in Perennial Ryegrass (*Lolium perenne*) and its Use in Trait Mapping. Annals of Botany 101: 845–853
- Kloosterman B, Oortwijn M, uitdeWilligen J, America T, de Vos R, Visser R GF and Bachem C WB (2010). From QTL to candidate gene: Genetical genomics of simple and complex traits in potato using a pooling strategy. BMC Genomics 11:158
- Kosambi DD (1944). The estimation of map distances from recombination values. Ann Eugenics 12: 172-175
- Kubat Z (2007). Chromosome walking with BAC clones as a method of genome mapping. Plant Soil Environ. 53: 447-450
- Kuleung C, Baenziger PS, Dweikat I (2004). Transferability of SSR markers among wheat, rye, and triticale. Theor Appl Genet 108: 1147-1150
- Kurata N, Moore G, Nagamura Y, Foote T, Yano M, Minobe Y, Gale M (1994). Conservation of genome structure between rice and wheat. Bio/Technology 12: 276-278
- Laido G, Barabaschi D, Tondelli A, Gianinetti A, Stanca AM, Li Destri Nicosia O, Di Fonzo N, Francia E and Pecchioni N (2009). QTL alleles from a winter feed type can improve malting quality in barley. Plant Breeding 128 : 598-605
- Lamkey KR and Lee M (1993). Quantitative genetics, molecular markers, and plant improvement. In BC Imrie and JB Hacker (ed.) Focused plant improvement: Towards responsible and sustainable agriculture. Proc. 10<sup>th</sup> Australian Plant Breeding Conf., Gold Coast. pp: 104-115
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE and Newburg L (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181
- Lander E and Botstein D (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199
- Lem P and Lallemand J (2003). Grass consensus STS markers: an efficient approach for detecting polymorphism in Lolium. Theor Appl Genet 107: 1113-1122
- Li J, Thomson M and McCouch SR (2004). Fine Mapping of a Grain-Weight Quantitative Trait Locus in the Pericentromeric Region of Rice Chromosome 3. Genetics 168: 2187-2195
- Liu GF, Yang J, Zhu J (2006). Mapping QTL for Biomass Yield and Its Components in Rice (*Oryza sativa* L.). Acfa Genetica Sinica 33 (7): 607-616

- Liu T, Mao D, Zhang S, Xu C and Xing Y (2009). Fine mapping *SPP1*, a QTL controlling the number of spikelets per panicle, to a BAC clone in rice (*Oryza sativa*). Theor Appl Genet 118: 1509-1517
- Lyttle TW (1991). Segregation distorters. Annu. Rev. Genet. 25: 511-557
- Mackay TFC (2001). The genetic architecture of quantitative traits. Annu Rev Genet 35: 303-339
- Maloof JN (2003). QTL plant growth and morphology. Curr Opin plant Biol 6: 85-90
- Marschner H (2002). Mineral nutrition of higher plants. New York: Academic Press. 880 p.
- Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD and Waterston RH (1997). High Throughput Fingerprint Analysis of Large-Insert Clones. Genome Res 7: 1072-1084
- Mauricio R (2001). Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. Nature 2: 370-381
- Melchinger AE, Utz HF, Schön CC (1998). Quantitative Trait Locus (QTL) Mapping Using Different Testers and Independent Population Samples in Maize Reveals Low Power of QTL Detection and Large Bias in Estimates of QTL Effects. Genetics 149: 383–403
- Mester DI, Ronin YI, Hu Y, Peng J, Nevo E, Korol AB (2003). Efficient multipoint mapping: making use of dominant repulsion-phase markers. Theor Appl Genet 107: 1102-1112
- Meyers BC, Scalabrin S and Morgante M (2004). Mapping and sequencing complex genomes: let's get physical! Nature 5: 578-589
- Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR and Sasaki T (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3: 87-103,
- Mollinari M, Margarido GRA, Vencovsky R and Garcia AAF (2009). Evaluation of algorithms used to order markers on genetic maps. Heredity 103: 494-502
- Moore G, Devos KM, Wang Z, Gale MD (1995). Grasses, line up and form a circle. Current Biol 5: 737-739
- Morgante M and Salamini F (2003). From plant genomics to breeding practice. Current Opinion in Biotechnology 14: 214-219
- Mott R (2006). Quantitative Trait Loci (QTL) Mapping Methods. Encyclopedia of Life Sciences.

- Murray AW and Skostak JW (1983). Construction of artificial chromosomes in yeast. Nature 305: 189-193
- Muylle H, Baert J, Van Bockstaele E, Pertijs J, Roldan-Ruiz I (2005). Four QTLs determine crown rust (*Puccinia coronata* f. sp. *lolii*) resistance in a perennial ryegrass (*Lolium perenne*) population. Heredity 95: 348-357
- Nam YW, Lee J-R, Song K-H, Lee M-K, Robbins MD, Chung S-M, Staub JE and Zhang H-B (2005). Construction of two BAC libraries from cucumber (*Cucumis sativus* L.) and identification of clones linked to yield component quantitative trait loci. Theor Appl Genet 111: 150-161
- Nimalgoda SD, Cloutier S and Walichnowski Z (2003). Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (Triticum aestivum L.) and validation of genome coverage using locus-specific primers. Genome 46: 870-878
- Oishi H, Takahashi W, Ebina M and Takamizo T (2010). Expression and gene structure of the cold-stimulated gene *Lcs19* of Italian ryegrass (*Lolium multiflorum* Lam.). Breeding science 60: 330-335
- O'Sullivan DM, Ripoll PJ, Rodgers M, Edwards KJ (2001). A maize bacterial artificial chromosome (BAC) library from the European flint inbred line F2. Theor Appl Genet 103: 425-432
- Parh DK, Jordan DR, Aitken EAB, Mace ES, Jun-ai P, McIntyre CL, Godwin ID (2008). QTL analysis of ergot resistance in sorghum. Theor Appl Genet 117: 369-382
- Paterson AH, Tanksley SD and Sorrells ME (1991). DNA markers in plant improvement. Advances in Agronomy 46: 39-90
- Pearson A, Cogan NOI, Baillie RC, Hand ML, Bandaranayake CK, Erb S, Wang J, Kearney GA, Gendall AR, Smith KF and Forster JW (2010). Identification of QTLs for morphological traits influencing waterlogging tolerance in perennial ryegrass (Lolium perenne L.). Theor Appl Genet DOI 10.1007/s00122-010-1473-8
- Peleg Z, Saranga Y, Suprunova T, Ronin Y, Röder MS, Kilian A, Korol AB, Fahima T (2008). High-density genetic map of durum wheat x wild emmer wheat based on SSR and DArT markers. Theor Appl Genet 117: 103-115
- Peters JL, Cnudde F and Gerats T (2003). Forward genetics and map-based cloning approaches. TRENDS in Plant Science 10 (8): 484-491
- Peterson DG, Tomkins JP, Frisch DA, Wing RA, Paterson AH (2000). Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. Journal of Agriculture Genomics 5: www.ncgr.org/research/jag
- Pflieger S, Lefebvre V and Causse M (2001). The candidate gene approach in plant genetics: a review. Molecular Breeding 7: 275-291

- Powell W, Machray GC and Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends in plant science 1(7): 215-222
- Qi X, Bakht S, Qin B, Leggett M, Hemmings A, Mellon F, Eagles J, Werck-Reichhart D, Schaller H, Lesot A, Melton R and Osbourn A (2006). A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. Proc. Natl. Acad. Sci. 103: 18848-18853
- Ramsay L, Macaulay M, Cardle L, Morgante S, degli Ivanissevich S, Maestri E, Powell W, Waugh R (1999). Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley
- Ramsay L, Macaulay M, degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000). A Simple Sequence Repeat-Based Linkage Map of Barley. Genetics 156: 1997-2005
- Rauh BL, Basten C and Buckler IV ES (2002). Quantitative trait loci analysis of growth response to varying nitrogen sources in Arabidopsis thaliana. Theor Appl Genet 104:743-750
- Renwick JH (1972) The mapping of human chromosomes. Annu. Rev. Genet. 5: 81-120
- Ronin Y, Korol A, Shtemberg M, Nevo E, Soller M (2003). High-resolution mapping of quantitative trait loci by selective recombinant genotyping. Genetics 164: 1657-1666
- Rossini L, Vecchietti A, Nicoloso L, Stein N, Franzago S, Salamini F and Pozzi C (2006). Candidate genes for barley mutants involved in plant architecture: an in silico approach. Theor Appl Genet 112: 1073–1085
- Rozen S and Skaletsky HJ (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods* and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, 365-386
- Rust AG, Mongin E and Birney E (2002). Genome annotation techniques: new approaches and challenges. DDT 7 (11): 70-76
- Saha MC, Rouf M, Zwonitzer JC, Chekhovskiy K and Hopkins AA (2005). An SSRand AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.). Theor Appl Genet 110: 323-336
- Saha MC, Cooper JD, Rouf Mian MA, Chekhovskiy K, May GD (2006). Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor Appl Genet 113:1449-1458
- Salvi S and Tuberosa R (2005). To clone or not to clone plant QTLs: present and future challenges. Trends in Plant science 10 (6): 297-304

- Sax K (1923). The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics 8: 552-560
- Schmidt T and Heslop-Harrison JS (1998). Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends in Plant Science 3: 195-199
- Schmidt R (2002). Plant genome evolution: lessons from comparative genomics at the DNA level. Plant Molecular Biology 48: 21-37.
- Schwarzacher T (2003). DNA, Chromosomes and in situ hybridization. Genome 46: 953-962
- Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, and Melchinger AE (2004). Quantitative Trait Locus Mapping Based on Resampling in a Vast Maize Testcross Experiment and Its Relevance to Quantitative Genetics for Complex Traits. Genetics 167: 485-498
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, and Miller W (2000). PipMaker-A Web Server for Aligning Two Genomic DNA Sequences. Genome Res 10: 577-586
- Semagn K, Bjornstad A, Skinnes H, Maroy AG, Tarkegne Y, William M (2006). Distribution of DArT, AFLP, and SSR markers in a genetic linkage map of a doubled-haploid hexaploid wheat population. Genome 49: 545-55
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using a F-factor-based vector. Proc. Natl. Acad. Sci. 89: 8794-8797
- Sim S, Chang T, Curley J, Warnke SE, Barker RE, Jung G (2005). Chromosomal rearrangements differentiating the ryegrass genome from the Triticeae, oat, and rice genomes using common heterologous RFLP probes. Theor Appl Genet 110: 1011-1019
- Skøt L, Humphreys MO, Armstead I, Heywood S, Skøt KP, Sanderson R, Thomas ID, Chorlton KH and Hamilton NRS (2005). An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). Molecular breeding 15: 233-245
- Skøt L, Humphreys J, Humphreys MO, Thorogood D, Gallagher J, Sanderson R, Armstead IP and Thomas ID (2007). Association of Candidate Genes With Flowering Time and Water-Soluble Carbohydrate Content in *Lolium perenne* (L.). Genetics 177: 535-547
- Slate J (2005). Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Molecular Ecology 14: 363-379
- Soderlund C, Longden I and Mott R (1997). FPC: a system for building contigs from restriction fingerprinted clones. CABIOS 13: 523-535.

- Soderlund C, Humphray S, Dunham A, French L (2000). Contigs Built with Fingerprints, Markers, and FPC V4.7. Genome res 10: 1772-1787
- Sorrels ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD et al. (2003). Comparative DNA sequence analysis of wheat and rice genome. Genome Research 13:1818-1827
- Studer B, Jensen LB, Hentrup S, Brazauskas G, Kölliker R and Lübberstedt T (2008). Genetic characterisation of seed yield and fertility traits in perennial ryegrass ( *Lolium perenne* L.). Theor Appl Genet 117: 781-791
- Studer B, Kölliker R, Muylle H, Asp T, Frei U, Roldán-Ruiz I, Barre P, Tomaszewski C, Meally H, Barth S, Skøt L, Armstead IP, Dolstra O and Lübberstedt T (2010). EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (Lolium spp.). BMC Plant Biology 10:177
- Takahashi Y, Teshima KM, Yokoi S, Innan H and Shimamoto K (2009). Variations in *Hd1* proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. PNAS 106: 4555-4560
- Tan YD and Fu YX (2007). A New Strategy for Estimating Recombination Fractions Between Dominant Markers From an F2 Population. Genetics 175: 923-931
- Tan GX, Weng QM, Ren X, Huang Z, Zhu LL and He GC (2004). Two whitebacked planthopper resistance genes in rice share the same loci with those for brown planthopper resistance. Heredity 92: 212-217
- Tanksley SD (1993). Mapping polygenes. Annu. Rev. Genet. 27: 205-233
- Thorogood D, Kaiser WJ, Jones JG, Armstead I (2002). Self-incompatibility in ryegrass 12. Genotyping and mapping the *S* and *Z* loci of *Lolium perenne* L. Heredity 88, 385-390
- Thorogood D, Armstead IP, Turner LB, Humphreys MO, Hayward MD (2005). Identification and mode of action of self-compatibility loci in *Lolium perenne* L. Heredity 94: 356-363
- Tomaszewski C, Byrne SL, Foito A, Kildea S, Kopecký D, Doležel J, Heslop-Harrison JS, Stewart D and Barth S (2012) Genetic linkage mapping in an F2 perennial ryegrass population using DArT markers. Plant Breeding. doi: 10.1111/j.1439-0523.2011.01944.x
- Turner LB, Cairns AJ, Armstead IP, Ashton J, Skøt K, Whittaker D and Humphreys MO (2006). Dissecting the regulation fructan metabolism in perennial ryegrass (*Lolium perenne*) with quantitative trait locus mapping. New phytologist 169: 45-58

- Turner LB, Cairns AJ, Armstead IP, Thomas H, Humphreys MW and Humphreys MO (2008). Does fructan have a functional role in physiological traits? Investigation by quantitative trait locus mapping. New Phytologist 179: 765-775
- Turner LB, Farrell M, Humphreys MO and Dolstra O (2010). Testing water-soluble carbohydrate QTL effects in perennial ryegrass (*Lolium perenne* L.) by marker selection. Theor Appl Genet 121: 1405-1417
- Vales MI, Schön CC, Capettini F, Chen XM, Corey AE, Mather DE, Mundt CC, Richardson KL, Sandoval-Islas JS, Utz HF, Hayes PM (2005). Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. Theor Appl Genet 111: 1260-1270
- Van Daele I, muylle H, Van Bockstaele E and Roldán-Ruiz I (2008). Mapping of markers related to self-incompatibility, disease resistance, and quality traits in *Lolium perenne* L. Genome 51: 644-656
- Van Ooijen JW (2006). JoinMap® 4.0, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.
- Van Ooijen JW (2009). MapQTL 6. Software for the mapping of quantitative trait loci in experimental populations of diploid species.Kyazma BV: Wageningen, Netherlands
- Van Ooijen JW and Voorrips RE (2001). JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands.
- Van Ooijen JW, Boer MP, Jansen RC and Maliepaard C (2002). MapQTL® 4.0, Software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, the Netherlands.
- Varshney RK, Hoisington DA and Tyagi AK (2006). Advances in cereal genomics and applications in crop breeding. TRENDS in Biotechnology 24 (11): 490-499
- Vincze T, Posfai J and Roberts RJ (2003). NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Res. 31: 3688-3691
- Visser RGF, Bachem CWB, de Boer JM, Bryan GJ, Chakrabati SK, Feingold S, Gromadka R, van Ham RCHJ, Huang S, Jacobs JME, Kuznetsov B, de Melo PE, Milbourne D, Orjeda G, Sagredo B and Tang X (2009). Sequencing the Potato Genome: Outline and First Results to Come from the Elucidation of the Sequence of the World's Third Most Important Food Crop. American Journal of Potato Research 86 (6): 417-429
- Voorrips RE (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. The Journal of Heredity 93 (1): 77-78
- Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Homes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23(21): 4407-4414

- Wang S, Basten CJ, and Zeng Z –B (2010). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. ( http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)
- Wegel E, Koumproglou R, Shaw P and Osbourn A (2009). Cell Type-Specific Chromatin Decondensation of a Metabolic Gene Cluster in Oats. The plant cell 21: 3926-3936
- Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinhofs A, Kilian A (2004). Diversity Arrays Technology (DArT) for whole-genome profiling of barley. PNAS 101 (26): 9915-9920
- Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesná J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E and Kilian A (2006). A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7: 206
- Wilkins PW (1991). Breeding perennial ryegrass for agriculture. Euphytica 52: 201-214
- Wilkins PW and Humphreys MO (2003). Progress in breeding perennial forage grasses for temperate agriculture. Journal of Agricultural Science 140, 129–150.
- Wittenberg AHJ, van der Lee T, Cayla C, Kilian A, Visser RGF, Schouten HJ (2005). Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Molecular Genetics and Genomics. 274: 30-39.
- Whisson SC, van derLee T, Bryan GJ, Waugh R, Govers F and Birch PRJ (2001). Physical mapping across an avirulence locus of *Phytophthora infestans* using a highly representative, large-insert bacterial artificial chromosome library. Mol Genet Genomics 266: 289-295
- Woo SS, Jiang J, Gill BS, Paterson AH, Wing RA (1994). Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. Nucleic Acids Research 2: 4922-4931
- Xia L, Peng K, Yang S, Wenzl P, de Vicente C, Fregene M, Kilian A (2005). DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theor Appl Genet 110: 1092-1098
- Xian-Liang S, Xue-Zhen S, Tian-Zhen Z (2006). Segregation distortion and its effect on genetic mapping in plants. Chinese Journal of Agricultural Biotechnology 3 (3): 163-169
- Xie C and Xu S (1999). Mapping quantitative trait loci with dominant markers in fourway crosses. Theor Appl Genet 98: 1014-1021

- Xu Z, Zou F and Vision TJ (2005). Improving Quantitative Trait Loci Mapping Resolution in Experimental Crosses by the Use of Genotypically Selected Samples. Genetics 170: 401-408
- Yamada T, Jones ES, Cogan NOI, Vecchies AC, Nomura T, Hisano H, Shimamoto Y,Smith KF, Hayward MD, and Forster JW (2004). QTL Analysis of Morphological, Developmental, and Winter Hardiness-Associated Traits in Perennial Ryegrass. Crop Sci. 44: 925-935
- Yamada T, Forster JW, Humphreys MW and Takamizo T (2005). Genetics and molecular breeding in *Lolium/Festuca* grass species complex. Grassland Science 51 89-106
- Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y and Sasaki T (2000). *Hd1*, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene CONSTANS. The Plant Cell 12: 2473-2483
- Young ND (1996). QTL mapping and quantitative disease resistance in plants. Annu. Rev. Phytopathol. 34:479–501
- Young ND (2000). Constructing a plant genetic linkage map with DNA markers. DNA-Based markers in plants. RL Philips and JK Vasil (eds). Kluwer Academic Publishers.
- Ytournel F, Gilbert H and Boichard D (2008). Comment affiner la localisation d'un QTL? INRA Prod. Anim. 21 (2): 147-158
- Zeng ZB (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. 90 : 10972-10976
- Zhang HB, Choi S, Woo SS, Li Z and Wing RA (1996). Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Molecular Breeding 2 (1): 11-24
- Zhang HB and Wu C (2001). BAC as tools for genome sequencing. Plant Physiol. Biochem. 39: 195–209
- Zhang ZH, Li P, Wang LX, Hu ZL, Zhu LH, Zhu YG (2004). Genetic dissection of the relationships of biomass production and partitioning with yield and yield related traits in rice. Plant Science 167: 1–8
- Zhao S (2001). A comprehensive BAC resource. Nucleic Acids Research 29: 141-143