Application of molecular markers in selected breeding material and plant genetic resources of *Lolium perenne* L.

Siyang Liu

Göttingen 2015

Application of molecular markers in selected breeding material and

plant genetic resources of *Lolium perenne* L.

Dissertation
to obtain the Ph. D. degree
in the International Ph. D. Program for Agricultural Sciences in Goettingen (IPAG)
at the Faculty of Agricultural Sciences,
Georg-August-University Göttingen, Germany

presented by

Siyang Liu

born in Taiyuan, Shanxi Prov., China

Göttingen, May 2015

D7

1. Name of supervisor: Prof. Dr. Heiko C. Becker

2. Name of co-supervisor: Prof. Dr. Gunter Backes

Date of dissertation: 12. May. 2015

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2x	Diploids
4x	Tetraploids
ADF	Acid detergent fiber content
ADL	Acid detergent lignin content
AFLP	amplified fragment length polymorphism
AH	Absence of heads
AM	Association mapping
AMOVA	Analysis of molecular variance
BPH	Better parent heterosis
CMS	Cytoplasmic male sterility
CV	Coefficient of variation
DArT	Diversity array technology
DMY	Dry matter yield
DNDF	Digestive neutral detergent fiber content
DUS	Distinctness, uniformity and stability
EG	Early spring growth
FMY	Fresh matter yield
GCA	General combining ability
GD	Genetic distance
GEBV	Genomic estimated breeding value
GS	Genomic selection
HP	Hybrid performance
ISSR	Inter-simple sequence repeat

JD	Jaccard genetic dissimilarity distance
LD	Linkage disequilibrium
LG	Linkage group
LSD	Least significant Difference
MAF	Minor allelic frequency
MAS	Marker assisted selection
MPH	Mid-parent heterosis
MPP	Mid-parent performance
MRD	Modified Roger´s genetic dissimilarity distance
NDF	Neutral detergent fiber content
NIRS	Near infrared reflectance spectroscopy
OMD	Organic matter digestibility
PCo	Principal coordinate
PCoA	Principal coordinate analysis
PIC	Polymorphism information content
QTL	Quantitative trait loci
RAPD	randomly amplified polymorphic DNA
RC	Resistance to crownrust
SD	Sward density
SI	Self incompatibility
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
WH	Winter hardiness

Chapter 1

General Introduction

1.1 The biology of *Lolium perenne* L. and its economic importance

Lolium perenne L. or perennial ryegrass is one of the most important forage grasses in temperate regions including northwest Europe, New Zealand, Japan, Australia, South Africa and South America (Wilkins and Humphreys 2003). It contains seven pairs of chromosomes (2n = 2x =14) and has a relatively large genome size (Boller et al. 2010). Tetraploid perennial ryegrass cannot be not found in nature but can be induced by colchicine treatment (Nair 2004). Its obligate outbreeding reproductive system is assured by an efficient two-loci self-incompatibility (SI) mechanism (Cornish et al. 1979) therefore perennial ryegrass populations are highly heterozygous and heterogeneous.

Perennial ryegrass is native to Europe and taxonomically classified as genus *Lolium*, tribe *Poeae*, sub-family *Pooideae*, family *Poaceae* and related to many important cereals like rice, wheat, barley, rye (Kellogg 2001) and forage crops like tall fescue and meadow fescue (Shinozuka et al. 2012). Other important *Lolium* spp. include Italian ryegrass (*Lolium multiflorum* L.) and hybrid ryegrass (*Lolium x hybridum* L.) generated from interspecific crosses between perennial and Italian ryegrass. Morphologically *L. perenne* differs from *L. multiflorum* in the folded leave buds and awnless spikelets. Moreover, perennial ryegrass is more persistent and its leaves and stems are more digestible than those of other grass species (Boller et al. 2010).

Apart from its major usage as silage or pasture owing to high digestibility and palatability, perennial ryegrass can be also used in nutrient recycling systems, soil conservation and as turf which correspond to its high growth rate under fertile soil, the fibrous root system and good wear tolerance (Hannaway et al. 1999). It might also serve as an alternative or complementary plant for bioenergy production which is currently dominated by maize (Aguirre et al. 2012; Salces et al. 2013). The merits of perennial ryegrass in bioenergy production include high yield potential, persistence, easy management requirements and relatively lower inputs (Lewandowski et al. 2003; Searchinger et al. 2008; Aguirre et al. 2012).

The economic importance of perennial ryegrass compared with other *Lolium spp.* can be manifested by the number of listed varieties and annual seed production: in 2007, OECD listed a total of 1156 *Lolium perenne*, 472 *Lolium multiflorum* and 110 *Lolium x hybridum* cultivars; since 2000, European countries produced on average 83,660 t seeds per year for perennial ryegrass

compared to 39,110 t for Italian ryegrass (Boller et al. 2010).

1.2   Breeding of *Lolium perenne* L.

The domestication and breeding of *Lolium perenne* have a relatively short history. According to Bolaric et al. (2005), the initial breeding of perennial ryegrass started in the 1920s. The major interest and effort were paid to the improvement of persistency, yield, digestibility, biotic and abiotic resistance (Luetke Entrup 2008; Boller et al. 2010). Most remarkable achievements in the past years include: 1) the enhancement of sugar content; 2) increased N usage efficiency; 3) prolonged growth season (Wilkins and Lovatt 2004); 4) resistance to rust (*Puccinia* species) by which the sugar content, yield and digestibility could be largely increased (Potter 1987).

Another major achievement is the induction of tetraploids in perennial ryegrass. Since the first tetraploid variety developed in 1950s, the proportion of tetraploid varieties in the market constantly increased. In 2007, 50% of the registered perennial ryegrass varieties in Germany were tetraploid (Humphreys et al. 2010). The chromosome doubling rendered enhanced resistance, digestibility as well as fresh matter yield in tetraploid perennial ryegrass (Hannaway et al. 1999). However, due to higher water content in comparison to the diploid counterpart, the increase of dry matter yield over diploid perennial ryegrass is not significant (Humphreys et al. 2010).

The genetic gain for dry matter yield obtained over the last 90 years was estimated to be 4% per decade (Humphreys 2005) which is four times lower than the gains estimated for many cereals (Aguirre et al. 2012). The relatively low yield improvement might be attributed to 1) the long breeding cycles required due to the perennial nature; 2) the different final products: in grasses only the biomass yield is of interest for yield instead of grain yield in cereals, and biomass yield cannot be improved by improving the harvest index as in cereals; 3) incomplete utilization of heterosis: yield improvement is achieved by recurrent selection or synthetic cultivars which only exploit heterosis partially ; 4) the consideration of a wide variety of traits which are not correlated or even negatively related with yield (Casler 2001) and 5) the poor correlation

between competitive sward condition used in practice and the spaced-plant condition often used in the selection process which reduce the efficiency of the selection (Foster 1973).

1.3 Conventional breeding methodology

Owing to the efficient SI system, ryegrass is mainly developed as population or synthetic cultivars. Recurrent selection plays a major role in the population improvement, either with or without progeny tests. The former is denominated as phenotypic selection or mass selection because the selection decision is based on the performance of the individual or its derivative clones (Posselt 2010a). Therefore the environmental influence is large for some traits. Progeny testing is suitable for narrow based populations because the half-sib or full-sib families are able to enlarge the genetic variance and facilitate the selection, especially when the heritability is low (Poehlman 1994). Based on a simulation study, Casler and Brummer (2008) proposed that among and within family selection is superior to mass selection because it provides a higher expected genetic gain.

Two characteristics for perennial ryegrass breeding are the possibilities of using clones in the selection and the requirement to conduct both space-plant and plot trials in breeding programs. The former one could help to separate the environmental variance from the genetic variance (Posselt 2010a). The latter one sets special difficulties for breeding due to the low agreement between space-plant performance used in the selection and the actual performance revealed in the competitive sward condition (Pesek and Baker 1971; Casler et al. 1996). Therefore it is recommended that field trials in micro plot conditions should be observed simultaneously with space-plant tests in order to get rid of this inconsistency (Wilkins 1991).

After several rounds of selection, the favorable alleles will be accumulated or fixed in the breeding material. For mass selection, the improved material could be registered as variety or used as advanced breeding material (Posselt 2010a). The parents used in composing synthetic varieties could be also derived from this advanced material. Poly-cross or Top-cross followed by

progeny testing can be used to select the parents. Based on the number of individuals used in the formation of the synthetics, broad-based and narrow-based synthetics could be constructed. Higher number of components reduces the inbreeding depression but might be not favorable for DUS (Distinctiveness, Uniformity and Stability) whereas low number of components enhances the selection intensity. According to synthetic prediction of the best synthetics based on experimental data, Posselt et al. (2001) concluded that 5 to 15 components are the optimum number of components for a perennial ryegrass synthetics.

1.4   Molecular markers and their application in perennial ryegrass

Conventional breeding approaches rely heavily on the selection based on phenotypes of plants or its progenies without knowing the mechanisms behind the phenotypic variation and the interaction between genotypes and environmental factors. Molecular markers build the bridge between the traits and their causative genes. The development of low cost and high throughput genotyping systems permits precise targeting of the genes and could be used in marker-assisted selection (MAS) (Jones et al. 2009).

The application range of molecular markers includes the estimation of genetic distances to assess the genetic diversity, the estimation of relatedness between genotypes; bi-parental linkage mapping as well as association mapping to localize QTL for MAS and genomic selection. In grasses, the application of markers lagged behind that in cereals (Humphreys et al. 2010). In recent years, however, progress has been made, especially in diversity studies and linkage mapping.

Genetic diversity and genetic distances estimated by molecular markers can be used in genebank management (Börner et al. 2012), cultivar protection (Roldán-Ruiz et al. 2001), genetic conservation (Montilla-Bascón et al. 2013) and to identify population structure (Van Inghelandt et al. 2010). The correlation between genetic distance and mid-parent heterosis was also revealed by Melchinger (1999), which might facilitate the selection of parents in current breeding programs. The diversity of perennial ryegrass germplasm or cultivars has been

examined by AFLPs (Roldàn-Ruiz et al. 2000; Guthridge et al. 2001; Elazreg et al. 2011), ISSRs (Ghariani et al. 2003; Hu et al. 2011; Safari et al. 2014), RAPDs (Bolaric et al. 2005a; Bolaric et al. 2005b) and SSRs (Kubik et al. 2001; Brazauskas et al. 2011). The merit of selecting parents by genetic distance was revealed by Kölliker et al. (2005) who reported that the more diverse synthetics produced by selecting parents with higher genetic distance outperformed the narrow synthetics in yield.

The initial step of linkage mapping is the construction of the linkage maps. Unlike inbreeding species in which double haploid populations or F2 segregation population can be used, in other species like perennial ryegrass, one-way or two-way pseudo test-crosses have to be exploited. The first comprehensive linkage map in perennial ryegrass was based on SSR, AFLP, and RFLP markers (Jones et al. 2002). QTL for plant architecture, herbage yield, quality characters, cold tolerance, heading date variation and seed production were identified with this map (Yamada et al. 2004; Cogan et al. 2005; Armstead et al. 2008). Other mapping populations have been also developed. Shinozuka et al. (2012) summarized 560 QTL found in *Lolium perenne* on various morphological, physiological and resistance or tolerance traits.

As in linkage mapping with bi-parental progenies, association mapping also looks for markers that are tightly linked with QTL. The difference lies in the amount of linkage disequilibrium (LD) for both methods. In bi-parental progenies, the recombination events are limited and LD is usually large (Al-Maskri et al. 2012). Association mapping exploits LD bearing in broader range of material over many years and the recombination events could be massive. Therefore LD is smaller within the mapping populations (Al-Maskri et al. 2012). This could in turn enhance the mapping resolution. One challenge of association mapping in the estimation process are the inflated false positives caused by structure in the mapping population. Mixed linear model fitting a random term that accounts for relatedness within the mapping population proved to be efficient in solving this problem (Yu et al. 2006).

Genomic selection (GS) was proposed firstly in animal breeding (Hayes et al. 2001). The general idea is that instead of selecting significant markers based on significance thresholds and conduct MAS solely based on the selected markers, all markers should be used for prediction. Many QTL with small effects might not be able to reach the threshold in the selection process as in linkage mapping or association mapping. However, the total of these small effects QTL might be

accountable for a large portion of the total genetic variance (Heffner et al. 2009). The consideration of all the markers might improve the prediction. Currently, GS in plants has already been applied in cereals like maize (Zhao et al. 2012) but rarely reported in perennial ryegrass.

1.5   Introduction of the current study

Improvement of biomass yield for perennial ryegrass via hybrid breeding has been proposed but a systematic examination of heterosis as well as identification of potential heterotic patterns was not conducted. On the other hand, molecular markers are able to provide estimations of genetic diversity and the relatedness between materials thereby might be used to facilitate the selection of parents for hybrid breeding.

In the present study, a large germplasm set was compiled including breeding material, cultivars, ecotypes and landraces. Considering the large within population variation, a bulked sampling strategy was applied, which allows an extensive germplasm examination with moderate cost. Although various molecular markers were previously used in diversity studies of *Lolium perenne*, it was not known which one would be more suitable for bulked samples and whether consistent results could be obtained by different marker types. Therefore DArT, SNP, and SSR markers were applied for the genotyping of the germplasm. The results on genetic diversity, genetic distances among accessions as well as strategies for bulk sampling are presented and discussed for different marker types in Chapter 2.

With a subset of the germplasm presented in Chapter 2, population hybrids were constructed using CMS (cytoplasmic male sterility). The parental accessions as well as their hybrids were phenotyped for FMY (fresh matter yield) and DMY (dry matter yield) at five locations. The variation, inheritance and heterosis for these two traits were reported. In addition, the correlation between heterosis, hybrid performance based on field experiments and genetic distances based on markers is presented and discussed in Chapter 3.

In addition to FMY and DMY, 12 other traits were also recorded in some of the locations. With

genotypic and phenotypic data for the parental accessions obtained in this study, an association study as well as genomic selection for DArT, SNP and SSR markers were conducted for a preliminary examination of these two methods. The detected markers in the association study and prediction accuracy of genomic selection are shown in Chapter 4.

Chapter 5 presents a general discussion on the genetic diversity within the germplasm pool and the heterotic effects revealed in the field experiments. The further perspectives of hybrid breeding in perennial ryegrass are also discussed.

Chapter 2

Genetic Diversity in Breeding Material and Plant Genetic Resources of *Lolium perenne* L. analyzed with DArT, SNP, and SSR Markers with Bulk Sampling of Populations

2.1 Introduction

Perennial ryegrass (*Lolium perenne* L.) is an outcrossing diploid species (2n = 2x = 14). The genus *Lolium*, which shares high evolutionary lineage with economically important crops like rice, wheat and barley (Kellogg 2001), is generally deemed as the major forage grass in Northwest Europe, New Zealand and temperate region of Japan, Australia, South Africa and South America (Wilkins and Humphreys 2003). Perennial ryegrass is one of the most important *Lolium* species in terms of numbers of registered varieties and global seed production (Boller et al. 2010). Apart from the economic importance in world fodder production, perennial ryegrass also serves as turf grass or amenity grass (Altpeter et al. 2000). Compared with other *Lolium* species like Italian ryegrass (*Lolium multiflorum*), and hybrid ryegrass (*Lolium x hybridum*), perennial ryegrass displays higher persistence and digestibility (Boller et al. 2010). Additionally, it possesses high resistance against biotic and abiotic stresses(Wilkins 1991) and high yield potential. Therefore it has been also proposed as candidate plant for biogas production (Aguirre et al. 2012; Salces et al. 2013).

The breeding of perennial ryegrass can be traced back to the 1920s (Bolaric et al. 2005a). The major breeding achievements include the improvement of yield and persistency, the increase of nutritional value (Wilkins and Humphreys 2003) as well as the induction of tetraploidy by treatment with colchicine (Dewey 1980). However, the gain in yield is not comparable to that in cereal crops in the same periods (Humphreys 2005; Casler and Brummer 2008). To fully exploit heterosis, cytoplasmic male sterility (CMS) systems were also proposed and developed (Ruge et al. 2003), but currently no hybrid cultivar is available on the market (Aguirre et al. 2012).

The advent of DNA markers provides powerful tools for scientific research and breeding programs (Jonah et al. 2011). On one hand, it is possible to create genetic maps to identify QTL and single genes for important traits (King et al. 2013). Many different marker types have been used in perennial ryegrass for genetic map construction including RFLPs, AFLPs, RAPDs, SSRs and DArTs (Hayward et al. 1998; Bert et al. 1999; Jones et al. 2002; Thorogood et al. 2002; Tomaszewski et al. 2012; King et al. 2013). On the other hand, polymorphic markers could provide reliable characterization for certain germplasm resources (Cruz et al. 2013) and therefore they benefit gene bank management (Börner et al. 2012), cultivar protection (Roldán-Ruiz et al. 2001), genetic conservation(Warburton et al. 2008; Montilla-Bascón et al. 2013) and

facilitate the identification of germplasm structure (Van Inghelandt et al. 2010). These merits could be utilized by breeders to increase the efficiency in developing cultivars (Montilla-Bascón et al. 2013), which is especially valuable in ryegrass breeding because the morphological differences are usually small between bred cultivars and ecotypes (Boller et al. 2010) and the pedigree information is rarely recorded (Hayes et al. 2013).

To date, the diversity of various perennial ryegrass germplasm or cultivars has been examined with different marker types including AFLP (Roldàn-Ruiz et al. 2000; Guthridge et al. 2001; Elazreg et al. 2011), ISSR (Ghariani et al. 2003; Hu et al. 2011; Safari et al. 2014), RAPD (Bolaric et al. 2005a; Bolaric et al. 2005b) and SSR (Kubik et al. 2001; Brazauskas et al. 2011). Many of those studies confirmed a large within population variability (Guthridge et al. 2001; Bolaric et al. 2005b; Elazreg et al. 2011) which indicates that multiple individuals have to be genotyped to ensure the representativeness for a certain accession. Therefore, the number of accessions being tested is usually limited due to the requirement to analyze many individual plants for a specific accession. On the other hand, studies analyzing a broader range of material with restricted number of samples per accession might suffer from the bias caused by sampling effects. Bulk sampling offers a potential solution for this dilemma, for instance, Guthridge et al. (2001) studied six perennial ryegrass populations with a bulk sampling strategy and AFLP markers; Nestmann et al. (2011) investigated the influence of grassland composition on the differentiation of *Lolium perenne* populations with bulked samples by using SNP markers; Cheng-Xiang et al. (2012) examined the diversity of *Castanea mollissima* bulked sample with SSR markers. To our knowledge, however, a diversity study of perennial ryegrass for a large germplasm set with bulk sampling has not been published yet.

The three marker types chosen for the genotypic characterization share common features, like available automated platforms, but also bear several differences. For DArTs, the detection of polymorphisms is not relying on prior knowledge of sequences (Jaccoud et al. 2001), therefore it is particularly ideal for species with publicly limited genome information, but the inherent dominant nature reduces information content (Reif et al. 2005; James et al. 2008). SNPs and SSRs both are co-dominant marker types and highly polymorphic, but currently the publicly available primers or sequences for perennial ryegrass are restricted.

Here we studied the genetic diversity of 297 perennial ryegrass accessions with DArT, SNP and

SSR markers. The objectives of this study were a) to assess the feasibility of bulk sampling for diversity studies of a large germplasm set of *Lolium perenne*; b) to examine the underlining population patterns and genetic diversity within the collections; c) to compare the performance of DArT, SNP and SSR markers in analyzing bulked ryegrass samples.

2.2 Material and Methods

2.2.1 Plant material

A set of 297 accessions of *Lolium perenne* including varieties, breeding material and ecotypes were chosen. The source of the material includes the three German breeding companies Deutsche Saatveredelung (DSV), Saatzucht Steinach (SZS), Norddeutsche Pflanzenzucht (NPZ) as well as the Genebank of the IPK (Leibniz Institute of Plant Genetics and Crop Plant Research). The ploidy status and geographical origin are available for the majority of the accessions (see Tab. 2.1 and Appendix). It is expected that this material represents a broad range of possible variation existing within the German *Lolium perenne* breeding pool as a whole. We denote these 297 accessions as set I. Additional to set I, one variety of *L. multiflorum* (Gordo) with one genotyping replication, one *L. x hybridum* ecotype (GR7418) and eight accessions of the *L. perenne* mapping population VrnA (Jensen et al. 2005) including two parents, two F1 offsprings and four F2 individuals were also included for comparison. Moreover, apart from the replicated *L. multiflorum* accession (Gordo), six replicated samples from set I were also genotyped to test the repeatability for marker systems. These 17 samples together were denominated as set II.

Table 2.1 Summary of set I with 297 *Lolium perenne* L. accessions classified by contributor, geographical origin, ploidy and breeding status

Germplasm Contributor			
DSV	126	IPK	48
NPZ	90	SZS	27
Other[a]	6		
Geographical Origin[b]			
Western EU	197	Eastern EU	10
Northern EU	37	Oceania	5
Southern EU	5	Unknown	43
Ploidy			
Diploid	232	Tetraploid	65
Breeding Status			
Breeding Material	206	Ecotype	43
Variety	42	Landrace	2
Unknown	4		

[a] standard cultivars were not assigned to any particular contributor.
[b] the classification refers to United Nations Statistics Division.
EU: Europe.

Seeds of the complete plant material were sown and later leaves were harvested. Leaf tissue was punched out with a metal rod (ø 1.8 mm) in the upper half of the leaf lamina. For set I and set II, leaves from 30 individuals per accession were equally pooled for each sample to obtain approximately 100-mg of fresh leaf material following Nestmann et al. (2011). In order to assess the influence of bulk size, various sampling strategies, namely 1, 12, 24, 48 and 100 individuals per bulk, were applied to four genebank accessions from set I: GR2725 (13 samples), GR2915 (9 samples), GR3107 (8 samples), GR8502 (7 samples). We denoted these 37 samples as set III (Tab. 2.2). In total, set I, set II, set III summed up to 351 samples.

Table 2.2 Summary of set III with repeated bulk sampling

No. individuals per bulk	GR2725	GR2915	GR3107	GR8502
1 ind.	1	1	1	1
12 ind.	4	4	3	2
24 ind.	4	2	2	2
33 ind.	2	-	-	-
48 ind.	1	1	1	1
100 ind.	1	1	1	1
Total		37		

2.2.2 Molecular markers

The 351 samples were genotyped with DArT, SNP and SSR markers. DNAs of the samples of set I and II were initially extracted at Saaten Union Biotech (Leopoldshöhe, Germany) where the SSR markers were subsequently genotyped. DNAs of set III were extracted at IPK. All prepared DNA samples were also distributed to Trait Genetics GmbH (Gatersleben, Germany) for SNP marker and Diversity Array Pty Ltd. (Canberra, Australia) for DArT marker genotyping, respectively. Pre-selection of markers was conducted in the corresponding company.

DArT markers were scored as presence/absence. Among all the DArT markers applied in this study, 114 were mapped (King et al. 2013), consisting of 18, 17, 16, 21, 22, 11, 9 markers on Chromosomes 1 to 7, respectively. The location of the rest of the DArT markers was unknown.

For SNP markers, allelic frequencies for each sample at each locus were directly inferred from the relative intensity scanned in pyro sequencing procedure. Based on the genetic maps (University of Aarhus and IPK Gatersleben, personal communication by K.J. Dehmer from IPK), 23, 22, 29, 38, 17, 20, 27 markers were located on Linkage Group 1 – 7, the location of the remaining 6 markers was unknown.

For SSRs, the electrophoresis spectra were recorded by a scanner but scored manually. Weak peaks were not scored due to their presumably low influence on the allelic frequencies for a locus. The initial scoring was stored as presence or absence of specific alleles, and then transformed into allelic frequencies. For example, if 4 alleles were found for a certain locus, each allele will be assigned with an allelic frequency of 0.25. A number of 5, 7, 12, 7, 5, 5 and 7 markers, respectively, were allocated on LGs 1 to 7 according to map constructed by Studer et al. (2010).

Marker loci containing 30% or more missing values across all the samples were excluded from the dataset leading to a final marker dataset containing 1384 DArT, 182 SNP and 48 SSR marker loci with an average missing value rate of 5.2%, 3.3% and 3.2%, respectively.

2.2.3 Statistical analysis

Genetic distances (GD) were calculated for the entire samples. Due to the differences between marker types, different distance measures were calculated following Reif et al. (2005). Distances of co-dominant SNP and SSR markers were assessed by Modified Roger`s Distance (MRD) (Wright 1978) whereas dominant DArT markers were calculated with Jaccard Distance (JD) (Jaccard 1908). The Pearson correlation coefficient was determined between all pairs of distances for three marker types. The mean values of GD for set I and for the replicated genotyped samples in set II were also calculated. Nei's genetic diversity (Nei 1987), mathematically equals to polymorphic information content (PIC) (Botstein et al. 1980), was calculated for each locus separately for set I with the formula:

$$\text{PIC} = 1 - \sum_{i=1}^{n} p_i^2$$

where n is the number of alleles and p_i represents the frequency for the ith allele. The average Nei's diversity over all the loci was calculated to determine the total genetic diversity.

Based on the GD matrices, phenograms of the 351 samples were constructed with Unweighted Pair Group Method with Arithmetic means (UPGMA) to check the applicability of bulk sampling strategy. The cophenetic index was calculated as a pairwise Pearson correlation coefficient between the cophenetic matrix and genetic distance matrix to check the fitness of the constructed phenograms.

Cluster analysis was conducted for DArT markers by the software STRUCTURE (Pritchard et al. 2000). Potential subgroup number K = 1 – 10 was tested with three replicates each by applying admixture model. The burn-in time and iterations were both set to 20,000. The ad hoc criterion was utilized to determine the most likely group numbers (Earl et al. 2012). However, the allelic frequency data from SNP and SSR markers were not executable in STRUCTURE. To have a comparable platform which is applicable to all marker types, PCo-based clustering was conducted with principal coordinate (PCo) 1 – 100 calculated from corresponding GD matrices to examine the sub-groups with all three marker types (Fraley et al. 2007). Potential number of clusters was set to 1 – 20 in this analysis.

Analysis of Molecular Variance (AMOVA) based on GD was implemented for set I with different categories defined by contributor, geographical origin, ploidy levels and breeding status (see Appendix and Tab. 2.1). The accessions containing incomplete information were discarded. AMOVA was also applied to the replicate samples in set II in order to dissect the proportion of the variance within and among replications to compare the marker types.

To simplify the multivariate data set and visualize the population patterns, Principal Coordinate Analysis (PCoA) was conducted with two dimensions for three marker types based on their corresponding GD matrices. PCoA was plotted for a combination of set I and set II to compare the discriminative ability among different marker types. Additionally, a PCoA plot was also constructed for set III to inspect the variation caused by various bulk sizes (Tab. 2.2).

Bootstrapping analysis was carried out to detect the variance generated by resampling subsets of complete marker data sets and to determine the minimum marker number required to achieve the acceptable accuracy as the complete data set. Of the total marker numbers, 2, 4, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80 and 90% were randomly selected with 100 repetitions each to form subsets of the entire data set. MRD or JD was further calculated for each of the subsets. The Coefficient of Variation (CV) across replications was determined because CV is not influenced by data with different mean values and more suitable for comparison between different marker types (Van Inghelandt et al. 2010).

R platform was utilized for all calculations, simulations and graphics (R Core Team 2013). Specially, PCo-based clustering was conducted with R package mclust (Fraley et al. 2007). Graphics were prepared by R build-in graphic functions and package ggplot2 (Wickham 2009).

2.3   Results

For set I, 1380 of the 1384 DArT marker loci were polymorphic; all 182 SNP marker loci were polymorphic; the number of alleles for SSR ranged from 2 to 23 with an average number of 8.2 alleles per locus. The JD for DArT markers ranged from 0.00 to 0.73 resulting in a mean distance of 0.45; for SNPs, the MRD were between 0.03 and 0.52 with an average value of 0.34; for SSRs,

the MRD ranged from 0.26 to 0.76 with a mean value of 0.54. The distribution of the distance estimates for three marker types are plotted in Fig 2.1. Regarding the GD between replicated samples in set II, DArT markers exhibited distances from 0.00 to 0.08 with an average of 0.04, while SNP markers differed by 0.06 to 0.21 with a mean of 0.16; in comparison, SSR markers were less reproducible with a GD between replicates from 0.20 to 0.43 and 0.34 on average (Fig. 2.1). Nei's genetic diversity for the set I based on DArTs, SNPs and SSRs was 0.26, 0.32 and 0.54, respectively. For the entire samples, the correlation of corresponding GD estimates was 0.83*** for DArTs and SNPs, 0.68*** for DArTs and SSR, 0.70*** for SSRs and SNPs.

Three indices, namely unique alleles, average alleles per locus as well as genetic diversity based on groups classified in passport data are summarized in Tab. 2.3. Considering geographic origin, we observed higher numbers of unique alleles, average alleles as well as genetic diversity for material from Western and Northern Europe and this holds true for all three marker types. For ploidy levels, diploid material showed higher unique alleles with SSRs and DArTs, but the genetic diversity was similar. Breeding material, ecotypes and material from DSV and NPZ showed higher unique alleles as well as average alleles with DArT and SSR markers than other groups; with SNPs, however, no distinct differences can be identified.

Based on the entire data, phenograms were constrcuted for DArT, SNP and SSR markers. Set III and its corresponding accessions in set I were highlighted with four different colors (see Fig. 2.2). Cophenetic indices were 0.90, 0.90 and 0.76, respectively, indicating an ideal representation of the information contained in the GD matrices. For all marker types, four replicate groups were formed and they were clearly distinct from other material; for DArTs and SSRs, four samples clustered separately from their corresponding replicate groups: GR2725, GR3107 and GR8502 from set III with only one individual as well as GR8502 from set I as bulk of 30 individuals. SNPs showed the similar pattern: although a one-individual-based GR8502 sample was correctly clustered with the replicate group, it was located on the boundary and showed high distance to other members.

PCoA analyses on set III revealed higher consistency for the samples with larger bulk sizes (Fig. 2.3). For all marker systems, four groups could be clearly defined. The first and second principal coordinate explained 31.0% and 18.3% of the molecular variance for DArTs; 27.9% and 20.1% for SNPs; 30.3% and 12.7% for SSRs, respectively. For DArTs, one GR3107 and one GR8502

sample as bulks with few individuals were distinct from their corresponding replicate group, but all four groups were distinguishable. SNPs displayed similar parttern to that of DArTs but some extend of mixture of GR2725 and GR8502 could be observed. For SSRs, one GR2725, three GR3107, three GR2915 and one GR8502 low bulking samples differed from others, but all the four groups were distinct. The fact that high bulk size group is more converged than low size group holds true for all marker types.

For PCoA on sets I and II, the first two principal coordinates only explained 5.1% and 3.1% of the molecular variance for DArTs, 3.8%, 3.4% for SNPs and 3.8% and 3.3% for SSRs (Fig. 2.4). For DArTs and SNPs (Fig. 2.4(a), Fig. 2.4(b)), two accessions of *L. mulitiflorum* (Gordo), one accession of *L. x hybridum* (GR7418) and VrnA mapping population were separated from the rest of set I; for SSRs, however, Gordo and GR7418 were not clearly apart. For all three marker types, large variation could be identified but a clear identification of some potential groups was difficult.

The ad hoc criteria on STRUCTURE analysis for DArTs indicated three potential subgroups in set I (see Fig. 2.5). However, the majority of the material was assigned to the first group: if we define the membership probability threshold at 0.7, 250 accessions would be assigned to group 1 and only 15 accessions could be clearly assigned to the 2nd and 3rd group (Fig. 2.6).

PCo-based clustering revealed rather variable number of clusters when a small number of PCos was used for the analyses, but after a certain point, the estimated number of clusters tended to be constant. For DArTs, the number of clusters ranged from 2 to 16 with PCo 1 to 64 and stabilized at 4 clusters after PCo 65. For SNPs, the amount of groups varied in the range from 2 to 11 with PCo 1 to 63 and stabilized at 3 clusters after PCo 64. For SSRs, the stabilization was reached much earlier than for DArTs and SNPs: after PCo 16, only one cluster was suggested by the model (Fig. 2.7). Despite of several clusters identified after the stabilization for DArTs and SNPs, the majority of the material (over 95%) was only assigned to one group (Fig. 2.8).

AMOVA based on GD and classified by geographical origin, breeding status, ploidy status and contributors was applied to set I and replicated samples in set II, respectively. For set I, though all the factors were significant at p = 0.01, only small proportion of the variance could be explained by defined factors (Tab. 2.4). For replicates in set II, higher variance components were explained by DArTs (98.96%) compared to SNPs (69.95%) and SSRs (52.32%) (Tab. 2.5).

A combined bootstrapping analysis showed that the CV (Coefficient of variation) of the GD estimates among pairs of accessions exponentially decreased when the number of markers selected increased (Fig. 2.9). As a result, DArTs represented lower CV value than SNPs and the highest CV was always obtained in SSRs. 40% (554), 60% (110) and 75% (36) of the total DArT, SNP and SSR marker set were able to provide similar results like the complete data set.
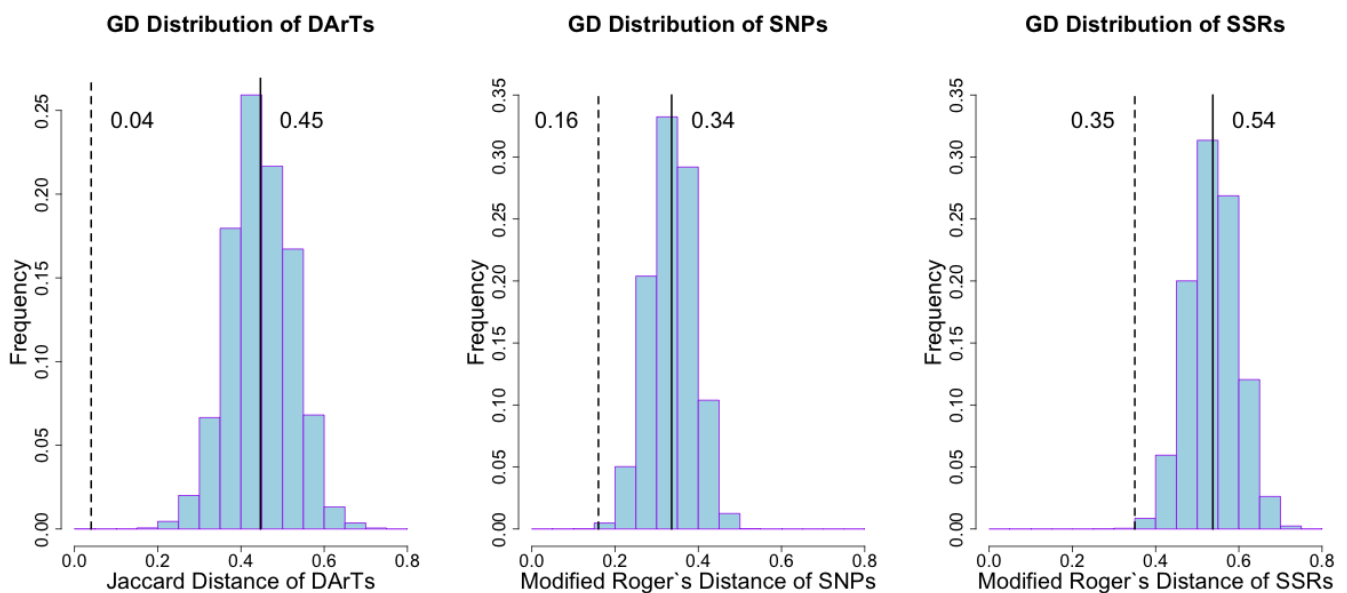
Figure 2.1 Distribution of genetic distances for DArT, SNP and SSR* markers for 297 perennial ryegrass accessions (set I)
Solid line indicates the mean value of the corresponding genetic distance based on 297 core collections.
Dashed line indicates the mean value of replicates sets.
* for SSR, one replicate sample was excluded due to high missing value rate.

Table 2.3 Number of unique alleles, average alleles per locus and genetic diversity based on DArT, SNP and SSR markers for 297 perennial accessions (set I) classified by geographical origin, ploidy level, breeding status and contributors

	DArTs			SNPs			SSRs		
	Unique Alleles	average alleles/locus	Genetic diversity	Unique Alleles[a]	average alleles/locus	Genetic diversity	Unique Alleles	average alleles/locus	Genetic diversity
Eastern	1	1.63	0.21	0	2.00	0.31	1	3.29	0.47
Northern	1	1.89	0.25	0	2.00	0.31	11	5.50	0.53
Oceania	0	1.52	0.21	0	1.99	0.30	4	3.17	0.46
Southern	0	1.56	0.22	0	1.97	0.29	1	3.00	0.48
Western	23	1.99	0.26	0	2.00	0.32	60	7.44	0.54
2x	86	1.99	0.26	0	2.00	0.31	118	8.00	0.54
4x	6	1.90	0.25	0	2.00	0.32	9	5.69	0.53
bre. ma.[b]	18	1.99	0.26	0	2.00	0.32	64	7.29	0.55
variety	1	1.88	0.22	0	2.00	0.31	8	5.25	0.51
ecotype	2	1.91	0.26	0	2.00	0.32	25	5.92	0.54
DSV	13	1.98	0.26	0	2.00	0.32	34	6.94	0.54
IPK	2	1.88	0.22	0	2.00	0.31	13	5.27	0.51
NPZ	3	1.94	0.27	0	2.00	0.32	26	6.54	0.55
SZS	1	1.83	0.23	0	2.00	0.31	4	4.69	0.53

[a] based on the standard that the specific group with the allelic frequency bigger than 0 or smaller than 1 when the allelic frequencies at the rest of material are 0 or 1.
[b] breeding material.

Table 2.4 AMOVA for 297 perennial ryegrass accessions (set I) based on genetic distance estimates of DArT, SNP and SSR markers and classified by geographical origin, breeding status, ploidy status and contributors

DArT				
Source of Variation	df	Sum of squares	Variance Component [c]	Variance%
Geographical Origin [a]				
Among geographical origin groups	4	0.640	0.003	2.64%
Within Geographical origin groups	249	24.210	0.097	97.36%
Breeding Status				
Among breeding status groups	4	0.820	0.003	2.81%
Within Breeding Status	292	29.575	0.101	97.19%
Ploidy				
Among ploidy groups	1	0.396	0.003	2.76%
Within ploidy groups	295	30.000	0.102	97.23%
Contributor [b]				
Among Contributor groups	3	1.101	0.004	3.83%
Within Contributor groups	287	29.004	0.101	96.17%
SNP				
Source of Variation	df	Sum of squares	Variance Component [c]	Variance%
Geographical Origin [a]				
Among geographical origin groups	4	0.289	0.001	1.36%
Within Geographical origin groups	249	13.530	0.054	98.64%
Breeding Status				
Among breeding status groups	4	0.353	0.001	1.49%
Within Breeding Status	292	16.773	0.057	98.51%
Ploidy				
Among ploidy groups	1	0.449	0.004	6.39%
Within ploidy groups	295	16.678	0.057	93.61%
Contributor [b]				
Among Contributor groups	3	0.394	0.001	1.88%
Within Contributor groups	287	16.582	0.058	98.11%

Table 2.4: *continued from previous page*

SSR				
Source of Variation	df	Sum of squares	Variance Component [c]	Variance%
Geographical Origin [a]				
Among geographical origin groups	4	0.928	0.004	2.60%
Within Geographical origin groups	249	35.306	0.142	97.40%
Breeding Status				
Among breeding status groups	4	0.974	0.003	1.88%
Within Breeding Status	292	42.300	0.145	98.12%
Ploidy				
Among ploidy groups	1	0.442	0.003	1.97%
Within ploidy groups	295	42.831	0.145	98.03%
Contributor [b]				
Among Contributor groups	3	1.171	0.004	2.51%
Within Contributor groups	287	41.440	0.144	97.49%

[a] material with unknown origin was removed.
[b] standard cultivar was removed.
[c] variance component was all significant at p = 0.01 after 1000 permutations.

Phenogram of DArTs

cophenetic index = 0.903

Phenogram of SNPs

cophenetic index = 0.893

Phenogram of SSRs

cophenetic index = 0.760
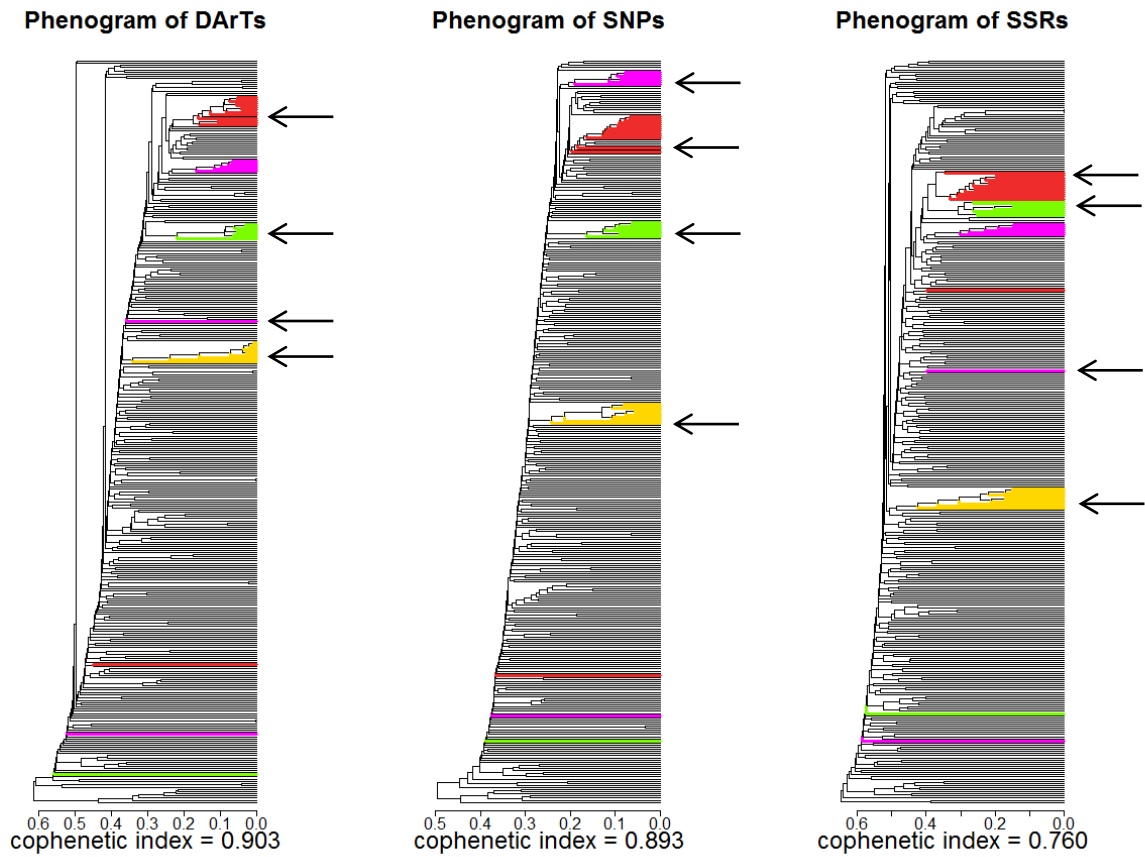
Figure 2.2 Phenograms for the entire germplasm set with set III and their corresponding samples in set I being highlighted

The *cophenetic index* was labeled underneath. *Arrows* indicts the four accessions in set I (corresponds to four accessions in set III) undergone normal 30 individual sampling process. *red*: GR2725; *green*: GR3107; *yellow*: GR2915; *purple*: GR8502.

23
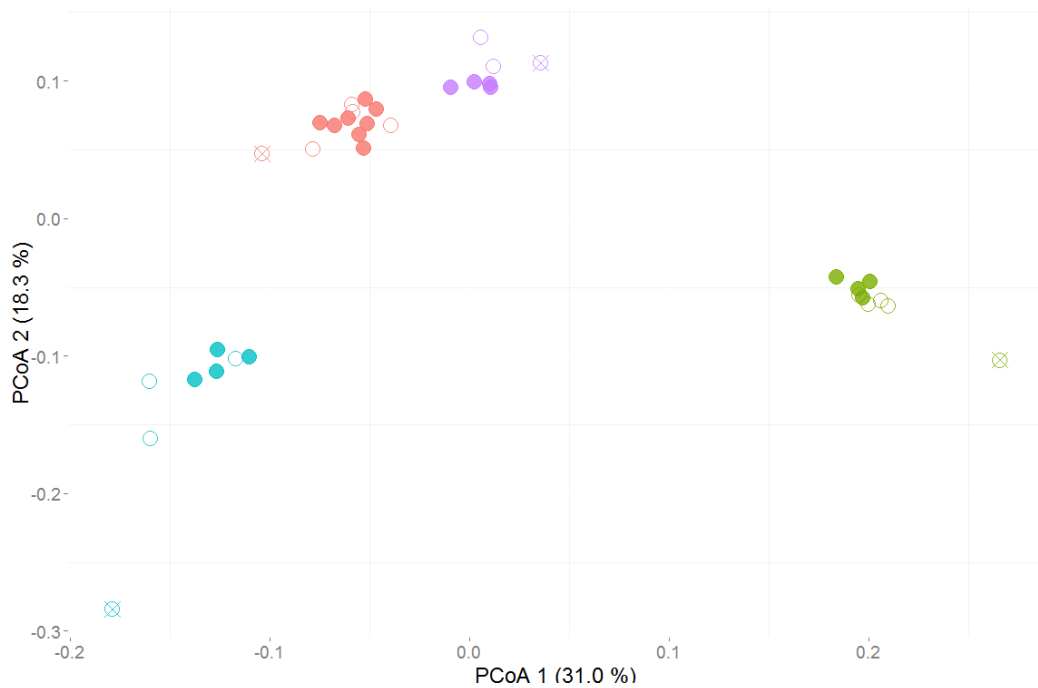
Figure 2.3 PCoA (Principal Coordinate Analysis) of set III (37 samples)
(a) for DArT markers
filled circle: bulk sampling with a sample size above 24; *empty circle*: bulk sampling with the sample size equal to or below 24; *circle with cross*: samples containing only one individual.
red: GR2725; *blue*: GR3107; *green*: GR2915; *purple*: GR8502.
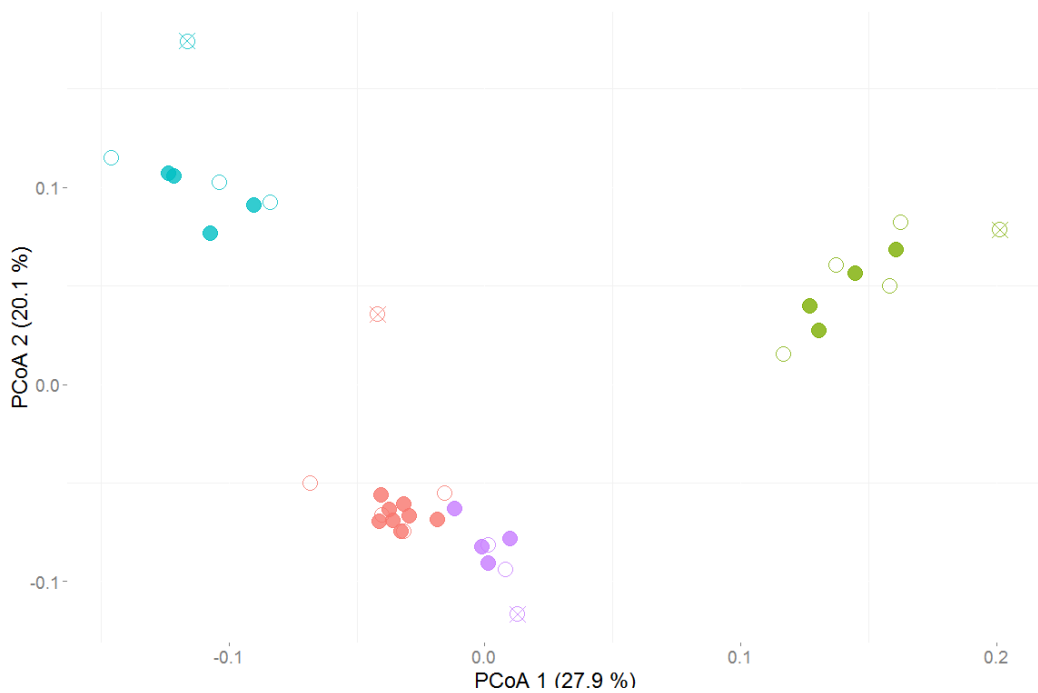
Figure 2.3 PCoA of Set III (37 accessions)
(b) for SNP markers
filled circle: bulk sampling with a sample size above 24; *empty circle*: bulk sampling with the sample size equal to or below 24; *circle with cross*: samples containing only one individual.
red: GR2725; *blue*: GR3107; *green*: GR2915; *purple*: GR8502.

24
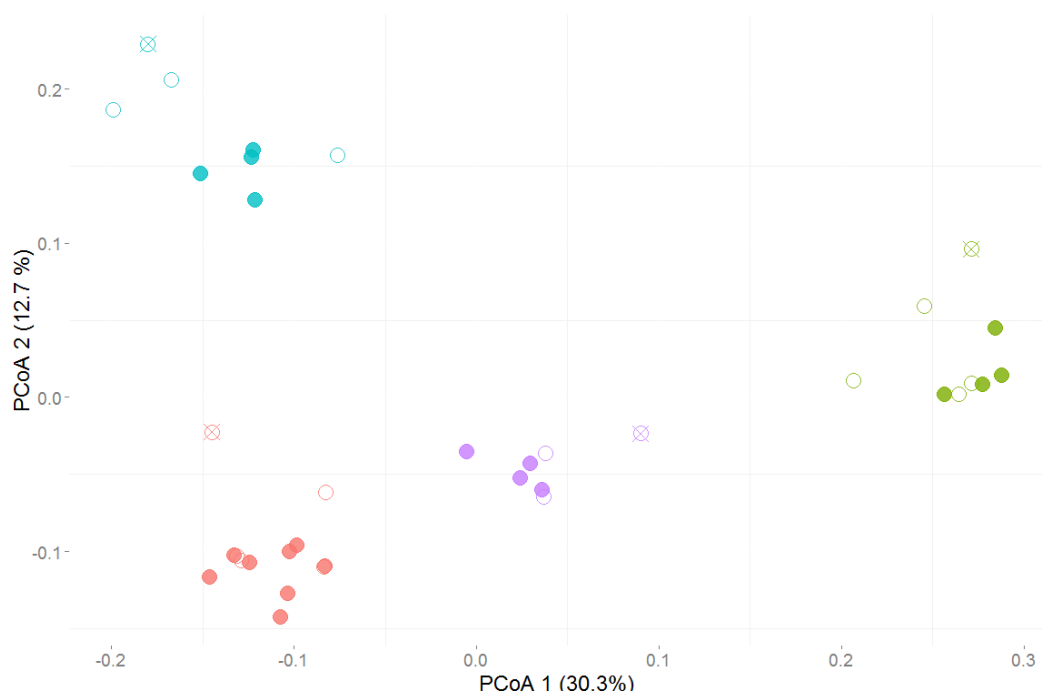
Figure 2.3 PCoA of Set III (37 accessions)
(c) for SSR markers
filled circle: bulk sampling with a sample size above 24; *empty circle*: bulk sampling with the sample size equal to or below 24; *circle with cross*: samples containing only one individual.
red: GR2725; *blue*: GR3107; *green*: GR2915; *purple*: GR8502.

Table 2.5 AMOVA for the replicated accessions (in set II) based on genetic distance estimates of DArT, SNP and SSR markers

	Source of Variation	df	Variance Component[a]	Variance%
DArT	Among replicates groups	5	0.086	98.96%
	Within replicates groups	7	0.001	1.04%
	Total	12		
SNP	Among replicates groups	5	0.033	69.95%
	Within replicates groups	7	0.014	30.05%
	Total	12		
SSR[b]	Among replicates groups	4	0.077	52.32%
	Within replicates groups[b]	6	0.070	47.68%
	Total	10		

[a] all of the variance components were significant at p = 0.01 after 1000 permutations.
[b] In SSR, one replicate set was excluded due to a high missing value rate.
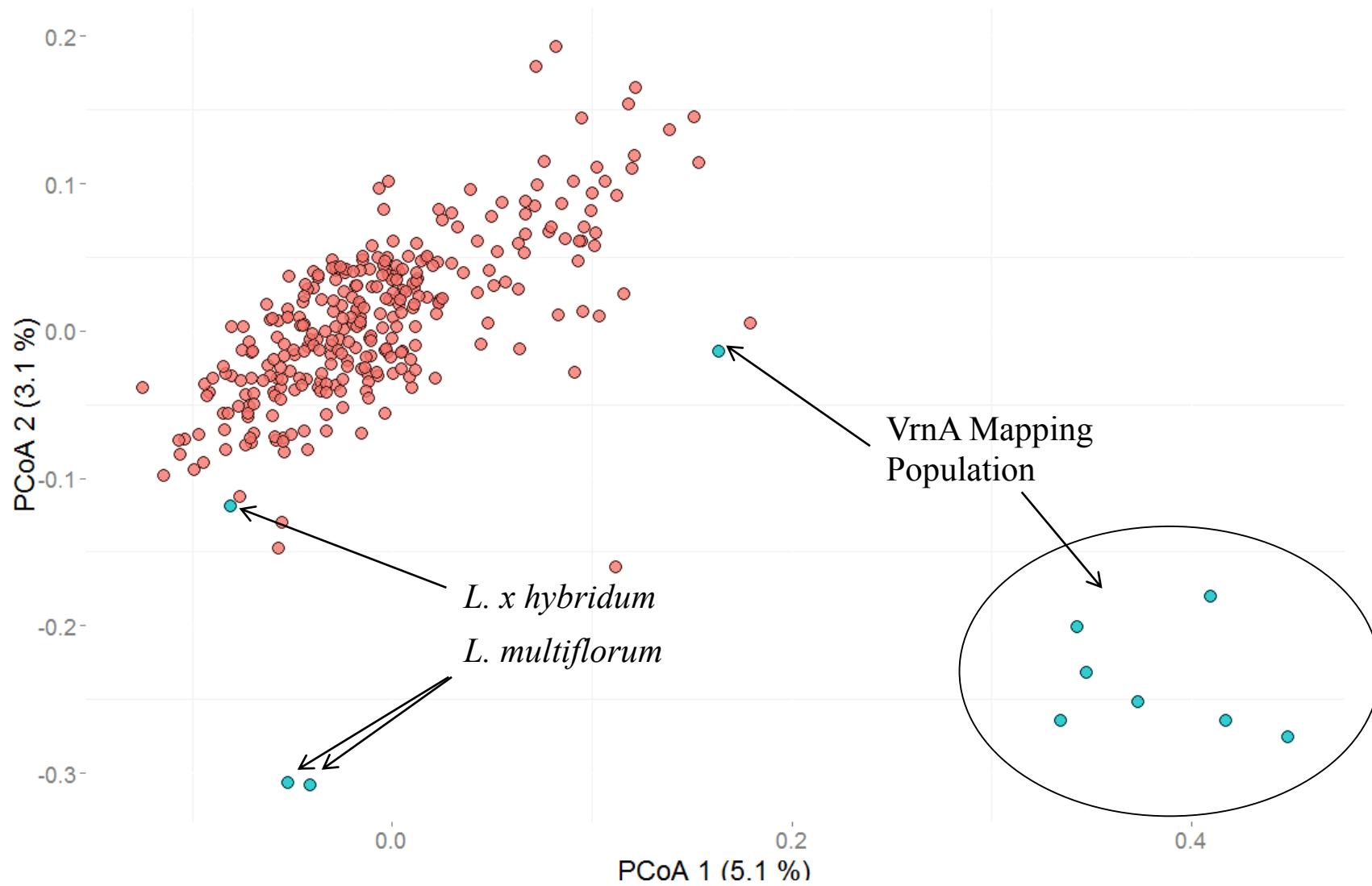
Figure 2.4 PCoA (Principal Coordinate Analysis) for 297 perennial ryegrass accessions (set I), two accessions of *L. multiflorum*, one accession of *L. x hybridum* and eight lines of VrnA mapping population (in set II)

(a) For DArT markers; *red*: set I; *green*: material in set II
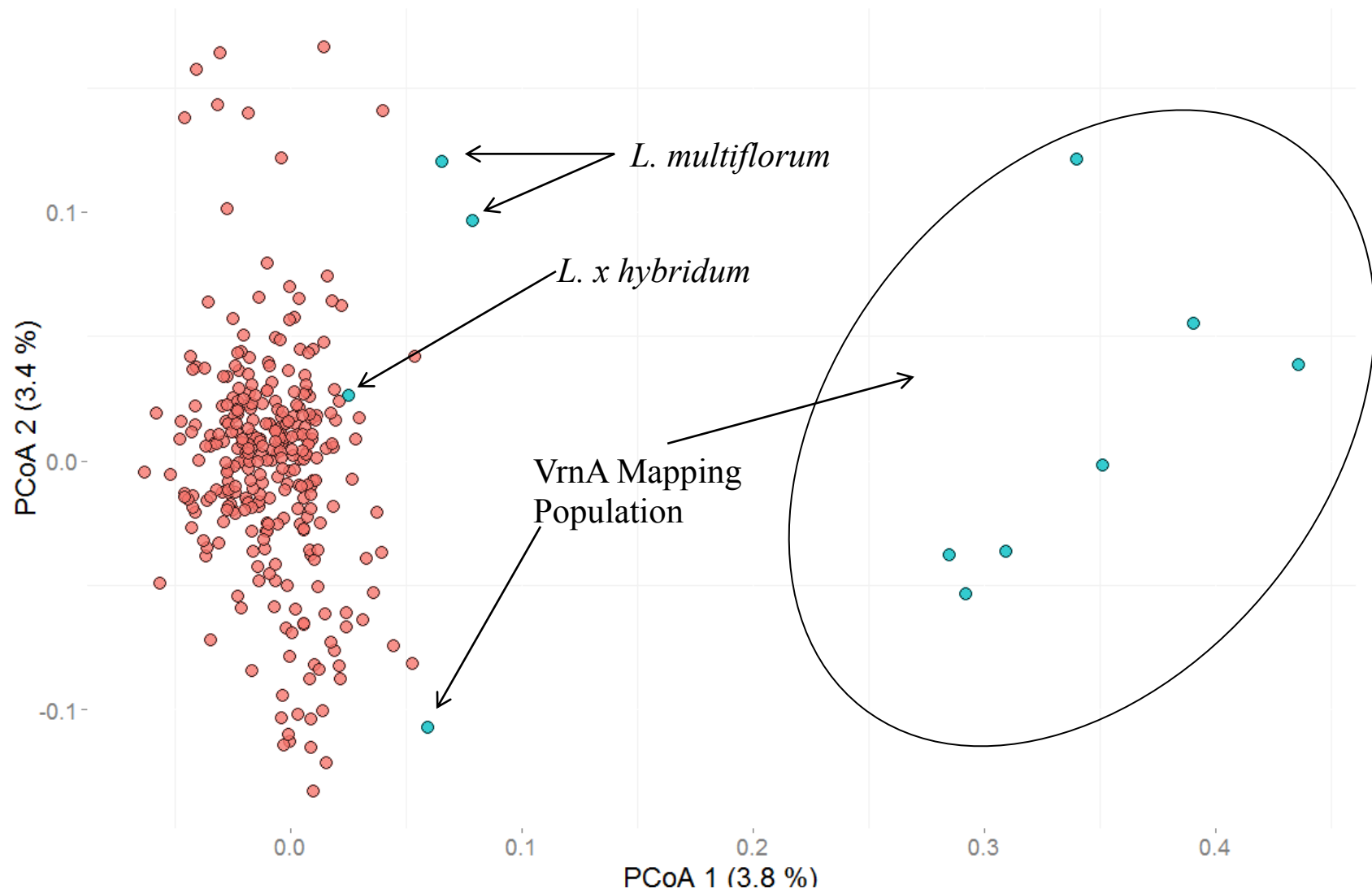
Figure 2.4 PCoA (Principal Coordinate Analysis) for 297 perennial ryegrass accessions (set I), two accessions of *L. multiflorum*, one accession of *L. x hybridum* and eight lines of VrnA mapping population (in set II)

(b) For SNP markers; *red*: set I; *green*: material in set II
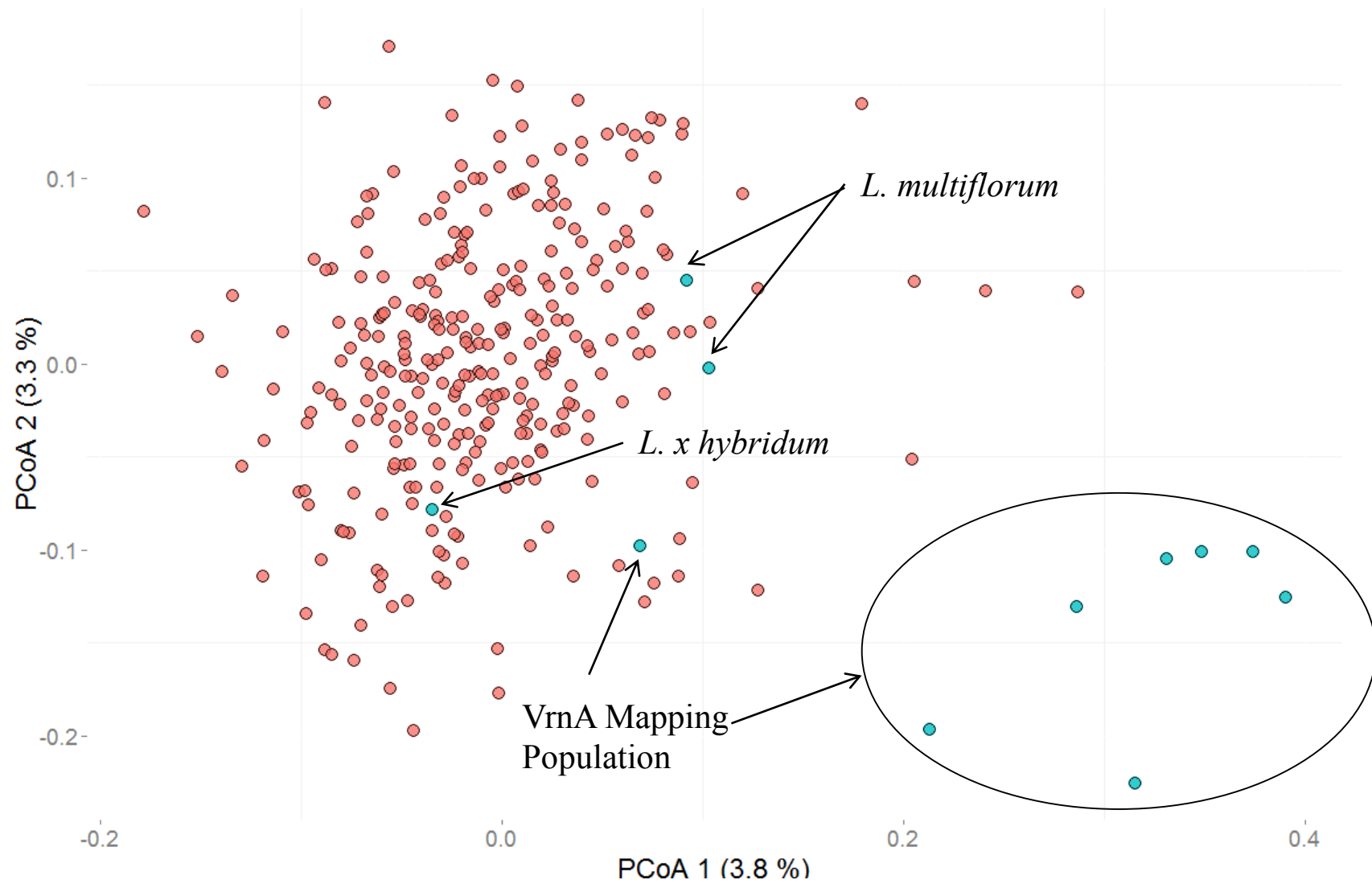
Figure 2.4 PCoA (Principal Coordinate Analysis) for 297 perennial ryegrass accessions (set I), two accessions of *L. multiflorum*, one accession of *L. x hybridum* and eight lines of VrnA mapping population (in set II)

(c) For SSR markers; *red*: set I; *green*: material in set II
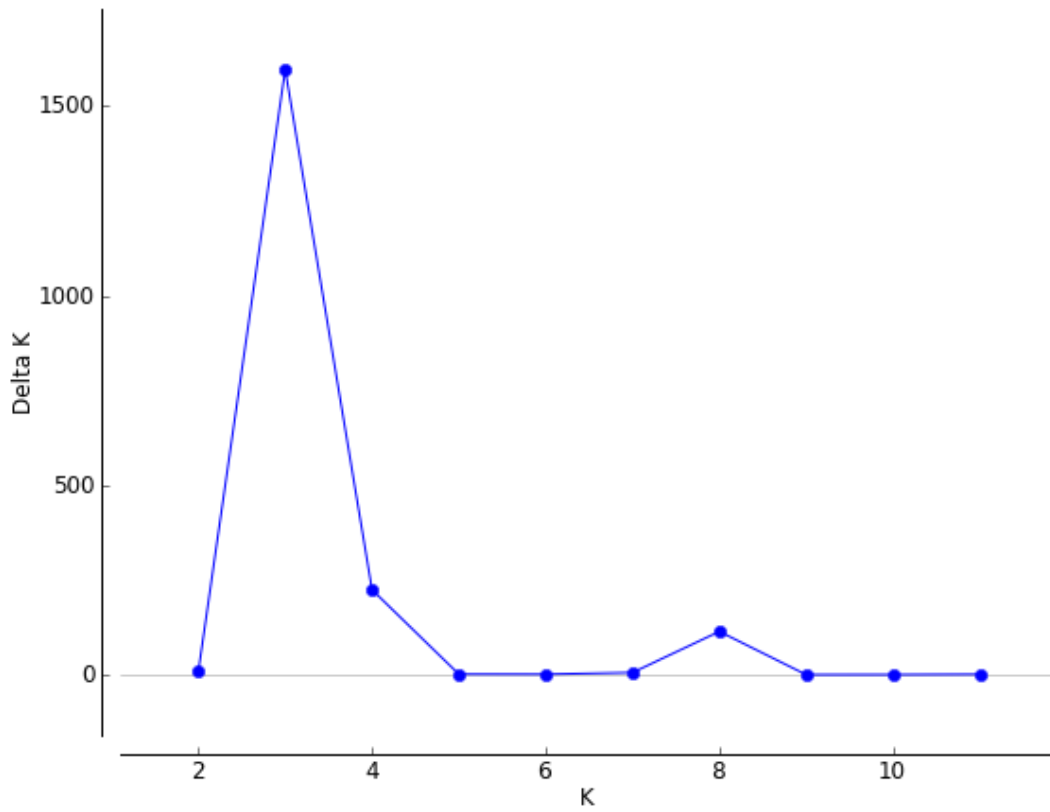
Figure 2.5 ΔK plot for set I based on STRUCTURE analysis of different numbers of subpopulations (K = 1 to 10) with DArT markers
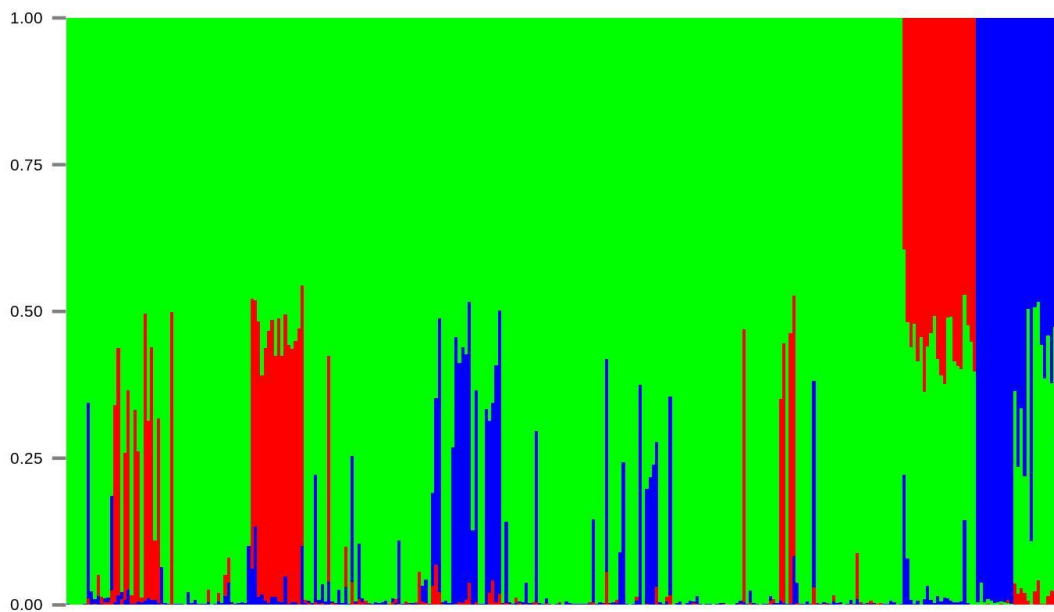
Figure 2.6 Membership probability plot for subgroup number K = 3 in set I
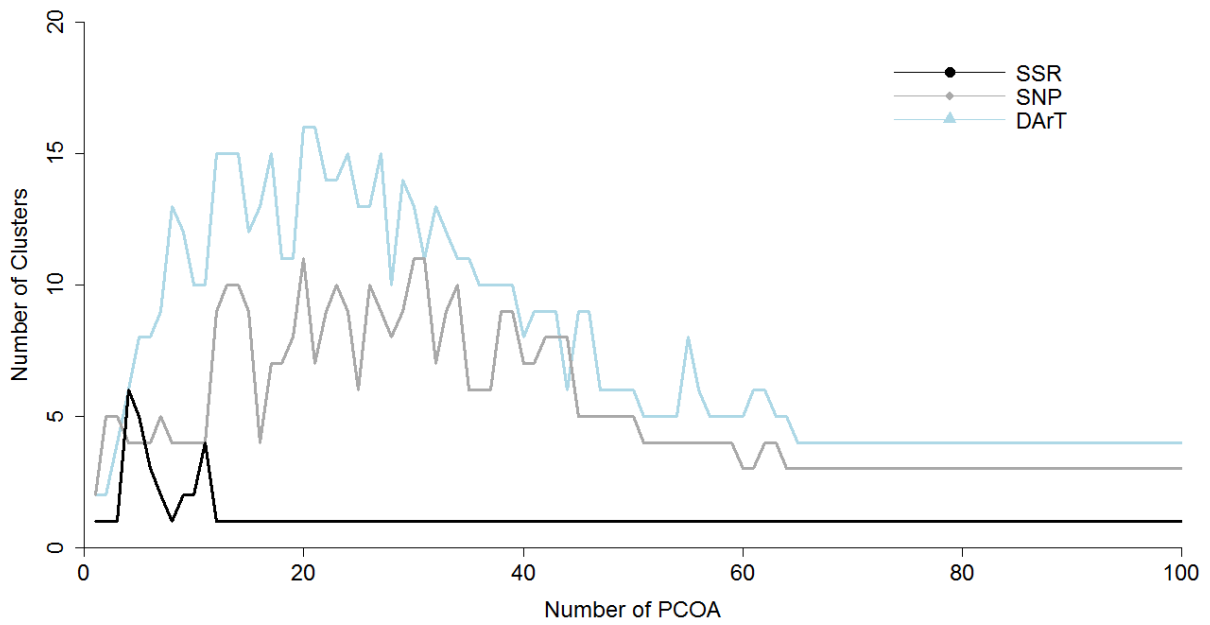
Figure 2.7 Number of Clusters identified by PCo based clustering with increasing numbers of PCos applied in the model
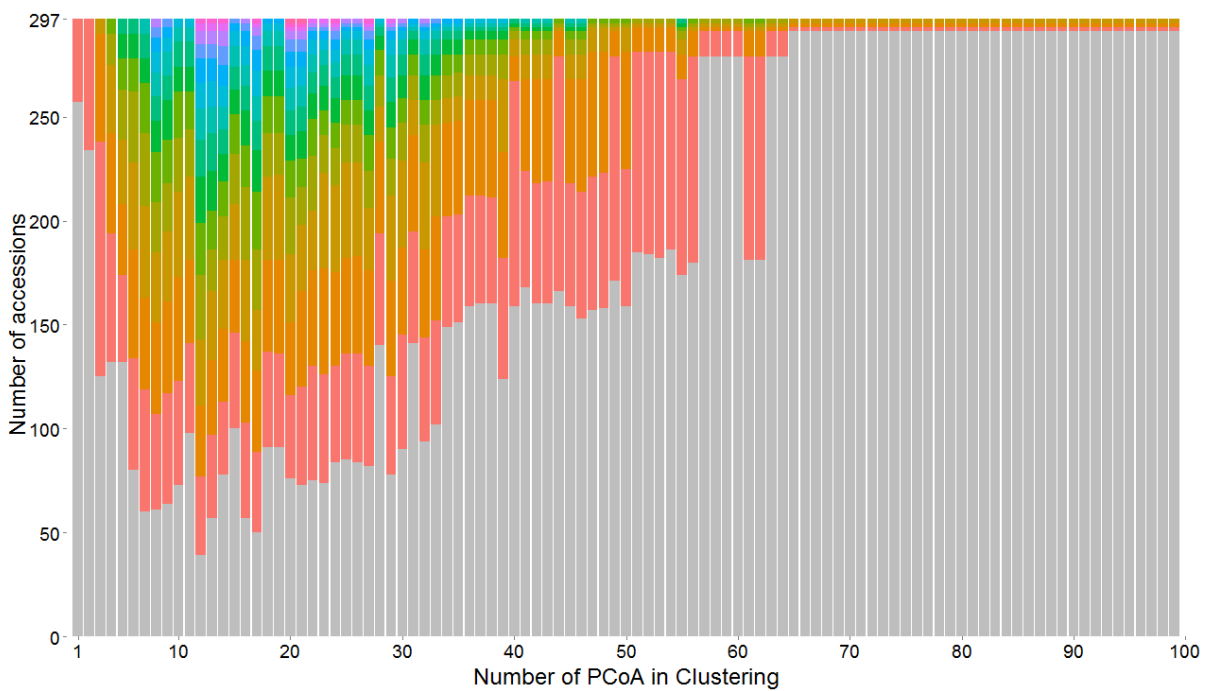
Figure 2.8 Barplot showing the assignment of subgroups with increasing numbers of PCos applied in the model
(a) for DArT markers. Each color represents a different group.
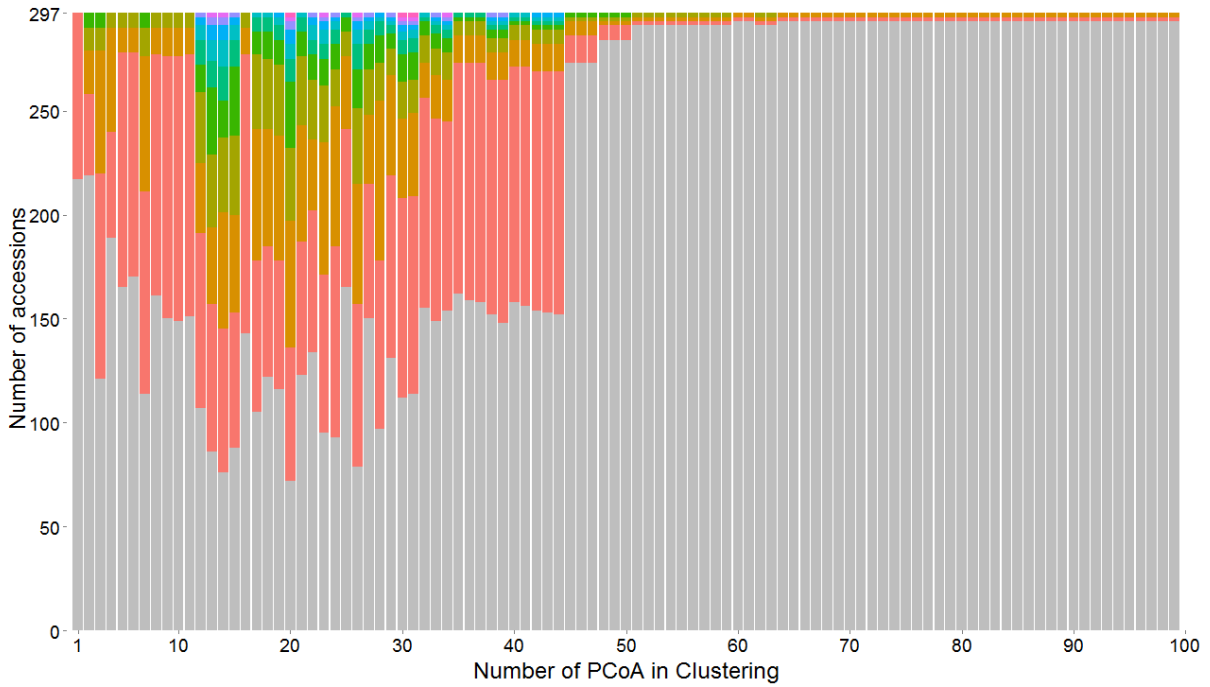
Figure 2.8 Barplot showing the assignment of subgroups with increasing numbers of PCo applied in the model
(b) for SNP markers. Each color represents a different group.
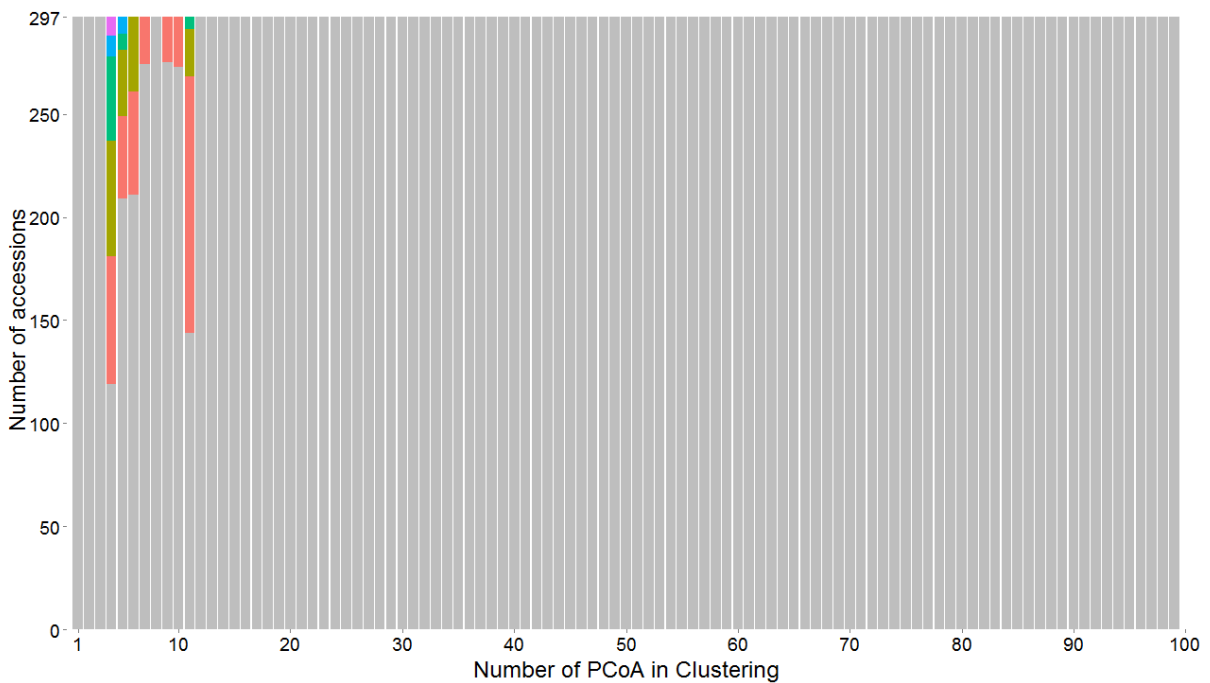
Figure 2.8 Barplot showing the assignment of subgroups with increasing numbers of PCo applied in the model
(c) for SSR markers. Each color represents a different group.
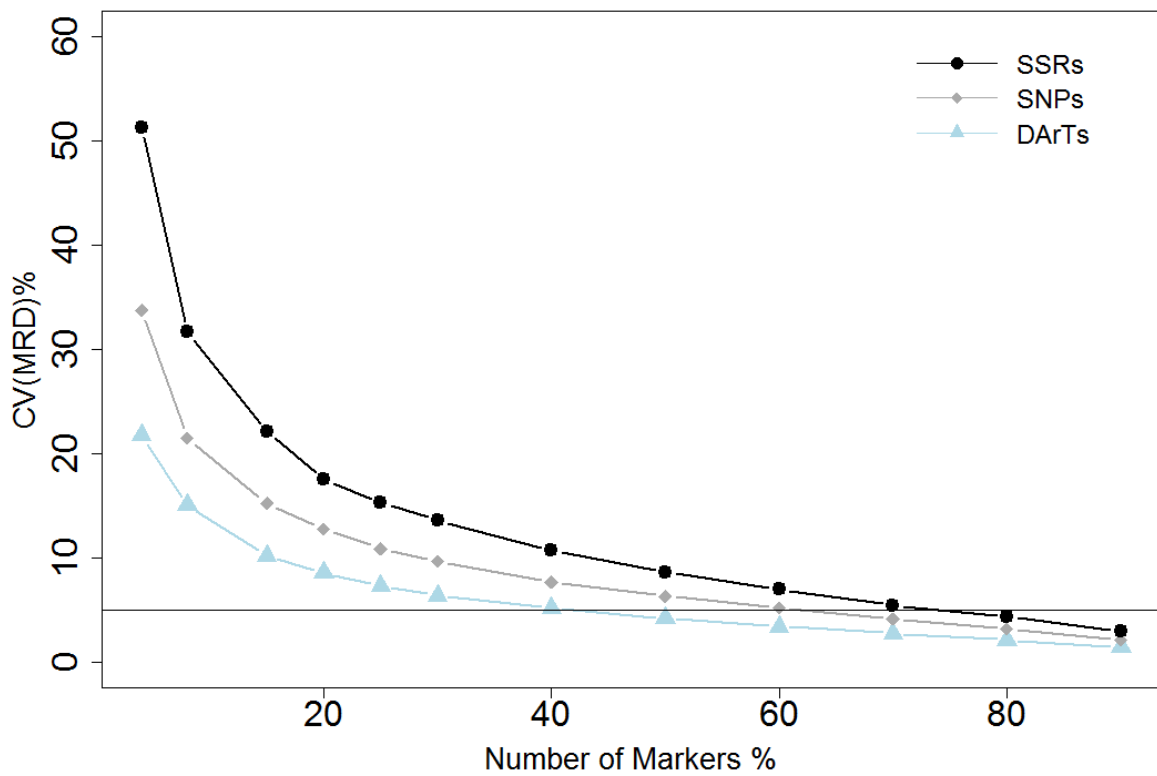
Figure 2.9 Combined bootstrapping analyses of DArT, SNP and SSR markers
Mean coefficient of variation (CV) of corresponding distance estimates over 100 replications was calculated by resampling a certain percentage of the entire marker dataset. The *horizontal line* indicates the threshold of 5% in CV.

2.4 Discussion

2.4.1 The feasibility of bulk sampling in genetic variation detection in *Lolium perenne* L.

In the current study, a large germplasm set was genotyped with bulk sampling. Because the ability to provide consistent distinctiveness between accessions is of fundamental importance for diversity and population structure study, a special set of samples (set III) was used to inspect the feasibility of bulk sampling strategy. In the phenograms on entire sample sets with set III being highlighted, GR2725, GR2915, GR3107 and GR8502 formed their own distinct clusters (Fig. 2.2), therefore all three marker systems should be suitable for diversity study with bulk sampling in perennial ryegrass. Our finding is consistent with Guthridge et al. (2001) who compared the discriminative capability in distinguishing cultivars with multiple individual samples and bulked sample (20 individuals/bulk) with AFLP markers and concluded that the results from bulk samples were consistent with those from individual samples.

In the phenogram, these four accessions from set I under common bulk sampling process with 30 seedlings per sample were also highlighted (Fig. 2.2). As a result, three of them did not show pronounced deviation from the clusters formed by set III, but GR8502 was the only exception. This phenomenon was observed for all the marker types. Random variation during sampling process could cause this disagreement, but since in each sample 30 individuals were pooled, such explanation might not be satisfactory. In addition, none of the other samples with high sampling size displayed such a deviation. Therefore it is suspected that probably some mistakes during the sampling process were made.

An appropriate sampling size is essential for the success of bulk sampling strategy. Owing to a highly efficient self-incompatibility reproduction system, a high level of within population variation compared to variation among populations is anticipated (Kölliker et al. 2005). In the present study, within population variation can be highly reflected in the set III with only single individual: they were either located outside of their corresponding groups or at the boundary (Fig. 2.2).

In order to further investigate the effect of bulk size, we divided each accession into groups with large bulk size vs. small bulk size in the PCoA on set III (Fig. 2.3). Samples containing 1, 12, 24 individuals were defined as small bulk sample and samples containing over 24 individuals were defined as high bulk samples. As expected, a clear trend was identified that bulks with higher individual numbers are more constant than bulks based on less individuals and it holds true for all three marker types. Like in phenograms, the extremes were always observed for samples containing only one individual. These samples could be very different from their corresponding group, such as in the case of GR3107. Based on this result, a bulk size above 24 individuals should be generally required and sufficient for a reliable, bulk-based estimation of the population structure in *Lolium perenne*. The bulk size of 30 individuals used in set I was above this threshold therefore the bulking procedure should be appropriate for our purpose. Related studies showed that although a small bulk with 3-5 individuals is appropriate for minor allele detection(Gilbert et al. 1999), 20-30 individuals per bulk are required for a reliable identification of accessions or cultivars (Guthridge et al. 2001; Bolaric et al. 2005b). Our results support these previous studies.

For SSRs, 8.2 alleles per locus were found on average, which is lower than 9.9 (Wang et al. 2009), 13.3 (Brazauskas et al. 2011) and 19.4 (Kubik et al. 2001) in other studies with single seedling sample strategy, even though a much higher number of accessions were examined in our study. These results might be attributed to two reasons. Unlike genotyping for individual sample, multiple peaks for a certain primer pair are possible for bulk samples of ryegrass accession during SSR data generation. To more reasonably and accurately estimate allelic frequency, the weak peaks in the banding profiles were eliminated. This reduces the ability to detect rare alleles. In addition, the bulk sampling is not ideal for the rare allele identification due to the sensitivity of the system (Kirkpatrick et al. 1991), it is also suggested that rare alleles are not detected if they comprise less than 4% in the PCR products (Michelmore et al. 1991). To detect rare alleles and more accurately characterize a certain accession, multiple small-bulk samples or multiple single seedlings out of this accession should be genotyped (Kraft and Säll 1999). From this perspective, the bulk sampling strategy should not be treated as a counterpart of the single seedling genotyping strategy but rather as a complementary method for the genotypic characterization of highly heterogeneous material, like *Lolium perenne*.

2.4.2 Diversity and structural patterns within the *Lolium perenne* L. germplasm

Molecular marker revealed a high polymorphism rate in the examined germplasm set: for DArTs, SNPs and SSRs, 1.99, 2.00 and 8.20 alleles per locus in average were obtained across germplasm set I. The JD for DArT markers ranged from 0.00 to 0.73; the MRD for SNPs ranged from 0.03 to 0.52; the MRD for SSR markers ranged from 0.26 to 0.76. The distribution of the genetic distance estimates was bell-shaped and alike for all the marker types (Fig. 2.1). The allelic polymorphism and wide range of genetic distances for each marker type indicate high genetic variation in the germplasm collection.

Genetic diversity for dominant DArT markers was found to be 0.26; for co-dominant SNP and SSR markers it was 0.32 and 0.45 across set I, respectively. Hu et al. (2011) observed a genetic diversity of 0.28 within 75 perennial ryegrass accessions with dominant ISSR markers, which is similar to what we found for dominant DArT markers. Brazauskas et al. (2011) observed a genetic diversity of 0.63 employing SSR markers on 37 *Lolium perenne* accessions, which is higher than what we obtained with SSR markers. At first glance, none of these results should be expected because a higher number of accessions were analyzed in our study. However, the usage of bulked samples should be taken into regard because the aforementioned studies were conducted at the individual genotype level. It has been well documented that in *Lolium perenne*, higher variation lies within the accessions rather than between the accessions (Guthridge et al. 2001; Bolaric et al. 2005b; Elazreg et al. 2011). Consequently, the discriminative power as well as distinction between accessions would be reduced (Kölliker et al. 2005) because the common alleles are likely to increase among bulk samples (Guthridge et al. 2001). Only among accession variation was tested in our study, therefore a relatively lower genetic diversity should not be too surprising.

Further we subdivided the germplasm set according to the corresponding passport data and compared unique alleles, average alleles per locus and genetic diversity for each subgroup (Tab. 2.3). For geographical origin, Western and Northern Europe exhibited a higher diversity than the other regions. However, a clear conclusion is difficult to draw because the loss of diversity was coupled with lower number of accessions in the germplasm set for Eastern (10 accessions), Southern Europe (5 accessions) and Oceania (5 accessions). It is known that for highly heterogonous material the amount of samples is a significant factor for the

determination of diversity (Brazauskas et al. 2011). For ploidy levels, no distinction in terms of genetic diversity could be made between diploids and tetraploids, which might underpin the relation between diploid and tetraploid perennial ryegrass since the modern tetraploid material was deviated from diploid material by chemical treatment (Boller et al. 2010). In a study on both 2x and 4x material from the same gene pool, a lack of distinction between ploidy levels was also found by Roldàn-Ruiz et al. (2000). Our finding confirms this in a broader range of material. Not surprisingly, in respect to breeding status, varieties were found to possess a lower diversity than breeding material and ecotypes. Successive selections have to be conducted in a breeding programs to meet DUS (Distinctness, uniformity, stability) criteria. During this process, a certain amount of alleles is fixed and this might reduce the available diversity (Bolaric et al. 2005b). The genebank material included here did not add extra diversity, which might on the one hand prove the effective maintenance of the diversity by breeders (Bolaric et al. 2005b) and, on the other hand, provide evidence for the intensive usage of ecotypes in practical breeding work (Boller et al. 2010).

Finally we conducted STRUCTURE, PCo-based clustering and AMOVA to inspect the potential structures in set I. In STRUCTURE analysis based on DArT markers, although the ad hoc criteria suggested 3 subgroups in set I, only 15 accessions could be assigned to group 2 or group 3 when the probability threshold was assigned to 0.7 (Fig. 2.6). In PCo-based clustering analysis, DArT and SNP markers identified 4 and 3 subgroups when more than 65 PCos and 64 PCos were incorporated in the model; but the majority of the material was assigned to only one group (Fig. 2.7, Fig. 2.8(a), (b)). For SSRs, the potential group number was estimated to be one after including over 11 PCos in the model (Fig. 2.7, Fig.2.8(c)), suggesting no population structure. It is worthwhile to mention that the incorporation of more PCos might be required in this analysis due to the small portion of variance explained by the first two PCos. Among different marker types, AMOVA analysis provided incongruous results for different factors. For example, contributor explained 3.83% of the total variance which is the highest among all the factors for DArTs; however, ploidy level (6.39% of the total variance) and geographical origin (2.60% of the total variance) were the most influential factor for SNPs and SSRs, respectively. Despite of the inconsistency, none of the factors could explain the variation within set I to a relevant extent even though the variance components for all the factors were significant at p = 0.01.

Based on these results, there is little evidence supporting the existence of a strong structure in the tested European germplasm. In the analysis of a subset of 80 accessions of *Lolium perenne*, Calsyn et al. (2005) found similar results, i.e. geographical origins could only explain 3% of the total variation. In an association study with relatively limited or highly selected germplasm resources, a stronger structure based on origin might be identifiable (Skøt et al. 2007; Yu et al. 2011), but it is generally not expected due to the intensive usage of natural resources in breeding (Bolaric et al. 2005b; Brazauskas et al. 2011) and lack of maintenance of heterotic pools (Brummer 1999). Our finding is in congruence with those statements.

2.4.3   The comparison of marker systems

The application of different marker types to the same germplasm set provides opportunities to compare the performance of different marker types in diversity studies. In the present study, the estimated genetic diversity was highest for SSRs, followed by SNPs and DArTs. The discrepancies should be attributable to the inherent nature of the markers and the methods used to calculate the diversity. DArT and SNP markers could be located at both coding and non-coding regions therefore they should be less variable in comparison to SSRs mainly located in non-coding regions (Abdel-Mawgood 2012). Furthermore, Nei's diversity measurements would favor a multi-allelic marker system like SSRs over the bi-allelic SNP and dominant DArT marker types. This was also confirmed by Van Inghelandt et al. (2010) and Simko et al. (2012) in their genetic diversity study with different marker types. For multi allelic marker types like SSRs, the maximum diversity value could approach 1 if loci are highly polymorphic; for a bi-allelic marker system like SNPs, the theoretical maximal value of this measurement could be only 0.75 in the case of equal share of both alleles.

Unlike crops in which the pedigree information or the prior population structure could be inferred, a solid reference that could be used to compare the accuracy of accession assignment is not available in the present study. Therefore we defined three indirect criteria for the comparison: 1) the discriminative ability to distinguish the exotic material in set II; 2) the reproducibility to provide constant results for the replicated samples in set II; 3) the consistency to provide similar results when only subsets of the data are used in a

bootstrapping process.

Exotic material used to test the discriminative ability were composed of two samples of *L. multiflorum* variety (Gordo), one sample of a *L. x hybridum* ecotype (GR7418) and eight lines of the VrnA mapping population (Jensen et al. 2005) including two parents, two F1 and four F2 individuals. *L. x hybridum* is derived from a cross between *L. multiflorum* and *L. perenne*, so it is supposed that it should lie between its parental species in terms of population structure (Boller et al. 2010). Though the mapping population VrnA is generated from a cross between two *Lolium perenne* genotypes for the QTL analysis of vernalization mechanism (Jensen et al. 2005), the samples we used in the experiment were rather weak in the field and quite different from other perennial ryegrass accessions (personal communication by Dr. Feuerstein from DSV) therefore their true origin is unclear and discrepancies in molecular markers should also be anticipated due to the morphological differences. As results from PCoA of DArTs and SNPs (Fig. 2.4(a), (b)), two accessions of *L. multiflorum* and all VrnA entries are obviously separated from the other material; the single accession of *L. x hybridum* is intermediate between set I and Gordo, exactly as expected. For SSRs, however, the distinction was largely diminished especially for GR7418. DArTs and SNPs seem to outperforming SSRs regarding their differential potential.

Owing to the elimination of sampling effect, replicated sample in set II should reveal intrinsic reproducibility of the different marker systems. Here, we observed a high consistency of DArT markers. Within the seven replicated samples in set II, an average JD of 0.004 was obtained and around 99% of the variation could be explained by the replicated accessions indicating an excellent reproducibility and rather low systematic error (Fig. 2.1, Tab. 2.5). SSRs performed worst (average MRD 0.35 and only half explainable variance) and SNPs were moderately well (average MRD of 0.16 and around 70% explainable variance). These findings might help to explain the lower correlations between SSRs and the other two marker types. Although DArT, SNP and SSR markers were all reported as highly reproducible in many studies (Jones et al. 2007; Cruz et al. 2013), the estimation is mainly based on studies in individual genotypes which is different from pooled samples in the present study. The SSRs were scored manually and stutter peaks in the banding profile was commonly observed in this case (Berg et al. 2000). Moreover, for SSRs, the theoretical allele numbers for each locus could be very large for bulk samples. Therefore all the detected peaks, either normal or stutter,

should be recorded if there is no a solid reference band profile available. This might give rise to the occurrence of higher error rate in comparison to DArTs and SNPs.

In the bootstrapping analysis of the entire sample set, similar patterns among marker types were observed. The CV decreased fast when the number of markers to be resampled is small and gradually the decrease tends to flatten along with the increase of the number of markers. In all scenarios, DArTs performed superior than SNPs while SSRs always displayed relatively lower consistency. The decrease pattern observed in this study is similar to that described by Van Inghelandt et al. (2010) and Garcia et al. (2004). They found that after a certain threshold a further increase of the number of markers will only slightly influence genetic distance estimates. If we set the threshold to CV of 5% as the acceptable precision for genetic distance estimation, 554 DArT markers (40%), 110 SNP markers (60%) and 36 SSR marker loci (75%) are required. The number of SNPs required for reliable diversity estimation is in accordance with the study of Li et al. (2011) in sugar beet diversity study. The ratio between SNPs and SSRs (about 3:1), however, are much lower than the ratio of 7-11 times more SNPs than SSRs proposed by Van Inghelandt et al. (2010) in a study on maize inbred lines with 8244 SNPs and 359 SSRs. This might be attributed to high reproducibility in SNPs, the usage of bulked populations and a much lower number of SSR and SNP applied. In diversity study on sugar beet, Simko et al. (2012) suggested a ratio of 4.9-13.3 between DArTs and SSRs which is lower than what we have obtained. It appears that, in dealing with bulked sample, a higher amount of DArT markers is required to compensate the loss of information due to the dominant nature. Nevertheless, DArTs in the present study outperformed the other two marker types in CV simulation owing to immense number of markers used.

Despite of discrepancies among marker types, we have to stress that they differ mainly quantitatively, but not qualitatively in this study. All the marker systems provided similar evidence about the germplasm collections: certain amount of diversity and polymorphism rate, lack of structure, ability to distinguish accessions. Genetic distance estimates generated by different marker types are also significantly correlated with a high correlation coefficient. However, due to the higher consistency, better coverage of the genome and technical low dependence of the prior knowledge of the sequences, DArT markers appeared to be better suited to deal with diversity studies with bulked sampling in *Lolium perenne*.

2.5   Conclusion and further perspectives

With bulk sampling strategy, high level of genetic diversity was found within the germplasm set in the present study. However, based on clustering analyses as well as AMOVA on the passport data, a clear structure within the germplasm set was not found. All three marker types showed the capability to conduct diversity studies, although DArTs appeared to be superior in terms of discriminative ability, repeatability and consistency. With this knowledge, the diversity study could be conducted further with more genotypes from certain selected accessions of interest because there is still large extent of diversity harboring within each accession which cannot be detected by bulk sampling. In addition, the correlation between genetic distance and heterosis could be tested for further study and this will provide information about the potential usage of genetic distance in breeding programs.

Chapter 3

Breeding of *Lolium perenne* L. for Biogas Production: Heterosis of Biomass Yield and its Correlation with Genetic Distance

3.1   Introduction

The generation and consumption of sustainable energy obtains increasing attention in these years in Europe. Currently, 2% of the arable land was used for biogas plant cultivation in Germany (Ofori and Becker 2008). However, the majority of biomass used for biogas production originates from maize (Weiland 2007) for which no production of biomass substrate over the winter is possible (Ofori and Becker 2008). Other potential shortcomings for the overwhelming usage of maize lie in the compromise to food/feed production (Aguirre et al. 2012) and the environmental and economic risk (Salces et al. 2013). *Lolium perenne* L. might serve as an alternative or supplement for biomass production due to its high yield potential, persistence, easy management and lower input requirement (Lewandowski et al. 2003; Searchinger et al. 2008; Aguirre et al. 2012).

For this purpose, the improvement of yield related traits are of fundamental importance. However, owing to the strictly outcrossing nature as well as the existence of a self-incompatibility system controlled by two loci (Cornish et al. 1979), the common breeding practice for perennial ryegrass is still relying on either population improvement or construction of synthetics which only partially use the potential heterosis (Aguirre et al. 2012). Meanwhile, longer breeding cycles (Casler and Brummer 2008) as well as poor consistence between spaced plant evaluations used in the selection and competitive sward condition in the actual cultivation (Wilkins and Humphreys 2003) set further hindrance for yield improvement. As a consequence, the average genetic gain in yield was rather limited: since early 20th century, the advances of annual dry matter yield due to breeding were estimated to be solely 3.8% per decade (Humphreys 2005). With this background, novel techniques or breeding schemes are required to produce highly productive cultivars.

Considering the success of yield improvement in many other crops like maize (Duvick 1992), a better exploitation of heterosis for perennial ryegrass was proposed since many years (Vogeland and Pedersen 1993). Many studies were therefore conducted in an attempt to assess the hybrid performance and heterosis under spaced plant or sward conditions in ryegrass (Foster 1971a; Foster 1971b; Foster 1973; Posselt 2010). Noteworthy, the population hybrids constructed in these studies usually contained certain amounts of intra-population progenies due to the lack of pollination control. Chance hybrids (Burton 1948) or

semi-hybrids (Brummer 1999) might be more appropriate description in this situation. Because the ratio of inter and intra population progenies is generally unknown, semi-hybrids might limit the manifestation of the full heterosis. To better harness the hybrid vigor, a manipulation of the pollination process is needed but difficult to conduct in perennial ryegrass with artificial emasculation (Aguirre et al. 2012).

For a better pollination control, CMS (cytoplasmic male sterility) systems could be used. They have been proven to be efficient to produce hybrid seeds in many major crops like maize, rice, rapeseed, wheat etc. The initial CMS reported in perennial ryegrass was generated by an intergeneric cross with *Festuca pratensis* and an interspecific cross with *Lolium multiflorum* (Wit 1974). However, instability in pollen fertility (Boller et al. 2010) and the long time required to introduce CMS into advanced breeding material (Islam et al. 2014) constrained its application. Recently, chemically induced CMS was successfully implemented in perennial ryegrass (Gaue and Baudis 2006) and different molecular mechanisms causing CMS were identified (Islam et al. 2014). It can be thereby expected that CMS material will be more accessible in a near future.

Another aspect regarding the generation of elite hybrids is the identification of heterotic pools. Brummer (1999) pointed out the importance of maintaining heterotic groups and indicated how molecular markers are able to delimit them. With heterotic patterns being rarely studied like in perennial ryegrass, it was suggested that, genetic distances based on molecular markers could be used to provide pre selections of hybrid parents (Boller et al. 2010). A theoretical correlations between genetic distance and mid-parent heterosis was suggested by Melchinger (1999). However, to our knowledge, though many studies investigated the genetic diversity in *Lolium perenne* L. with various marker types (Kubik et al. 2001; Kopecky et al. 2009; Elazreg et al. 2011; Hu et al. 2011), heterosis and their relationships with molecular markers were not reported.

In the current study, *Lolium perenne* L. population hybrids were constructed using CMS. Their corresponding parental populations were genotyped with DArT, SNP and SSR markers. Thereafter, both parents and F1 hybrids were phenotyped under competitive sward condition. The major objectives of this study were: 1) to evaluate the genetic diversity within the collected germplasms by molecular markers and compare different marker types; 2) to

estimate the extent of population heterosis for fresh matter and dry matter yield in hybrids produced using CMS; 3) to estimate the correlation between genetic distances and heterosis which might benefit further breeding work.

3.2 Material and Methods

3.2.1 Parental accessions and crosses

A CMS system was employed to construct the hybrid populations. In total, 76 accessions were tested as pollinators and nine CMS lines were utilized as maternal material. Their corresponding passport data summarizing geographical origin, contributor, ploidy levels and breeding status are given in Tab. 3.1. Parental material and their resultant hybrids were sowing in 2010 and 2011 independently. For the 2010 sowings, 55 F1 hybrids (29 diploids and 26 tetraploids) were produced from 35 pollinators (29 diploids and six tetraploids) and eight CMS accessions (but only two diploid and four tetraploid CMS accessions per se were tested in the field). For the 2011 sowings, all 48 F1 hybrids were diploids and they were composed of crosses between 45 pollinators and four CMS accessions (among them, three were tested in the field). In the 2010 and 2011 trials, 4 and 5 cultivars, respectively, were planted as checks. For the diploids, the majority of the hybrids were generated by crossing diverse pollinators with one of the three CMS accessions: CMS_236, CMS_237, CMS_238; for tetraploid hybrids, all three CMS accessions (CMS_234, CMS_235 and CMS_239) were crossed with six common pollinators (NPZ_FNR1_(2010), Maint._msF1_384, Syn_06_4503, PC_09_4902, Syn_08_4606, Syn_08_4607). The hybrids as well as pollinators tested in both sowings were different therefore these two sowings were analyzed as different experiments. It is noteworthy that the parental material used in the field trial was selected prior to the genotyping.

3.2.2 Field experiments

Sown in late 2010 and 2011, F1 hybrids and their parents were planted under competitive sward conditions at five locations, namely Hof Steimke, Malchow, Moosburg, Steinach and Ven Zelderheide using a randomized complete block design with two blocks at each location. Here, two different nitrogen levels - 100% nitrogen input as common agricultural practice and 60% of the common nitrogen level – were applied. Due to the perennial nature, each entry was harvested and evaluated in two consecutive years, that is, in 2011 and 2012 for the 2010 sowings and in 2012 and 2013 for the 2011 sowings. The number of cuts per growth year ranged from 3 to 5 depending on the growth situation at different locations. The harvesting was conducted by plot harvester with a Near Infrared Spectroscopy (NIRS) machine equipped on it. Two phenotypic traits were recorded: fresh matter yield (FMY) and dry matter yield (DMY). DMY was obtained from the product of FMY and dry matter content, determined by NIRS estimation based on subsamples of every harvested plot.

3.2.3 Molecular markers

Parental material used for genotyping were planted and maintained at the IPK Genebank. 100 mg fresh leaf tissue per plant was obtained after the harvest of the leaves. 30 leaves were pooled for each accession to form bulked samples, which could promise an adequate representativeness of each accession. Subsequently parental bulked samples were genotyped by 1384 DArT (Diversity Array Pty Ltd., Canberra, Australia), 182 SNP (Traits Genetics GmbH, Gatersleben, Germany) and 48 SSR marker loci in parallel. For SNP and SSR markers, allelic frequency data were recorded, while for DArT markers, presence/absence of the alleles was scored. For details of the molecular analyses, see 2.2.2

3.2.4 Statistical analysis

Genetic distances were calculated for all pairwise combinations of parental material. For co-

dominant SNP and SSR markers, Modified Roger`s Distance (MRD) was calculated while for dominant DArT markers, Jaccard`s Distance (JD) was estimated (Reif et al. 2005). Mutual correlations between all pairs of distances by different marker systems were evaluated by Pearson`s correlation coefficient. Based on genetic distance, Principal Coordinate Analysis (PCoA) was conducted with PCo1 and PCo2 for each marker system.

FMY and DMY were analyzed for variance components, heritability, entry means by software Plabstat version 3A (Utz 2011). Analysis of variance was calculated with the following model:

$$y_{ijk} = \mu + g_i + l_j + n_k + gl_{ij} + gn_{ik} + ln_{kj} + e_{ijk}$$

Where y_{ijk} is the observation of the genotype i at location j with the nitrogen level k; μ denotes the general mean; g_i represents the effect of genotype i; l_j is the effect of location j; n_k indicates the effect of nitrogen level k; gl_{ij}, gn_{ik}, ln_{kj} depict the genotype × location, genotype × nitrogen, location × nitrogen interaction; e_{ijk} characterize the residual error. Broad sense heritability (h^2) was calculated according to Hill et al. (1998). Pearson`s correlation coefficient was calculated among traits.

With the fitted parental yield and their resultant hybrid performance, panmictic Mid-parent heterosis (MPH) was calculated with the following formula (Lamkey and Edwards 1999):

$$\text{MPH} = \left(\frac{F1 - \mu_{parents}}{\mu_{parents}}\right) \times 100\ \%$$

Also panmictic Better-Parents Heterosis (BPH) was calculated as:

$$\text{BPH} = \left(\frac{F1 - P_{better\ parents}}{P_{better\ parents}}\right) \times 100\ \%$$

Where F1 represents the hybrid population performance, $\mu_{parents}$ denotes the average of the parents, $P_{better\ parents}$ indicates the performance of better parents. Pearson`s correlation was calculated between genetic distance of the parents and MPH as well as F1 hybrid performance.

MPH, BPH calculation, genetic distances estimation, correlation analysis as well as graphical presentation were all conducted on R platform (R Core Team 2013).

Table 3.1 Passport data for parental accessions of *Lolium perenne* L. utilized in this study

	Origin	Breeding_Status	Ploidy	Parental Status	Year of Sowing
Urspr.klon_CMS_234	DNK	Breeding Material	4x	CMS	2010
Urspr.klon_CMS_235	DNK	Breeding Material	4x	CMS	2010
Urspr.klon_CMS_233	NLD	Breeding Material	4x	CMS	2010
CMS_237	UNK	Breeding Material	2x	CMS	2010&2011
Urspr.klon_CMS_239	BEL	Breeding Material	4x	CMS	2010
CMS_213	UNK	Breeding Material	2x	CMS	2010&2011
2030872	DEU	Breeding Material	2x	Pollinator	2010
2030038	DEU	Breeding Material	2x	Pollinator	2010
Urspr.klon_CMS_231	FRA	Breeding Material	4x	CMS	2010
2030323	DEU	Breeding Material	2x	Pollinator	2010
2030350	DEU	Breeding Material	2x	Pollinator	2010
2030367	DEU	Breeding Material	2x	Pollinator	2010
CMS_236	UNK	Breeding Material	2x	CMS	2010&2011
2020548	DEU	Breeding Material	2x	Pollinator	2010
2030830	DEU	Breeding Material	2x	Pollinator	2010
2030926	DEU	Breeding Material	2x	Pollinator	2010&2011
2030337	DEU	Breeding Material	2x	Pollinator	2010
2030117	DEU	Breeding Material	2x	Pollinator	2010
2020795	DEU	Breeding Material	2x	Pollinator	2010
2040391	DEU	Breeding Material	2x	Pollinator	2010
106232	DEU	Breeding Material	2x	Pollinator	2010
2040121	DEU	Breeding Material	2x	Pollinator	2010
2040832	DEU	Breeding Material	2x	Pollinator	2010&2011
NPZ_FNR_1_(2010)	UNK	Breeding Material	4x	Pollinator	2010
Maint._MSF1_384	UNK	Breeding Material	4x	Pollinator	2010
Syn_06_4503	UNK	Breeding Material	4x	Pollinator	2010
PC_09_4902	UNK	Breeding Material	4x	Pollinator	2010
Syn_08_4606	FRA	Breeding Material	4x	Pollinator	2010
Syn_08_4607	UNK	Breeding Material	4x	Pollinator	2010
S4	FRA	variety	2x	Pollinator	2010
S11	DEU	Breeding Material	2x	Pollinator	2010
S12	DEU	Breeding Material	2x	Pollinator	2010
S14	DEU	Breeding Material	2x	Pollinator	2010
S15	DEU	Breeding Material	2x	Pollinator	2010
S21	DEU	Breeding Material	2x	Pollinator	2010
S22	DEU	Breeding Material	2x	Pollinator	2010
S23	DEU	Breeding Material	2x	Pollinator	2010
S26	DEU	Breeding Material	2x	Pollinator	2010
S35	DEU	Breeding Material	2x	Pollinator	2010
S40	DEU	Breeding Material	2x	Pollinator	2010
Syn_06_4505	UNK	Breeding Material	4x	Pollinator	2010

2030377	DEU	Breeding Material	2x	Pollinator	2010
2040371	DEU	Breeding Material	2x	Pollinator	2010
2090503	DEU	Breeding Material	2x	Pollinator	2011
GR8419	NLD	variety	2x	Pollinator	2011
GR7867	NLD	variety	2x	Pollinator	2011
2060497	DEU	Breeding Material	2x	Pollinator	2011
2060480	DEU	Breeding Material	2x	Pollinator	2011
2060123	DEU	Breeding Material	2x	Pollinator	2011
2060072	DEU	Breeding Material	2x	Pollinator	2011
2090516	DEU	Breeding Material	2x	Pollinator	2011
2090502	DEU	Breeding Material	2x	Pollinator	2011
2062153	DEU	Breeding Material	2x	Pollinator	2011
2062148	DEU	Breeding Material	2x	Pollinator	2011
2060452	DEU	Breeding Material	2x	Pollinator	2011
2060005	DEU	Breeding Material	2x	Pollinator	2011
GR3511	CZE	ecotype	2x	Pollinator	2011
GR3107	GBR	variety	2x	Pollinator	2011
GR2704	DNK	variety	2x	Pollinator	2011
GR2725	GBR	variety	2x	Pollinator	2011
GR2910	FIN	Breeding Material	2x	Pollinator	2011
GR3084	FRA	variety	2x	Pollinator	2011
GR3236	BEL	variety	2x	Pollinator	2011
GR3243	NLD	variety	2x	Pollinator	2011
GR3352	DEU	ecotype	2x	Pollinator	2011
GR8420	FRA	ecotype	2x	Pollinator	2011
GR5646	UNK	ecotype	2x	Pollinator	2011
GR3467	DEU	ecotype	2x	Pollinator	2011
GR3122	RUS	landrace	2x	Pollinator	2011
GR2929	RUS	Breeding Material	2x	Pollinator	2011
GR2915	GBR	landrace	2x	Pollinator	2011
GR2859	DEU	ecotype	2x	Pollinator	2011
2011-13_1	FRA	variety	2x	Pollinator	2011
2011-14_1	DEU	Breeding Material	2x	Pollinator	2011
2011-15_1	DEU	Breeding Material	2x	Pollinator	2011
2011-16_1	GBR	variety	2x	Pollinator	2011
Aberavon	GBR	variety	2x	Pollinator	2011
GR5100	DEU	ecotype	2x	Pollinator	2011
GR5015	DNK	variety	2x	Pollinator	2011
GR8611	FRA	ecotype	2x	Pollinator	2011
GR8502	IRL	variety	2x	Pollinator	2011
Urspr.klon_CMS_238	NLD	Breeding Material	2x	CMS	2011
PC_08_2902	ITA	ecotype	2x	Pollinator	2011
2060927	DEU	Breeding Material	2x	Pollinator	2011
2060956	DEU	Breeding Material	2x	Pollinator	2011

UNK: unknown

3.3 Results

3.3.1 Molecular characterization of parental material

The allelic polymorphisms and genetic distances provide information about the diversity of the parental material used in the field trial. Of 1384 DArT marker loci, 1310 were polymorphic indicating an average of 1.9 alleles per locus; all SNP marker loci were polymorphic; a total of 313 unique alleles were detected in 48 SSR marker loci suggesting an average of 6.5 alleles per locus. JD for DArT markers ranged from 0.13 to 0.59 with a mean of 0.40; MRD for SNP and SSR markers were in the range of 0.16 to 0.48 and 0.32 to 0.72 with the mean of 0.31 and 0.50, respectively. The distribution of genetic distance estimates was bell-shaped for all marker types (Fig. 3.1). In the hybrids, the genetic diversity could not be completely exploited because only CMS accessions were used as pollen recipient to control the pollination process so they are not freely selectable. Taking this into account, the theoretically possible range of the genetic distance was 0.25 - 0.57, 0.18 - 0.48, 0.37 - 0.67 for DArT, SNP and SSR markers, respectively (Tab. 3.2). It could be found that, despite of fact that hybrids was produced prior to the genotyping of the parents, the range of the selected parents was similar to theoretical values.

PCoA plots with two dimensions for DArT, SNP and SSR markers based on their corresponding genetic distance matrices are shown in Fig. 3.2. Approximately only 10% of the variance could be explained by the first two dimensions (5.51% and 4.25% for DArTs, 4.97% and 4.21% for SNPs, 5.35% and 4.35% for SSRs). In SNPs, a clear segregation of diploids and tetraploids could be observed (Fig. 3.2 (b)), for the two other marker types, neither a clear segregation nor subgroups were observed. The correlation coefficients between genetic distances estimates were 0.74 for DArT and SNP markers, 0.56 for DArT and SSR markers and 0.60 for SNP and SSR markers, all of these coefficients were significant at p = 0.01.

3.3.2 Genetic variation and variance components

Tab. 3.3 reveals the yield performance under different nitrogen levels, with reduced nitrogen input significantly reducing yield performance. Tab. 3.4 summarizes the FMY and DMY in the 2010 and 2011 sowings for all the tested material. In 2010, due to the differed performance in FMY, diploids and tetraploids were separated. For diploids, the FMY ranged from 100.35 t/ha to 137.21 t/ha; for tetraploids, a higher FMY from 126.35 t/ha to 156.99 t/ha was observed. In diploids, the DMY ranged from 25.01 t/ha to 29.96 t/ha, in tetraploids, the range of DMY was 26.65 t/ha to 31.21 t/ha. In spite of remarkable differences in FMY, tetraploids only slightly outperformed diploids in DMY. In the 2011 sowings, only diploid material was cultivated. The FMY was between 92.83 t/ha and 132.10 t/ha; the DMY ranged from 21.86 t/ha to 27.57 t/ha.

Broad-sense heritability was 0.67 (diploids 2010), 0.88 (diploids 2011) and 0.70 (tetraploids 2010) for FMY and 0.62(diploids 2010), 0.84 (diploids 2011) and 0.68 (tetraplodis 2010) for DMY, suggesting a relatively high repeatability and reliability of the measurements (Tab. 3.4). The effect of locations, nitrogen levels and genotypes were all found to be highly significant. Interaction between locations × nitrogen and genotypes × locations were also significant at $p = 0.01$ and their variance components were in a similar range as the variance components of genotypes. Inconsistent significance results were acquired for the effect of genotype × nitrogen for both sowings. Their corresponding variance components, however, were clearly smaller compared with the genotypic variance components.

3.3.3 Correlation between yield data

Because of its perenniality, performance across years is important for the evaluation of *Lolium perenne* materials. In this study, FMY and DMY are used as the sum of yield for two growth years. The final yield is the sum of 1Y (total yield in the first growth year) and 2Y (total yield in the second growth year). Their mutual correlations are shown in Tab. 3.5. Though all correlations were significant, discrepancies among correlation coefficients existed. Some similar patterns were obtained for both sowings, for example, extremes were

always observed between 1Y and total yield; correlations among DMY were usually lower than among FMY. Specially, correlation between DMY and FMY (0.73) for diploids were relatively low for the 2010 sowings compared to the 2011 sowings (0.90).

3.3.4 Heterosis and the correlation between parental genetic distances, heterosis and F1 hybrid performance

For the calculation of heterosis, both parents are required. Due to the limitation of CMS seeds, however, some CMS material was either not available or not sufficient for all the location × nitrogen combinations. This resulted in a reduction of the hybrids that are available for the heterosis estimation. Finally, only 27 and 39 valid hybrids were obtained for the 2010 sowings and the 2011 sowings, respectively. For the 2010 sowings, 8 diploid hybrids and 19 tetraploid hybrids were further separated.

For tetraploids from the 2010 sowings, because each CMS accession was crossed with same paternal accessions, GCA (General combining ability) of these parental accessions could be calculated and is shown in Table 3.6. Contradictory results for GCA were observed: for FMY, GCA of the pollinators was more variable than that of CMS accessions; for DMY, CMS accessions showed larger effects and higher variation. But the absolute GCA value and variation of the parental material in comparison with the mean performance were limited for both traits. The mutual correlation coefficient between F1 hybrid per se performance (HP), Mid-parent heterosis (MPH) and Mid-parent performance (MPP) revealed a significant positive correlation between HP and MPH, a significant negative correlation between MPH and MPP, and a non-negative correlation between HP and MPP (Tab. 3.7).

Apart from CMS_234×Syn_06_4503, CMS_234×Syn_08_4606, where a slight lower yield in comparison to mid-parent performance was detected for FMY, positive MPH was revealed by almost all the F1 hybrids from the 2010 sowings with a mean of 13.88% (9.41% - 19.58%) for diploids and 6.85% (-1.16% - 14.31%) for tetraploids (Fig. 3.3(a), Tab. 3.8). In the 2011 sowings, MPH ranged from -3.99% to 10.40% with a mean of 3.24%. In BPH both the ratio of superior hybrids and the mean heterosis was reduced: for diploids, BPH ranged from -2.31% to 14.24% with a mean of 3.23% (2010 sowings) and from -8.82% to 7.84% with an average

of -1.10% in the 2011 sowings. For tetraploids, BPH ranged from -4.2% to 13.91% with an average of 4.94%. For MPH and BPH in DMY, the patterns were similar to that in FMY (Fig. 3.3(b), Tab. 3.8). Moreover, for both sowings, several hybrids were higher yielding than the best check cultivars (Fig. 3.3).

Tab. 3.9 summarizes the correlation of genetic distance with heterosis and F1 performance. For diploids, the correlation between genetic distance and heterosis as well as F1 performance was positive in most cases. The most prominent correlations were found between MRD of SNPs and MPH for FMY (0.47) and DMY (0.60) in the 2010 sowings. Unfortunately, because there were only eight hybrids available, these correlations were not significant at p = 0.05. Significant correlations at p = 0.05 were found only in the 2011 sowings, including correlation between MPH and GD for all the marker types, correlation between F1 performance and GD for DArTs and SNPs in FMY, correlation between MPH and GD for SSRs and correlation between F1 performance and GD for DArTs and SSRs in DMY. However, all the correlation coefficients were below 0.4. For tetraploids, the lack of correlation with genetic distance was generally observed. In most cases, SNP markers provided comparatively high correlation coefficients, but these advances were rather small. Fig. 3.4 graphically shows these correlations. It can be seen that in the 2011 sowings where significant correlation was found, the majority of the tested hybrids were composed of parents with moderate genetic distances; only the parents of three hybrids were located at the high GD range.
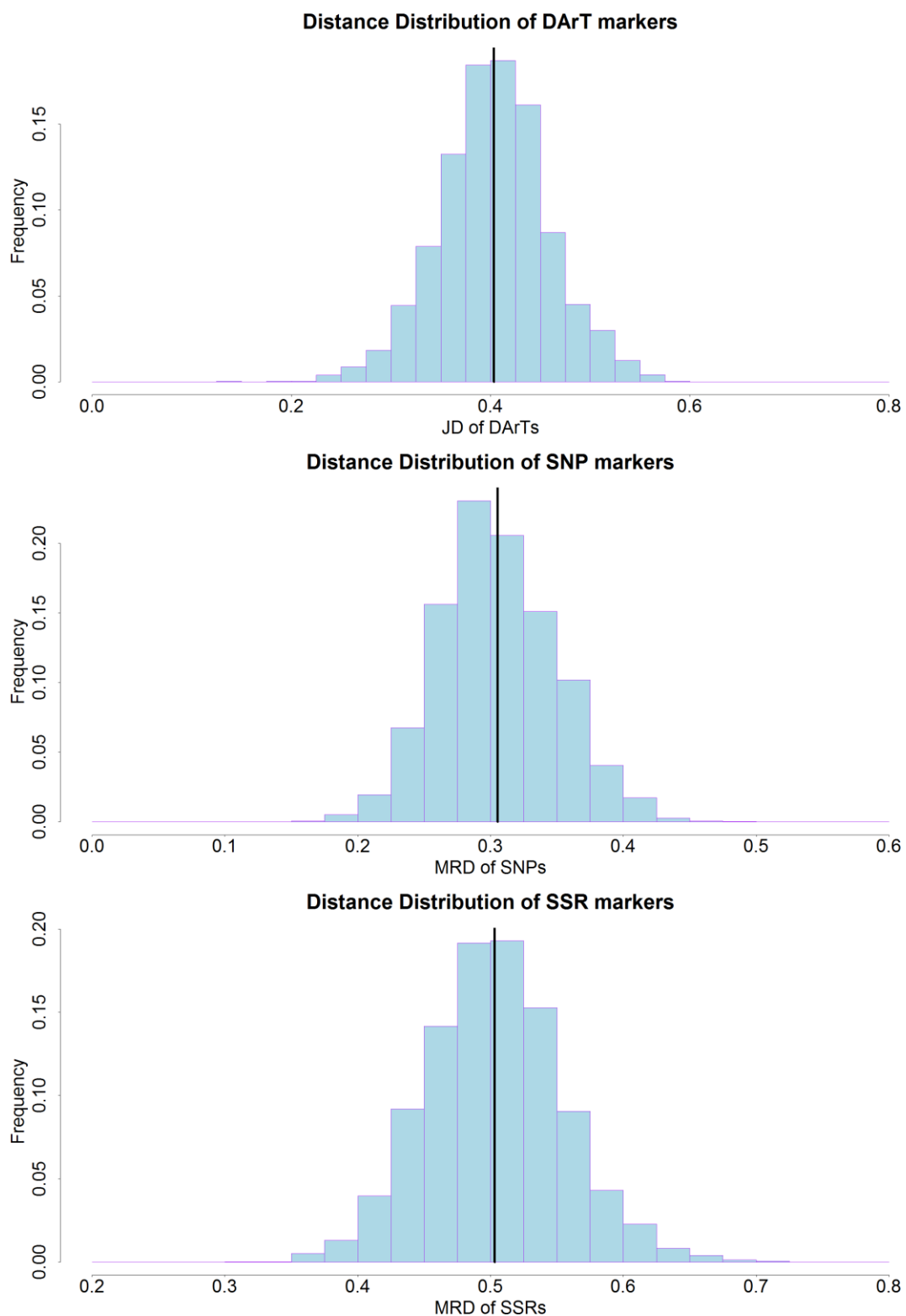
Figure 3.1 Distribution of genetic distances for DArT, SNP and SSR markers for parental material tested in the field

Vertical bar denotes the mean value of the corresponding genetic distance.

PCoA based on DArT markers
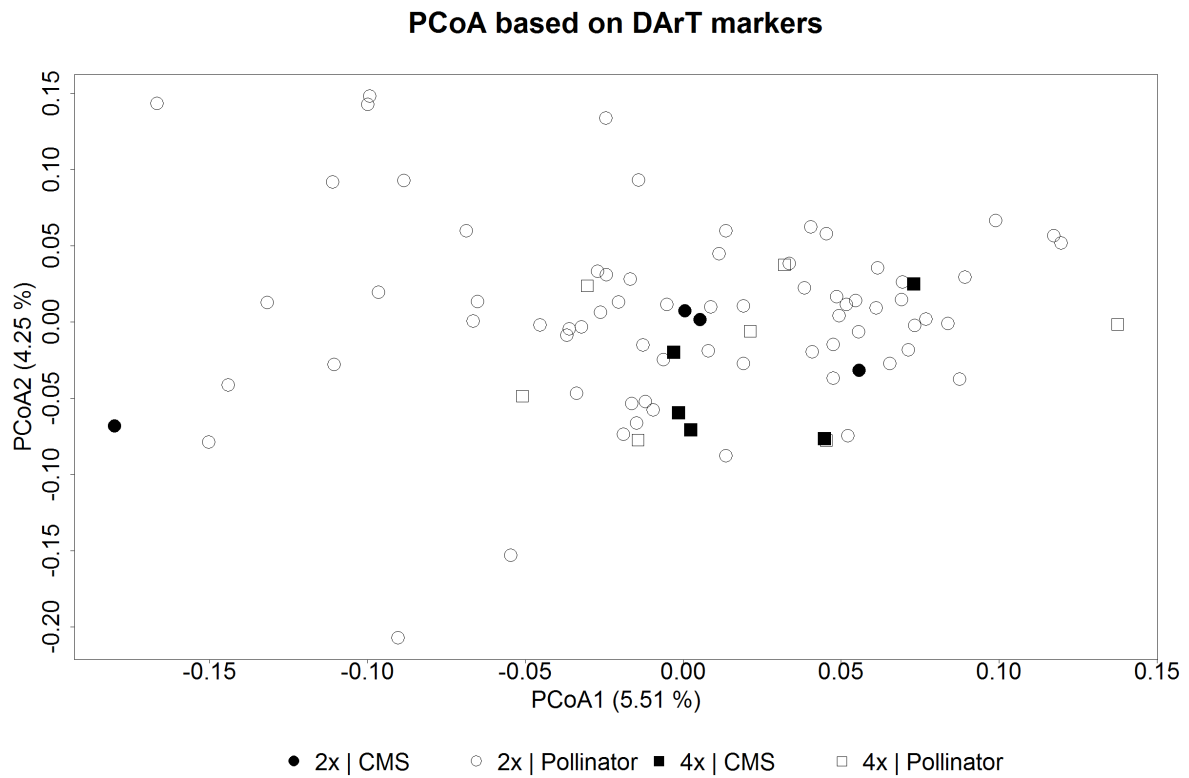
Figure 3.2 PCoA for parental material used in the field experiment
(a) based on DArT markers

PCoA based on SNP markers
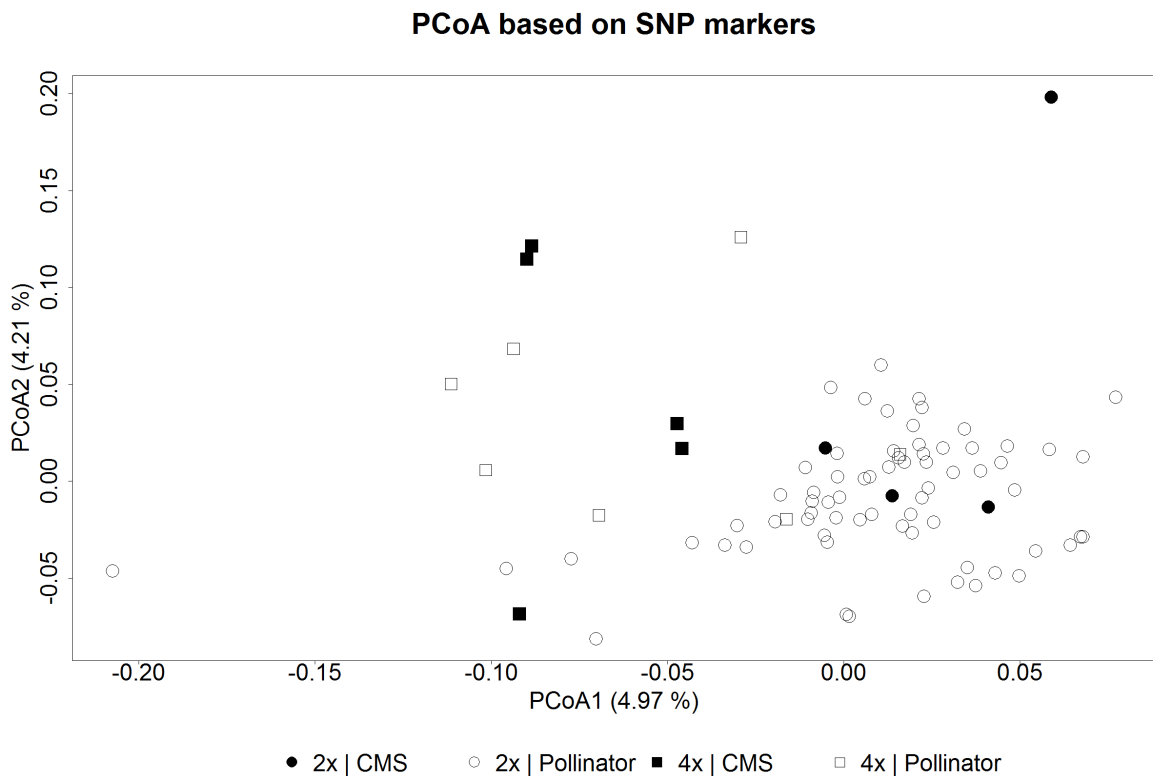
Figure 3.2 PCoA for parental material used in the field experiment
(b) based on SNP markers

PCoA based on SSR markers
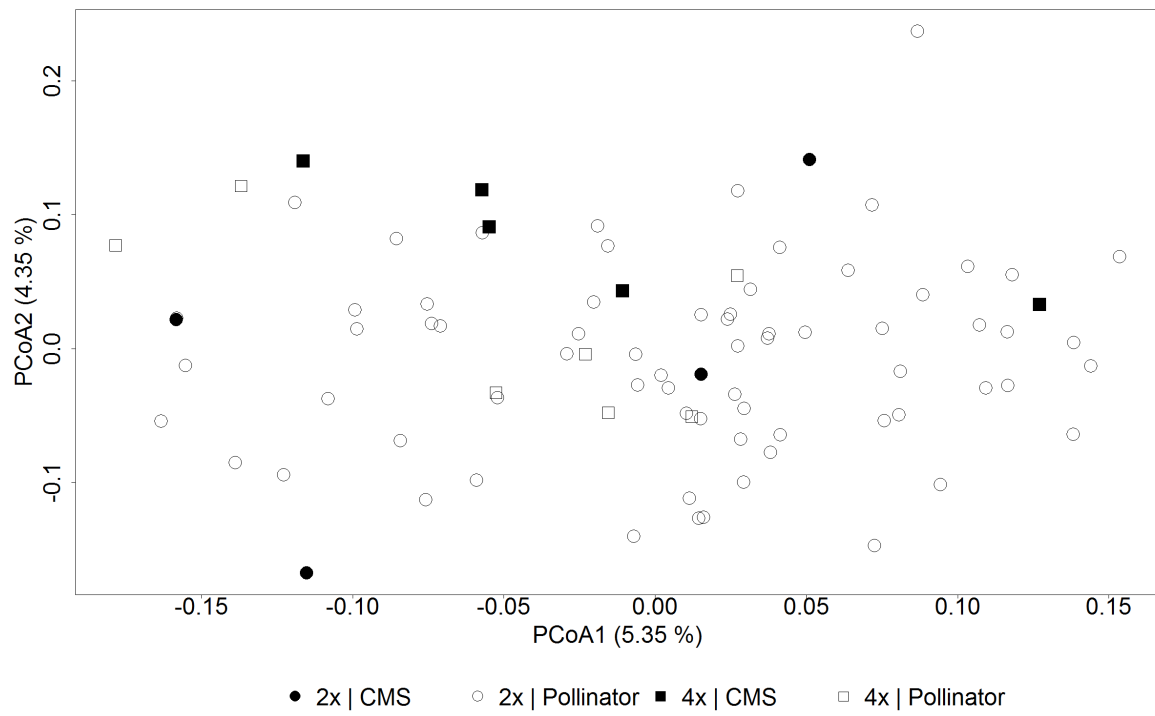
Figure 3.2 PCoA for parental material used in the field experiment
(c) based on SSR markers

Table 3.2 Range and mean $\pm$ SE for genetic distances between theoretical possible crosses and the actual crosses made in this study

			JD for DArT	MRD for SNP	MRD for SSR
Theoretical possible crosses [a]	diploids	range	0.25 - 0.57	0.18 - 0.48	0.37 - 0.67
		mean	0.41 $\pm$ 0.004	0.31 $\pm$ 0.003	0.52 $\pm$ 0.003
	tetraploids	range	0.37 - 0.47	0.28 - 0.42	0.39 - 0.60
		mean	0.43 $\pm$ 0.005	0.34 $\pm$ 0.006	0.52 $\pm$ 0.008
Actual crosses in the study [b]	diploids	range	0.25 - 0.57	0.18 - 0.48	0.37 - 0.66
		mean	0.39 $\pm$ 0.009	0.28 $\pm$ 0.007	0.50 $\pm$ 0.007
	tetraploids	range	0.38 - 0.47	0.28 - 0.37	0.39 - 0.57
		mean	0.43 $\pm$ 0.007	0.34 $\pm$ 0.007	0.50 $\pm$ 0.01

[a] Theoretical possible crosses include all the pairs of crosses containing CMS accessions utilized in the study.
[b] Actual crosses only include ones tested in the field trials.

Table 3.3 Average yield under different nitrogen levels

		Complete Nitrogen	Reduced Nitrogen
FMY	2010 sowings 2x	144.3	107.7
	2010 sowings 4x	161.9	123.8
	2011 sowings 2x	133.2	99.5
DMY	2010 sowings 2x	31.1	25.0
	2010 sowings 4x	31.7	25.8
	2011 sowings 2x	27.7	22.9

FMY: fresh matter yield; DMY: dry matter yield.

Table 3.4 Range, Mean $\pm$ SE, variance component estimates, heritability and LSD at 5% for fresh matter yield and dry matter yield (t/ha) for diploids and tetraploids in 2010 and 2011 sowings

	FMY			DMY		
	2x (2010 sowings)	2x (2011 sowings)	4x (2010 sowings)	2x (2010 sowings)	2x (2011 sowings)	4x (2010 sowings)
n	61	100	39	61	100	39
Range	100.35~137.21	92.83~132.10	126.35~156.99	25.01 ~ 29.96	21.86~27.57	26.65~31.21
Mean	125.92±3.77	116.36±2.65	142.58±3.80	28.04±0.70	25.28±1.36	28.80±0.60
σ_L^2	3143.78**	105.59**	4168.05**	186.03**	17.68***	204.98*
σ_N^2	665.32**	568.25**	717.52**	18.69**	11.29**	17.08**
σ_G^2	29.32**	49.48**	33.57**	0.80**	1.22**	0.75**
σ_{NL}^2	84.46**	10.61**	55.56**	3.51**	1.54**	2.28**
σ_{LG}^2	14.58**	22.04*	26.30**	0.90**	0.70**	0.71*
σ_{NG}^2 [a]	0	4.24*	0	0	0.22**	0.05
h^2	0.67	0.88	0.70	0.62	0.84	0.68
LSD5[b]	10.51	7.37	10.61	1.95	1.36	1.67

[a] negative but not significant variance components were adjusted to 0.

[b] least significant difference at p = 0.05.

*,** denotes significance at p = 0.05 and 0.01, respectively.

σ_L^2= variance components of location; σ_N^2= variance components of nitrogen; σ_G^2= variance components of genotypes; σ_{NL}^2= variance components of nitrogen × location interaction;σ_{LG}^2= variance components of location × genotype interaction;σ_{NG}^2= variance components of nitrogen × genotype interaction; h^2 denotes the broad-sense heritatbility.

Table 3.5 Phenotypic correlation coefficients between traits for 2010 and 2011 sowings

		FMY1Y	FMY2Y	FMY	DMY1Y	DMY2Y
FMY2Y	2010 sowings 2x	0.64				
	2011 sowings 2x	0.75				
	2010 sowings 4x	0.67				
FMY	2010 sowings 2x	0.96	0.84			
	2011 sowings 2x	0.95	0.92			
	2010 sowings 4x	0.94	0.88			
DMY1Y	2010 sowings 2x	0.70	0.50	0.69		
	2011 sowings 2x	0.91	0.64	0.84		
	2010 sowings 4x	0.73	0.61	0.74		
DMY2Y	2010 sowings 2x	0.32	0.82	0.54	0.47	
	2011 sowings 2x	0.58	0.87	0.75	0.59	
	2010 sowings 4x	0.47	0.88	0.69	0.69	
DMY	2010 sowings 2x	0.62	0.74	0.73	0.90	0.81
	2011 sowings 2x	0.85	0.83	0.90	0.91	0.87
	2010 sowings 4x	0.64	0.82	0.78	0.91	0.93

All the correlation coefficients were significant at p = 0.01.
FMY: fresh matter yield; DMY: dry matter yield.
1Y: total yield in the first growth year; 2Y: total yield in the second growth year.

Table 3.6 General combining ability (GCA) for tetraploid parents in 2010 sowings

FMY	CMS_234	CMS_235	CMS_239	GCA
NPZ_FNR1_(2010)	152.5	139.7	142.6	-1.9
Maint._msF1_384	145.7	144.2	143.7	-2.3
Syn_06_4503	143.7	156.99	146.4	2.2
PC_09_4902	145.5	147.1	150.6	0.9
Syn_08_4606	139.1	147.3	142.7	-3.8
Syn_08_4607	154.2	153.9	147.0	4.9
GCA	-0.1	1.4	-1.3	146.8[a]

DMY	CMS_234	CMS_235	CMS_239	GCA
NPZ_FNR1_(2010)	30.0	29.3	29.5	0.2
Maint._msF1_384	28.3	29.2	29.3	-0.4
Syn_06_4503	28.4	31.2	29.2	0.2
PC_09_4902	29.0	29.5	29.7	0.0
Syn_08_4606	28.0	30.1	29.5	-0.2
Syn_08_4607	29.5	29.9	29.5	0.3
GCA	-0.6	0.50	0.1	29.4[a]

[a] mean performance across all the hybrids.

Figure 3.3 Barplot for the performance of F1 hybrids and their corresponding mid-parent performance
 (a) for 2010 sowings. *left*: diploids; *right*: tetraploids; *red bar*: F1 hybrid performance; *green bar*: mid-parent performance. *Horizontal bar* showed the performance for the best checks.
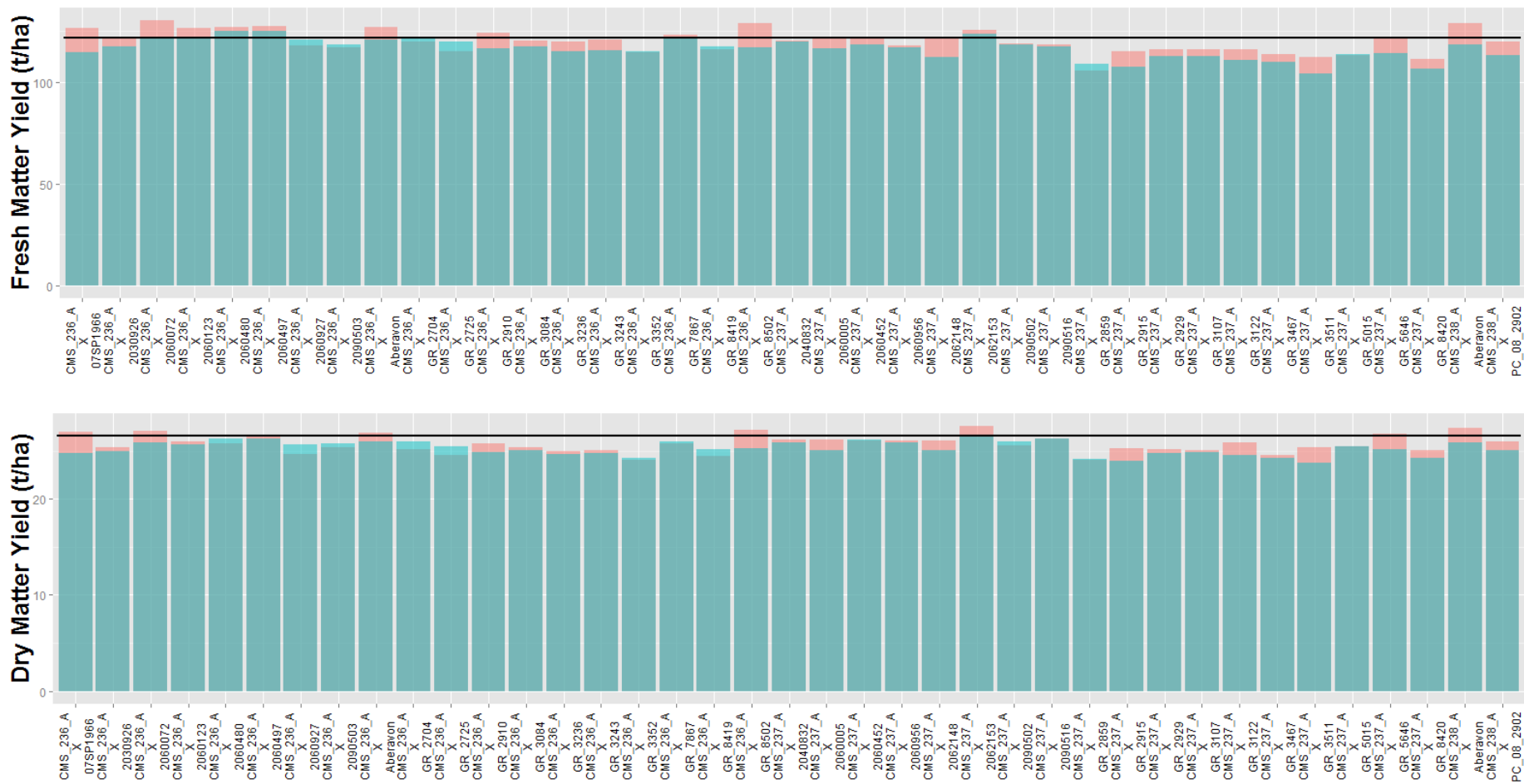
Figure 3.3 Barplot for the performance of F1 hybrids as well as their corresponding mid-parent performance
(b) for 2011 sowings. *red bar*: F1 hybrid performance; *green bar*: mid-parent performance. *Horizontal bar* showed the performance for the best checks.

Table 3.7 Correlation coefficient between hybrid performance and mid-parent heterosis, hybrid performance and mid-parent performance, heterosis and mid-parent performance for 2010 and 2011 sowings

| | | Correlation Coefficient | | |
		HP and MPH	HP and MPP	MPH and MPP
FMY	2010 sowings (8 diploids)	0.44	0.40	-0.63
	2011 sowings (39 diploids)	0.48**	0.69**	-0.31*
	2010 sowings (19 tetraploids)	0.74**	0.19	-0.52*
DMY	2010 sowings (8 diploids)	0.70	0.16	-0.60
	2011 sowings (39 diploids)	0.63**	0.54**	-0.31*
	2010 sowings (19 tetraploids)	0.79**	0.00	-0.61**

*,** significance at p = 0.05, 0.01, respectively.
FMY: fresh matter yield; DMY: dry matter yield; HP: hybrid performance; MPH: mid-parent heterosis; MPP: mid-parent performance.

Table 3.8 Summary of MPH and BPH for fresh matter yield and dry matter yield in diploids and tetraploid*s*

		MPH (%)					BPH (%)			
		Min	Max	Mean±SE	+/all[a]		Min	Max	Mean±SE	+/all[a]
FMY	2010.2x	9.41	19.58	13.88±0.53	8/8		-2.31	14.24	3.23±0.81	5/8
	2011.2x	-3.99	10.40	3.24±0.55	34/42		-8.82	7.84	-1.10±0.59	26/42
	2010.4x	-1.16	14.31	6.85±0.61	17/19		-4.2	13.91	4.94±0.65	17/19
DMY	2010.2x	4.63	11.63	8.57±0.43	8/8		-2.92	8.35	3.40±0.61	6/8
	2011.2x	-4.20	8.82	1.87±0.49	30/42		-7.28	5.59	-0.91±0.49	17/42
	2010.4x	-0.44	10.71	6.36±0.48	18/19		-3.66	10.16	4.42±0.55	17/19

[a] +/all indicates number of hybrids showed positive heterosis effect / total number of hybrids.

Table 3.9 Summary of correlation coefficients between genetic distance and mid-parent heterosis, genetic distance and hybrid per se performance for FMY and DMY in 2010 and 2011 sowings

		MPH			Hybrid per se performance		
		JD of DArTs	MRD of SNPs	MRD of SSRs	JD of DArTs	MRD of SNPs	MRD of SSRs
Diploids	2010 FMY (8 accessions)	-0.06	0.47	0.17	0.19	0.38	0.25
	2011 FMY (39 accessions)	0.34*	0.32*	0.35*	0.32*	0.32*	0.15
	2010 DMY (8 accessions)	0.24	0.60	0.19	0.00	0.29	0.07
	2011 DMY (39 accessions)	0.28	0.22	0.38*	0.39*	0.28	0.33*
Tetraploids	2010 FMY (19 accessions)	-0.24	0.23	0.09	0.04	0.16	-0.09
	2010 DMY (19 accessions)	-0.27	0.22	0.17	-0.2	0.00	-0.26

* denotes significance at p = 0.05; FMY: fresh matter yield; DMY: dry matter yield
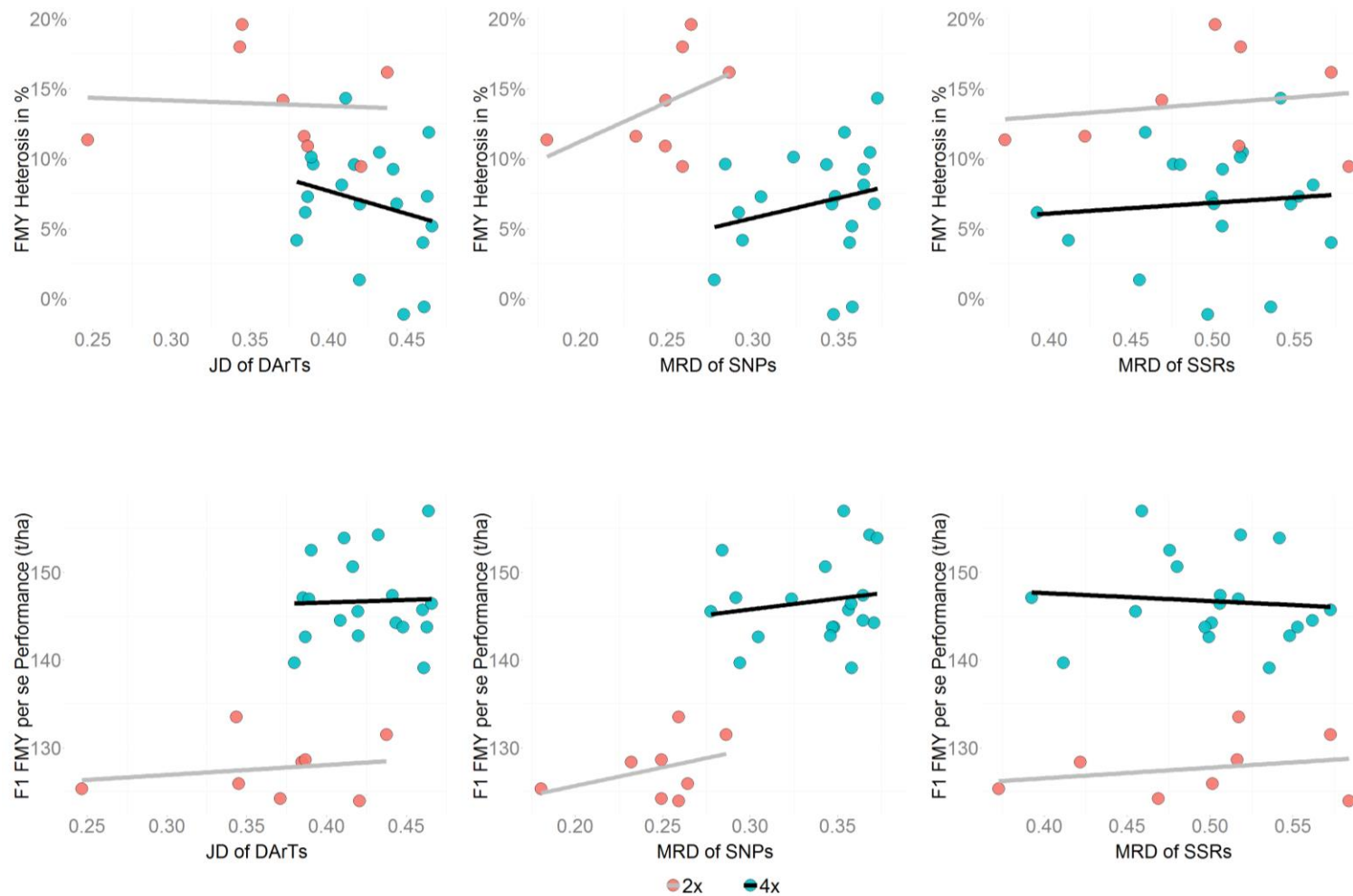
Figure 3.4 Graphical presentation of correlation between mid-parent heterosis, F1 per se performance and GD estimated by DArT, SNP and SSR markers (a) for FMY in 2010 sowings
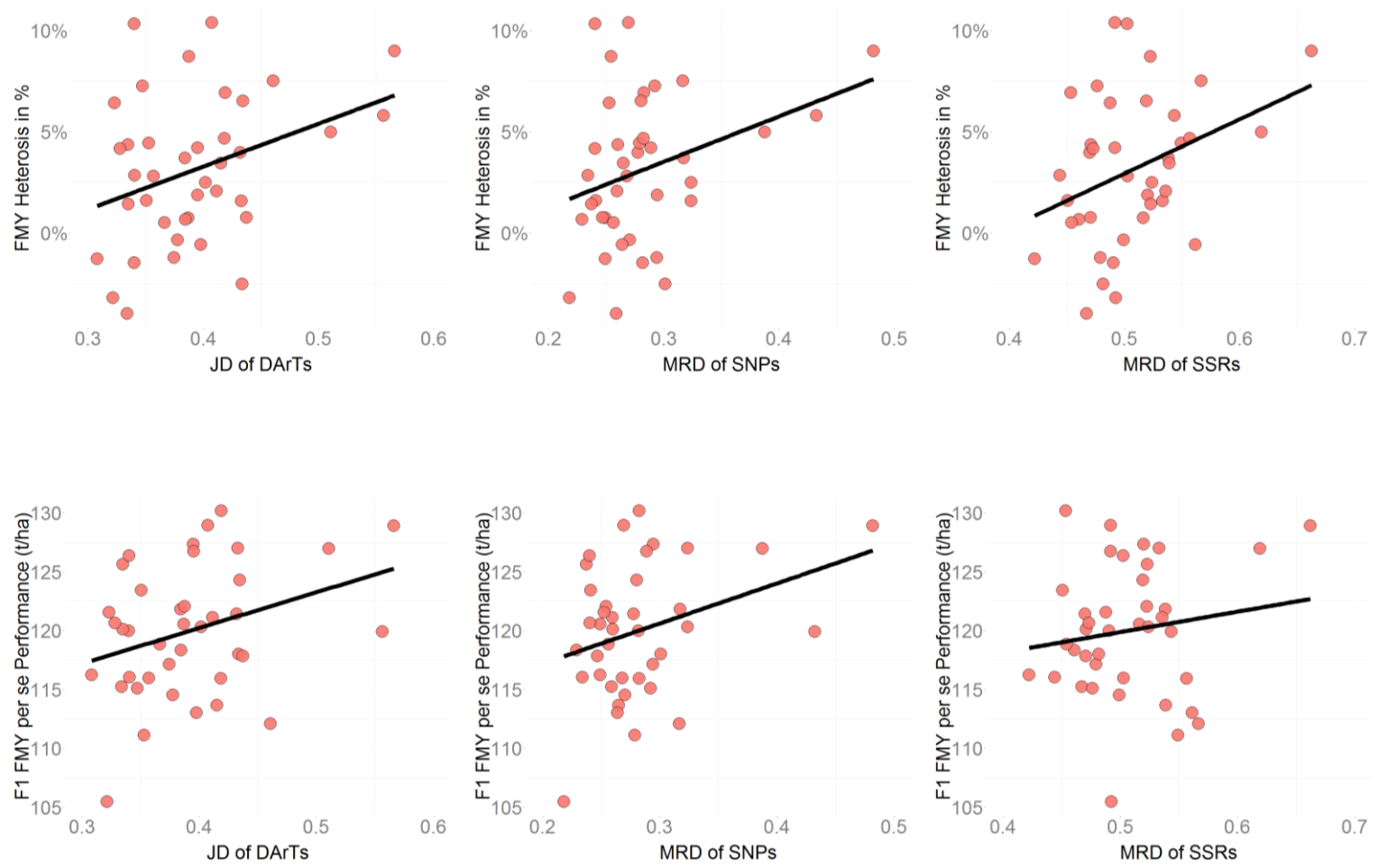
Figure 3.4 Graphical presentation of correlation between mid-parent heterosis, F1 per se performance and GD estimated by DArT, SNP and SSR markers (b) for FMY in 2011 sowings
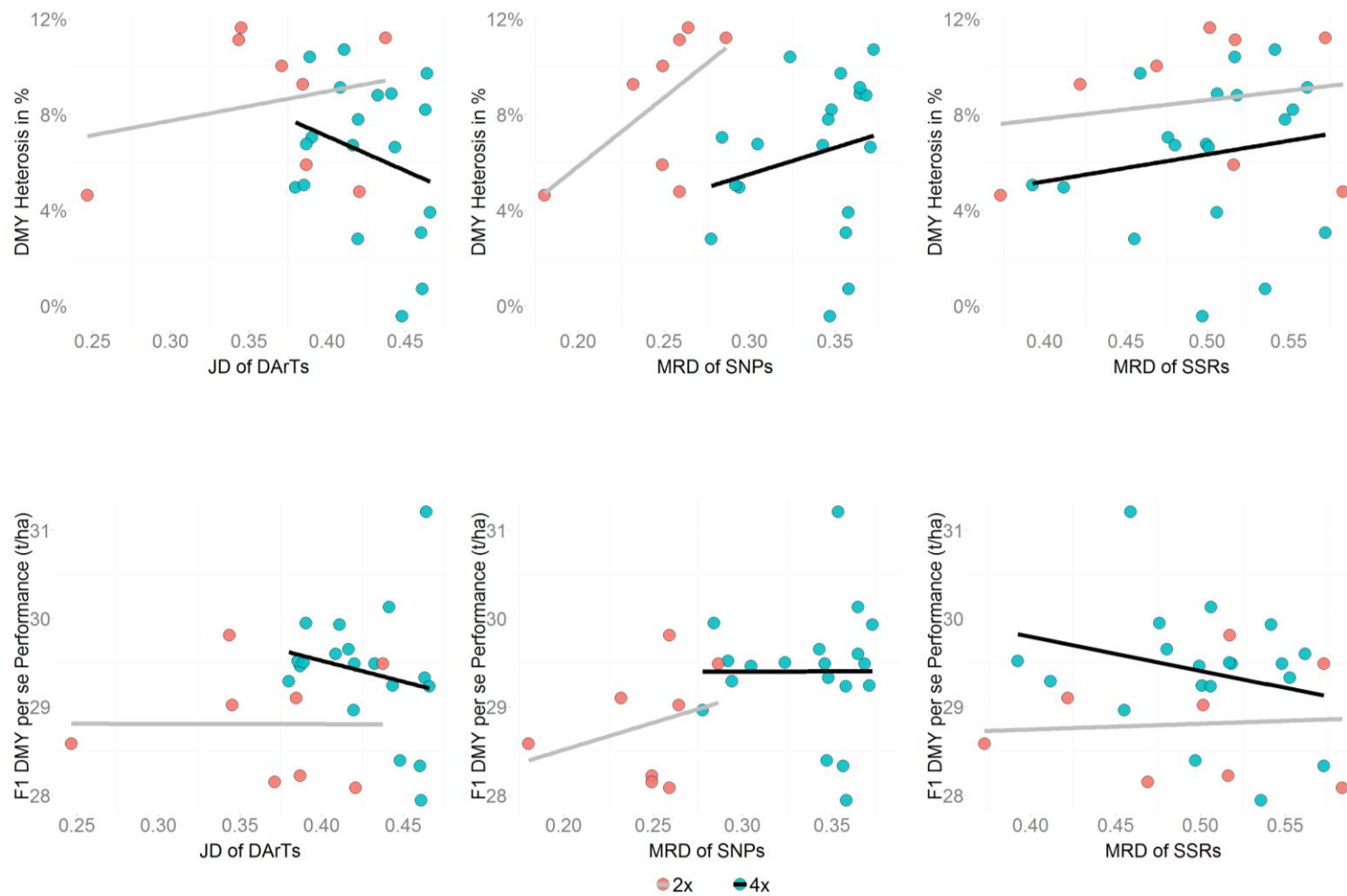
Figure 3.4 Graphical presentation of correlation between mid-parent heterosis, F1 per se performance and GD estimated by DArT, SNP and SSR markers (c) for DMY in 2010 sowings
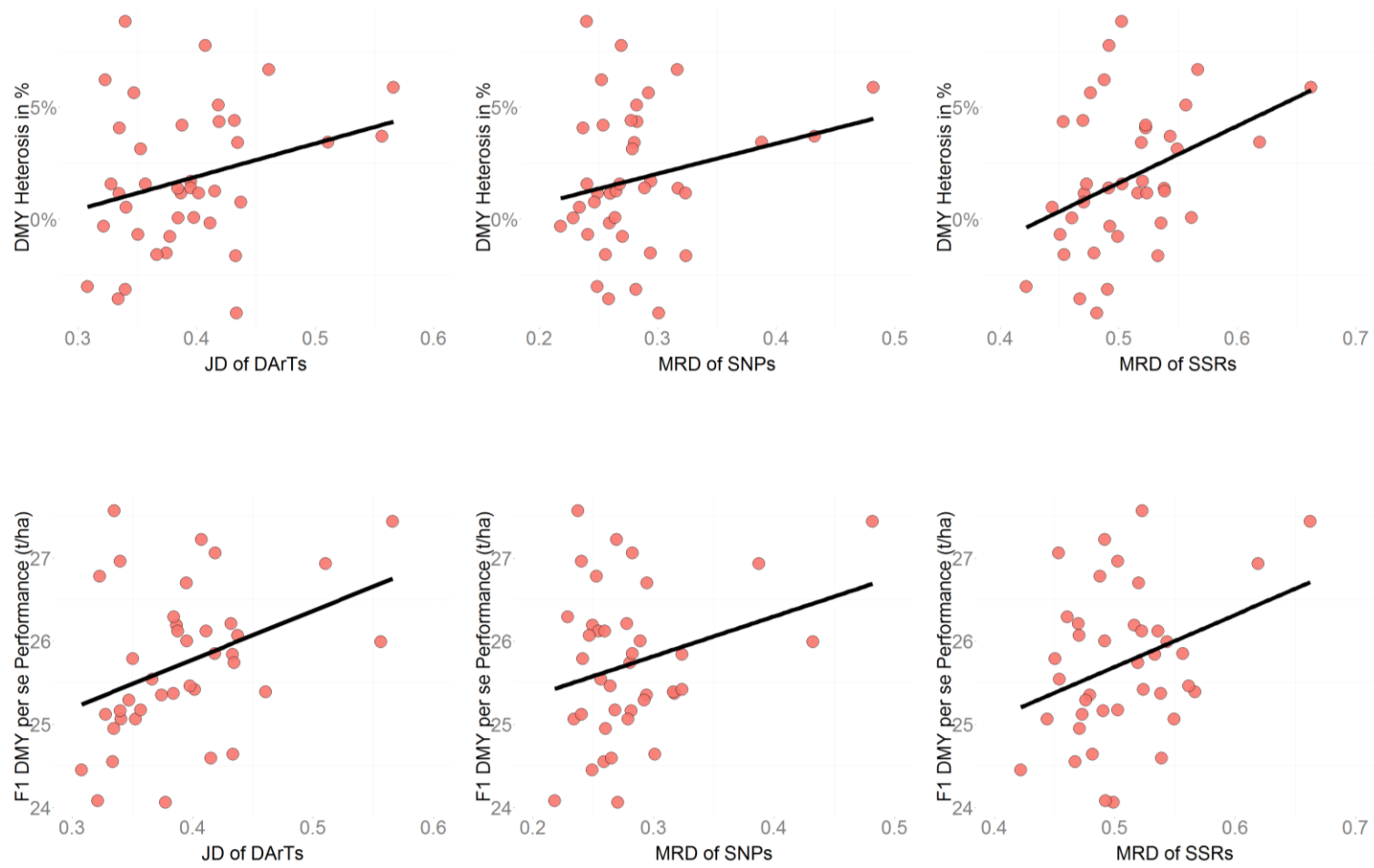
Figure 3.4 Graphical presentation of correlation between mid-parent heterosis, F1 per se performance and GD estimated by DArT, SNP and SSR markers (d) for DMY in 2011 sowings

3.4 Discussion

3.4.1 Genetic diversity revealed by DArT, SNP and SSR markers

The exploitation of bulked samples makes the extensive genetic exploration of highly heterozygous accessions possible at moderate costs. Despite of the difference in number of loci, attribution of marker types (co-dominant/dominant, multi-allelic/bi-allelic) and the methods to estimate the genetic distances, a certain extent of agreement between genetic structural patterns revealed by differed marker types was found. For instance, the distribution of the genetic distance was all bell-shaped (Fig. 3.1); all the marker types were highly polymorphic within the germplasm set; the mutual correlations between genetic distances were either moderately or highly significant at $p = 0.01$. In a similar study with AFLP, ISSR, RAPD and SSR markers in perennial ryegrass, significant correlations were only rarely observed (Posselt and Barre 2006). The improved agreement between marker types might be explained by higher number of marker loci used in the present study, which promises a better coverage of the genome. Additionally, the platforms used to generate marker data were either automate (SNPs and DArTs) or semi-automated (SSRs) which might reduce the errors during the scanning and/or scoring process.

In PCoA, large variation among germplasm accessions were observed for all the marker types, but structures clearly defining the subgroups of the material were difficult to identify in spite of the inclusion of both breeding material and ecotypes in the germplasm (Fig. 3.2). The lack of population structure in our germplasm set might be attributed to the extensive utilization of ecotype/landraces in ryegrass breeding since early 20th century (Roldán-Ruiz et al. 2001; Bolaric et al. 2005; Brazauskas et al. 2011). On the other hand, effort to maintain divergent improved populations were missing in most of the breeding programs (Brummer 1999), further diminishing potential structures within available breeding resources.

Interestingly, we observed a clear separation of tetraploid material from the diploid material in SNP markers, but for DArT and SSR markers such a differentiation was not found (Fig. 3.2(b)). If we transform the allelic frequencies in SNP into binary presence/ absence data by setting 50% allelic frequencies as the threshold, this separation disappeared (data not shown). To our knowledge, possibilities to distinguish ploidy levels using molecular markers

in perennial ryegrass were not reported before. From the summary of minor alleles in Chapter 2 (Tab. 2.3), we did not find many unique alleles in tetraploids and based on the breeding history, all of the existing tetraploid breeding material should be initially produced from diploids (Boller et al. 2010). This suggests that the differences might not be attributed to a special gene pool formed by tetraploids. Another possible reason is the alteration of heterozygosity caused by the complex interaction between homologous chromosomes in polyploids (Marsden et al. 1987). The sampling effect might also play a role: in a simple one locus two alleles case, diploid plants can only have three genotypes, but tetraploids five. Possibly the SNP genotyping platform used in this study was able to capture the slight differences in heterozygosity represented by allelic frequencies between diploid and tetraploid material, which was reflected as distinction in the PCoA plot.

3.4.2 Variance components, heritability and correlation between traits

Although heritability varied between sowing years, the range of broad–sense heritability obtained in the current study (0.67 to 0.88 for FMY and 0.62 to 0.84 for DMY, Tab. 3.4) was in agreement with related studies based on either conserved or frequent cutting management (Rhodes 1971; Devey et al. 1989; Conaghan et al. 2008) suggesting the success and usefulness of the harvest system used in the present study and rendering a reliable basis for further analysis. The heritability was higher for FMY than for DMY which is in agreement with inherent connections between these two traits: DMY is derived from the product of FMY and dry matter content estimated from small subsamples of each plot. Therefore extra errors should be introduced in addition to the error generated during FMY measurements (Conaghan et al. 2008). Superior heritability for FMY over DMY was also reported previously (Frandsen 1986).

Variance components attribute relative importance to different factors contributing to total variation. σ_N^2, σ_L^2, σ_G^2 and σ_{GL}^2 were significant at p = 0.05 in all cases and σ_N^2 and σ_L^2 provided larger variance components among them. For σ_{GN}^2, however, significance at p = 0.1 was only observed in 2011. Inconsistent levels of interactions between nitrogen levels and genotypes were also reported before by Wilkins (1989) in a study of four perennial ryegrass varieties for dry matter yield. The incorporation of several ecotypes as pollinators in 2011 (see Tab.3.1)

might be accountable for the enhanced nitrogen genotype interaction. In spite of the inconsistence of significance level, the absolute magnitude of σ_N^2 was much smaller than σ_G^2 suggesting that there is no need to analyze nitrogen levels separately in this experiment and the conclusions can be applied to both nitrogen levels.

In terms of the correlation between traits, we generally observed high correlation between yield in the first year and total yield, fresh matter yield with dry matter yield and relatively poor correlations between the yield in the 1st growth year and the 2nd growth year (Tab. 3.5). It is commonly observed that the total yield in the first growth year was always higher than in the second year (Wilkins 1989). In this experiment, the total yield is calculated as the sum of both years. Therefore yield in the first growth year became the most determinant factor and the variations in the second year have a lower influence on total yield. Considering multiple years' production, however, the correlation between first growth year and the total yield might be reduced. But in our experiment with only two years of growth, it seems that only considering the first growth year is adequate to provide nearly similar results.

According to O'Donovan and Dillon (1999), dry matter content should be between 16 and 18% if there is no surface moisture while cutting, which is not largely variable. The close correlation between fresh matter yield and dry matter yield in the present study (Tab. 3.5) might reflect this low variation in dry matter content. Tight correlation between FMY and DMY was also confirmed by Conaghan et al. (2008). They further proposed that breeders could use solely FMY as an indirect selection index to select genotypes for DMY (Conaghan et al. 2008).

3.4.3 Hybrid performance and heterosis

One of the characteristics of the present study is the strict pollination control rendered by the CMS system. The controlled pollination process avoids the possible intra population cross which tends to underestimate the heterosis. Apart from DMY in 2011 sowings, where only 17 out of 42 F1 hybrids were superior to their corresponding better performing parents, a high ratio of positive BPH was observed across traits and ploidy levels (Tab.3.8). Several hybrids

also outperformed the existing cultivars (Fig. 3.3). For diploids, BPH for FMY ranged from -8.82% to 14.24% with a mean of 3.23% and -1.10% in 2010 sowings and 2011 sowings, respectively. For DMY, the range covered -7.28% to 8.35% with a mean of 3.40% and -0.91% for both sowings. For tetraploids, a higher mean BPH for FMY (4.94%) and DMY (4.42%) was observed. Compared to the average genetic gain of around 0.4% per year in the past 90 years (Casler and Brummer 2008) the enhancement obtained by crossing population seems to be acceptable. However, considering the high cost in the process of multiplication of hybrid seeds (Brummer 1999), the results were not very appealing in a practical perspective.

The limited BPH could be also revealed by the mutual correlation between hybrid performance (HP), mid-parent heterosis (MPH) and mid-parent performance (MPP) giving the relationships that HP = MPH + MPP (Tab. 3.7). Even though HP showed positive relationships with both MPH and MPP indicating the necessity to select based on both MPH and MPP for yield improvement, the correlation between MPP and MPH was generally negative suggesting that selected better performing parents likely possess a smaller BPH therefore only limited gain in yield could be achieved finally.

Under spaced-plant conditions with semi-hybrids, Foster (1971a) reported a maximum BPH at 31%, but only 2 out of 15 hybrids displayed positive BPH; under sward condition, only one hybrid showed positive BPH within 15 hybrids with BPH of 17% (Foster 1971b). Recently, in a study of semi-hybrids constructed by eight parental populations, Barrett et al. (2010) reported an average of 1.6% increase in yield under sward condition and a maximum BPH of 7%. Comparatively, higher ratios of positive BPH and a similar level of average BPH were found in our experiment, but the maximum BPH was not as high as what was obtained by Foster (1971a, 1971b).

The hybrid performance as well as heterosis should be determined by four major factors for perennial ryegrass: 1) The GCA and SCA of the parents; 2) the design of the field experiments, either spaced-plants or high competitive sward conditions; 3) the crossing systems which including SI hybrid system, semi-hybrid system, CMS system etc.; 4) the existence of heterotic pools. In the present study, the general low level of heterosis (Fig. 3.3, Tab. 3.8) is on one hand caused by the highly heterozygous parental populations used in the cross, the expected BPH are thereby not as high as in hybrids from crossing two inbreds because the possible increase

in heterozygosity in our case is reduced. On the other hand, the lack of heterosis also reflects the lack of heterotic patterns within the germplasm which might be explained by the intensive usage of ecotypes and lack of maintenance of heterotic populations in breeding practice (Brummer 1999). It also corresponds to the lack of population structure revealed by molecular markers (Fig. 3.2). In addition, for spaced-plant experiment, the variance between genotypes could be fully presented. In contrary, under competitive sward conditions the differences between populations were reduced because only the more vigorous individuals could survive (Casler et al. 1996; Boller et al. 2010).

Compared to other studies, the often observed positive PBH (Tab. 3.8) could be explained by the exploitation of a CMS system against semi-hybrid systems in other studies. It excluded the possible intra-population pollination and therefore shows the full heterosis for population crosses.

The lower maximum BPH might be also attributed to the limitation of CMS resources in this experiment due to the large seed requirement for field trials under different nitrogen levels and at multiple locations. In both sowings, only six CMS accessions were used for the hybrid construction although a relative large amount of F1 hybrids were constructed. If the GCA of those CMS accessions was poor then the possibilities of obtaining a high BPH would be largely reduced. Eventually, based on the GCA estimated from tetraploid parental material, neither pollinators nor CMS accessions provided promising value (Tab. 3.6). The GCA of the diploid parents was not attainable.

The small number of CMS accessions has further impacts on the biases of the current study, because the calculation of MPH has to take both parents into account. Firstly, it might influence the correlation between HP, MPP and MPH (Tab. 3.7). If the GCA of these CMS parents changed, the correlation might be changed as well. Secondly, it might cause inconsistent results. For instance, in Tab. 3.8 we observed a large reduction from MPH to BPH for diploid hybrids in the 2010 sowings (for FMY from 13.88% to 3.23%, for DMY from 8.57% to 3.40%) compared with tetraploids (for FMY from 6.85% to 4.94%, for DMY from 6.36% to 4.42%). For the 2010 sowings, only one CMS accession (CMS_237) was used to construct the eight valid diploid hybrids and the MPH of these hybrids was largely influenced by this CMS accession.

3.4.4 The correlation between heterosis, F1 performance and genetic distances

Theoretically, the MPH should be linearly correlated with genetic distances (Melchinger 1999). For diploid hybrids in the current study, all but one of the correlation coefficients between genetic distance and MPH or F1 hybrid per se performance was positive, some of them were also significant at p = 0.05 level (Tab. 3.9). However, none of these correlation coefficients was high enough to make reliable predictions on heterosis based on genetic distances. In perennial ryegrass, similar studies to correlate MPH or F1 per se performance with genetic distance are rare. Kölliker et al. (2005) observed an increase in yield in poly crosses constructed by six parents with high genetic distance compared to those with low genetic distances, but the correlation coefficient was not mentioned. In other species, contradictory relationships between heterosis and genetic distances has been often reported (Joshi et al. 2001; Jaikishan et al. 2010; Wu et al. 2013). The lack of correlation might be attributed to the presence of epistasis, the lack of the association between the makers used in the GD estimation and QTL responsible for the specific traits and the poor correlation between GD and heterosis at QTL in the crosses examined (Reif et al. 2012).

The graphical presentation of these correlations revealed extra information (Fig. 3.4). Considering the correlations in the 2010 sowings, the distinction between tetraploids and diploids was crucial, because a clear separation of them was often observed. Under the separate consideration, correlations were difficult to identify for both ploidy levels. In the 2011 sowings, though the correlation coefficient was higher, this was mainly caused by three hybrids with high genetic distance. For hybrids between parents with narrower genetic distances, relatively high MPH could still be observed and there were rather poor association between GD and MPH within that range. To confirm these correlations, a more balanced distribution of the genetic distances of the parental material is required.

In tetraploids, SNP markers appeared to be superior to other marker types which might be accountable by their aforementioned ability to more accurately characterize tetraploids by capturing the differences in allelic frequencies. But even for SNP markers, the correlation coefficient was very small and non-significant. Therefore, it is difficult to draw a conclusion about which marker type is generally superior to others in terms of their correlation to heterosis.

3.5   Conclusion and further perspectives

In the present study, 85 parental accessions of perennial ryegrass were genotyped with bulk sampling strategy. This parental material and its deriving CMS F1 hybrids were field tested. In general, heterosis was limited in the tested hybrids. The correlation between genetic distance and mid-parent heterosis was positive but too low for any practical prediction. These positive correlations were mainly caused by three hybrids with both high MPH and high parental GD. Such correlations should be confirmed by further studies with higher numbers of hybrids deriving from highly diverse parents. At the same time, the development of more CMS accessions might be helpful in detection of higher heterosis as well as of heterotic patterns.

Chapter 4

Applications of Molecular Markers in *Lolium perenne* L.:

Association Mapping and Genomic Selection

4.1   Introduction

Perennial ryegrass is an obligate outbreeding diploid species and maintains high heterozygosity and heterogeneity. As one of the most important forage grasses it is mainly sown in Europe, temperate region of Asia, Australia, New Zealand, South Africa and South America (Boller et al. 2010). The broad usage and cultivation of perennial ryegrass could be attributed to its high level of quality, palatability, digestible energy, protein and minerals (Hannaway et al. 1999). Additionally, it can be used for soil conservation, nutrient recycling, turf, amenity grassland and as bioenergy plant (Hannaway et al. 1999). The major achievements in breeding of *Lolium perenne* include the improvement of yield, nutritional values and resistance against biotic and abiotic stresses. Conventional breeding based on recurrent selection or synthetics played a major role in this improvement without knowing the genetic control behind.

The fast development of molecular techniques permits a better understanding and manipulating of the genetic components underlining the traits, especially for complex traits controlled by many genes (namely quantitative trait loci, QTL) and therefore provided an alternative or supplement to the conventional breeding scheme (Moose and Mumm 2008). The identified QTL could be used for gene functionality research or be incorporated into a breeding program to conduct the marker assisted selection (MAS) (Moose and Mumm 2008). Through a two-way pseudo-testcross population, the first comprehensive linkage map in perennial ryegrass was constructed with SSR, AFLP, RFLP markers (Jones et al. 2002). Consequently, QTL for plant architecture, herbage yield, quality characters, cold tolerance, heading date variation and seed production were discovered in this population via linkage mapping approach (Yamada et al. 2004; Cogan et al. 2005; Armstead et al. 2008). In a QTL meta-analysis conducted in 2012, Shinozuka et al. summarized 560 QTL found in *Lolium perenne* on various morphological, physiological and resistance traits.

As an alternative methodology to linkage mapping, association mapping could be also conducted in analyzing complex traits. Different from linkage mapping, in which linkage disequilibrium (LD) was created from limited recombination events during the mapping population construction (Al-Maskri et al. 2012), association mapping exploits LD from historical recombinations. Therefore the markers identified in association mapping are

expected to be closer to the causal QTL, resulting in an enhanced resolution (Korte and Farlow 2013). On the other hand, linkage mapping is usually conducted within rather limited, in most cases highly selected material. Therefore the results in many cases cannot be applied to other natural or breeding populations (Korte and Farlow 2013). Association mapping is able to harness a broader range of material therefore the results could be used in a larger set of populations. The cons of association mapping lie in the higher amount of markers required and incapability to find effects caused by minor alleles.

Despite of several merits compared to linkage mapping, results from association studies generally suggested that many important quantitative traits like yield, quality and persistency are commonly explained by many QTL with only little variance explained by each of them (Heffner et al. 2009; Huang et al. 2009). The significant markers identified in association studies usually only explained a small proportion of the total variance, which constrains their implementation in MAS. In this context, another methodology named genomic selection was proposed and first applied in animal breeding (Hayes et al. 2001). The general idea of GS is that all the markers should be considered simultaneously instead of selecting the most significant markers as in association studies. The effect of all markers used in GS will be estimated from genotypes and phenotypes of a reference population and a prediction model could be created. Based on this model, genomic estimated breeding value (GEBV) could be inferred solely by the genotyping (Hayes et al. 2013). In simulation studies, the prediction accuracy of GEBV to the actual performance for grain yield could be as high as 0.58 in maize (Zhao et al. 2012) and 0.70 in rye (Wang et al. 2014). Currently, the application of GS in plants is already in use in some breeding programs.

The application of association mapping as well as genomic selection has been proposed in perennial ryegrass (Hayes et al. 2013). However, till now only few studies touched these topics. Within the frame of correlating the genetic distance and heterosis in the field experiment, we had phenotypic data of many parental accessions as well as their genotypic data estimated from bulked samples at hand as 'byproducts'. Although the design of the experiment was not optimized for association mapping or genomic selection, a preliminary application of these two methodologies was conducted.

4.2   Material and Methods

4.2.1   Plant material

The plant material used in the field trial is described in Chapter 3. We only used the parental material from the 2011 sowings because in the 2010 sowings, the incorporation of many tetraploids might complicate the analysis and cause unpredicted effects. The 2011 sowings consist 45 diploid pollinators and three diploid CMS accessions; among them, two pollinators were not genotyped. Therefore 46 parents were finally used in the association mapping and for genomic selection.

4.2.2   Field trial

Sowing was conducted in late 2011 and the measurements last for 2 growth years till end of 2013. The experiment was conducted at five locations including Hof Steimke (HS), Malchow (MA), Steinach (ST), Ven Ziderheide (VZ) and Moosburg (MO). At each location, two nitrogen levels, complete and reduced (60%), were applied. Plots were arranged with randomized complete block design and all accessions were tested under competitive sward condition.

4.2.3   Traits

Each trait was measured at both nitrogen levels in at least one location. The following traits were recorded:

1-7. Traits measured by Near Infrared reflectance spectroscopy (NIRS): Acid detergent fiber content (ADF in %), Acid detergent lignin content (ADL in %), Ash content (in %), Neutral detergent fiber content (NDF in %), Protein content (in %), in vivo Organic Matter Content (OMD in %), Digestibility of NDF content (DNDF in %). All of these traits were measured in 2012 during the first cut at Hof Steimke.

8. Absence of heads (AH): visually scored from many stems (scored as 1) to no stems (scored

as 9) with absent heads. Trait was measured in 2012 before the 3rd cut at Hof Steimke.

9. Early spring growth (EG): visually scored from bad (1) to good (9) in 2013 before the 1st cut at Malchow.

10. Sward density (SD): visually scored from low (1) to high (9) in 2013 before the 1st cut at Malchow.

11. Resistance to crown rust (RC): scored from susceptible (1) to resistant (9). RC was measured in 2012 before the 1st cut at Steinach.

12. Winter hardiness (WH): scored from bad (1) to good (9). WH was recorded in 2013 before the 1st cut at all five locations.

13. Fresh matter yield (FMY): harvested with plot harvester in 2012 and 2013 with multiple cuts in each year. The cutting frequency was determined at each location varying from 3 to 5. Total yield in t/ha over two harvesting year was recorded. FMY was measured at five locations.

14. Dry matter yield (DMY): calculated from the products of FMY and dry matter content estimated by NIRS equipped on the plot harvester. Total yield in t/ha over two harvesting year was recorded. DMY was measured at five locations, identical to FMY.

4.2.4 Molecular markers

The parental accessions were bulk sampled with 30 individuals per bulk.1384 DArT, 182 SNP and 48 SSR markers were applied. Further details about marker distribution and marker generation were described in Chapter 2. Based on the standard of Minor Allele Frequencies (MAF) > 0.1 and missing value rate < 0.3 within 46 parents in the 2011 sowings, 800 DArT, 134 SNP and 45 SSR marker loci (121 alleles) were screened.

4.2.5 Statistical analysis

Variance components, broad-sense heritability h^2 or repeatability and the entry means were estimated by using PLABSTAT (Utz 2011) , version 3A. For traits measured with more than one location, the following model was implemented:

$$y_{ijk} = \mu + g_i + l_j + n_k + gl_{ij} + gn_{ik} + ln_{kj} + e_{ijk}$$

where y_{ijk} is the phenotypic value of the genotype i at location j with the nitrogen level k; μ denotes the general mean; g_i represents the effect of genotype i; l_j describes the effect of location j; n_k indicates the effect of nitrogen level k; gl_{ij}, gn_{ik}, ln_{kj} depict the genotype $\times$ location, genotype $\times$ nitrogen, location $\times$ nitrogen interaction; e_{ijk} characterizes the residual error. According to Hill et al. (1998), broad sense heritability (h^2) was calculated as follows:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gl}^2}{N_l} + \frac{\sigma_{gn}^2}{N_n} + \frac{\sigma_e^2}{N_l \times N_n}}$$

where N_l, N_n denote the number of locations and number of nitrogen levels.

For traits measured at only one location, the following formula was applied:

$$y_{ik} = \mu + g_i + n_k + gn_{ik}$$

Repeatability was calculated as:

$$\text{repeatability} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_{gn}^2}$$

The correlations among traits were estimated by Spearman`s rank correlation to reduce the influence of possible outliers.

Association mapping was performed with R package 'rrBLUP' (Endelman 2011), version 4.3. To control the possible influence caused by population structure, a linear mixed model with

kinship matrix, denominated as A matrix, was implemented. The method used to calculate the A matrix are provided by Poland et al. (2012). Quantile–quantile (QQ) plots of the expected against observed P values were produced to check the effect of population control.

To control the false positive rate in multiple comparisons, false-discovery rate (FDR) was set to 0.2. Manhattan plot of the P value in the negative log_{10} scale against marker location in the genetic maps was used to visualize the results.

Genomic selection was performed by cross validation within 46 genotypes. For each simulation round, one subset of genotypes was used to train the prediction model (TS, training set); the rest of the genotypes were used to determine correlation between Genomic Estimate Breeding Value (GEBV) calculated from the prediction model and the actual phenotypes (validation set). Two scenarios were carried out: the first scenario is aimed to inspect the influence of the size of the training set on the prediction accuracy and determine the optimized training set. In comparison to other traits, FMY and DMY were well replicated therefore the 1st scenario was carried out on them. Accordingly, the optimized training set size applied to the other traits in Scenario 2. Ridge regression was used to estimate the parameters in the prediction model. The mixed model is shown as follows:

$$y = X\beta + Zu + \epsilon$$

where y is the vector of fitted genotypic value; β is the vector of fixed effect, and u is the vector of random marker effects. X is the design matrix for the fixed effect and Z is the marker matrix. ϵ denotes the residual error.

For each size of the training set, 500 cross validations were conducted.

4.3   Results

4.3.1 Phenotypic analysis

A total of 11 phenotypic traits for the 46 parental accessions were acquired from a single location with two nitrogen levels. In these traits, variance components for nitrogen and genotype are all significant at p = 0.01, except for absence of heads and ADL content. Due to the lack of repetition, nitrogen × genotype interaction could not be tested for significance (Tab. 4.1). The heritability or repeatability was between 0.60 and 0.80 for these traits.

For the traits that have been tested at multiple locations, the estimation of interaction is possible (Tab. 4.2). Nitrogen levels showed little effect on the winter hardiness since σ_N^2, σ_{GN}^2 , σ_{NL}^2 were either none significant or small in comparison to σ_G^2. Also σ_{GN}^2 was limited in fresh matter yield and dry matter yield compared to the corresponding genotypic variance. The rest of the factors were all significant at p = 0.01. Heritability of WH, FMY and DMY was 0.78, 0.88 and 0.84 respectively.

Additionally, the range, average, least significant difference at p= 0.05 for all the 14 traits were summarized in Tab. 4.3.

The correlations between traits displayed clear pattern of significance: most of the significant correlations were detected within NIRS traits and within non-NIRS traits (Tab. 4.4). Correlations between these two groups of traits were rarely observed. In NIRS traits, subgroups based on mutual correlation could be identified as well: subgroup containing ADL, NDF, ADL and subgroup containing protein, OMD, DNDF and ash content. High negative correlations were commonly observed between these two subgroups. In non NIRS traits, highly significant correlations were found between EG and WH, FMY, DMY; EG and SD with FMY, DMY; between FMY and DMY.

4.3.2 Association mapping

A total of 9 markers/alleles were found to be significantly correlated with 6 traits. For FMY,

one DArT and one SNP marker were identified. The same SNP marker was also detected in DMY. Two SNP markers were identified for AH and EG each. In each of ADL and ADF, one SSR allele was detected. The detected markers/alleles and their corresponding linkage group, map position, effects were summarized in Tab. 4.5.

Additionally, Manhattan plots and the QQ plot for these markers/alleles were shown in Fig. 4.1 and Fig. 4.2, respectively. Manhattan plots revealed the negative logarithm of p value and the location of the identified markers in the linkages groups. QQ plot was used to compare the inflation of the p values for none significant markers to assess the possible false positive rate by plotting the observed negative log p values against the estimated ones from normal distribution. As results, the applied model efficiently controlled the false positive rate because majority of the non-significant markers converged on the y = x line.

4.3.3 Genomic selection

In Scenario 1, the prediction accuracy increased with the increase of the training set size. Though the upward trend tends to reach a plateau after n = 24, the highest r is always acquired at n = 30 (Tab. 4.6). The highest prediction accuracy for DArT, SNP and SSR markers is 0.40, 0.32, 0.48 for FMY and 0.46, 0.33, 0.53 for DMY, respectively. In the bar plot of scenario 1 for both traits (Fig. 4.3, Fig.4.4), the variation of each bar is gradually reduced from n = 3 to n = 15, but is increased from n = 18 to n = 30.

Due to the highest r obtained at n = 30 in scenario 1, the training set size of 30 was fixed for scenario 2. Of these 12 traits, ash, AH, EG, WH showed the highest r with DArTs, while ADF, NDF, protein, OMD, SD and RC were better predicted by SSRs. The performance of SNP markers exceeded the other marker types only for ADL and DNDF. However, the general prediction accuracy for traits in scenario 2 was low: none of these traits exceeded the highest prediction obtained in scenario 1, which was 0.48 for FMY and 0.53 for DMY.

Table 4.1 Variance components and heritability for all the traits estimated at a single location

	σ_N^2	σ_G^2	σ_{GN}^2	h^{2a}
ADF content	2.31**	3.23**	1.63	0.80
ADL content	0.01*	0.05**	0.04	0.68
Ash content	0.10**	0.30**	0.22	0.73
NDF content	9.26**	5.12**	5.35	0.66
Protein content	0.23**	1.04**	1.41	0.60
OMD-VIVO content	3.93**	3.18**	3.06	0.68
DNDF-content	1.52**	3.58**	2.53	0.74
Absense of heads	0.02	1.89**	1.01	0.79
Early spring growth	0.04**	0.47**	0.30	0.76
Sward density	0.16**	0.42**	0.33	0.72
Resistance to crownrust	0.12**	0.88**	0.73	0.71

[a] also repeatability in this case.
*, ** denote the significance at p = 0.05, 0.01 respectively.

Table 4.2 Variance components and heritability for all the traits estimated at multiple locations

	σ_L^2	σ_N^2	σ_G^2	σ_{GN}^2	σ_{LG}^2	σ_{NL}^2	h^2
Winter hardiness	1.13**	0.00[a]	0.14**	0.02*	0.37**	0.00[a]	0.78
Fresh matter yield	105.59**	568.25**	49.48**	4.24*	22.04**	10.61**	0.88
Dry matter yield	17.57**	11.37**	1.25**	0.21**	0.64**	1.60**	0.84

[a] negative variance component were adjusted to 0.
*, ** denote the significance at p = 0.05, 0.01, respectively.

Table 4.3 Descriptive statistic of the 46 parental perennial ryegrass accessions

Trait	Min	Max	Mean	LSD5
Estimated in one location				
ADF content (%)	27.67	38.35	31.04	2.53
ADL content (%)	1.98	3.19	2.41	0.41
Ash content (%)	8.95	12.06	10.82	0.93
NDF content (%)	48.26	64.35	54.64	4.59
Protein content (%)	5.47	15.25	8.87	2.36
OMD-VIVO content (%)	67.07	79.69	75.31	3.47
DNDF-content (%)	67.88	80.76	75.53	3.15
absense of heads	1	8	4.71	1.99
early spring growth	4.5	8	6.34	1.08
sward density	3	8	7.02	1.13
resistance to crownrust	1	6	4.17	1.7
Estimated at multiple locations				
winter hardiness	5.5	7.6	6.65	0.56
Fresh Matter Yield (t/ha)	92.83	132.1	113.12	7.37
Dry Matter Yield (t/ha)	21.86	27.24	24.83	1.37

*, ** denote the significance at p = 0.05, p = 0.01 respectively.
LSD5: least significant difference at the level of P = 0.05.

Table 4.4 Spearman's rank correlation of all the 14 traits (n = 46)

	ADF	ADL	Ash	NDF	Protein	OMD	DNDF	AH	EG	SD	RC	WH	FMY
ADL	0.68**												
Ash	-0.49**	-0.55**											
NDF	0.98**	0.66**	-0.39**										
Protein	-0.58**	-0.30*	0.59**	-0.49**									
OMD	-0.94**	-0.73**	0.42**	-0.94**	0.55**								
DNDF	-0.78**	-0.73**	0.52**	-0.75**	0.65**	0.91**							
AH	-0.39**	-0.37*	0.24	-0.36*	0.16	0.37*	0.31*						
EG	0.17	-0.10	0.01	0.10	-0.31*	-0.18	-0.20	-0.04					
SD	-0.08	-0.10	-0.10	-0.13	-0.09	0.09	-0.04	0.41**	0.21				
RC	-0.10	-0.27	0.24	-0.13	0.00	0.11	0.09	-0.04	0.40**	0.20			
WH	0.03	-0.16	0.08	-0.06	-0.12	-0.07	-0.16	-0.02	0.68**	0.32*	0.42**		
FMY	0.02	-0.33*	0.24	-0.02	-0.21	-0.06	-0.11	0.02	0.69**	0.38**	0.67**	0.74**	
DMY	0.00	-0.24	0.08	-0.06	-0.27	-0.05	-0.18	0.02	0.70**	0.45**	0.63**	0.72**	0.92**

*, ** denote significance at p = 0.05, 0.01 respectively.
ADF: Acid detergent fiber content; ADL: Acid detergent lignin content; Ash: Ash content; NDF: Neutral detergent fiber content; Protein: Protein content; OMD: *in vivo* organic matter content; DNDF: Digestibility of NDF content; AH: Absence of heads; EG: Early spring growth; SD: Standing ability; RC: Resistance to crown rust; WH: Winter hardiness; FMY: Fresh matter yield; DMY: Dry matter yield.
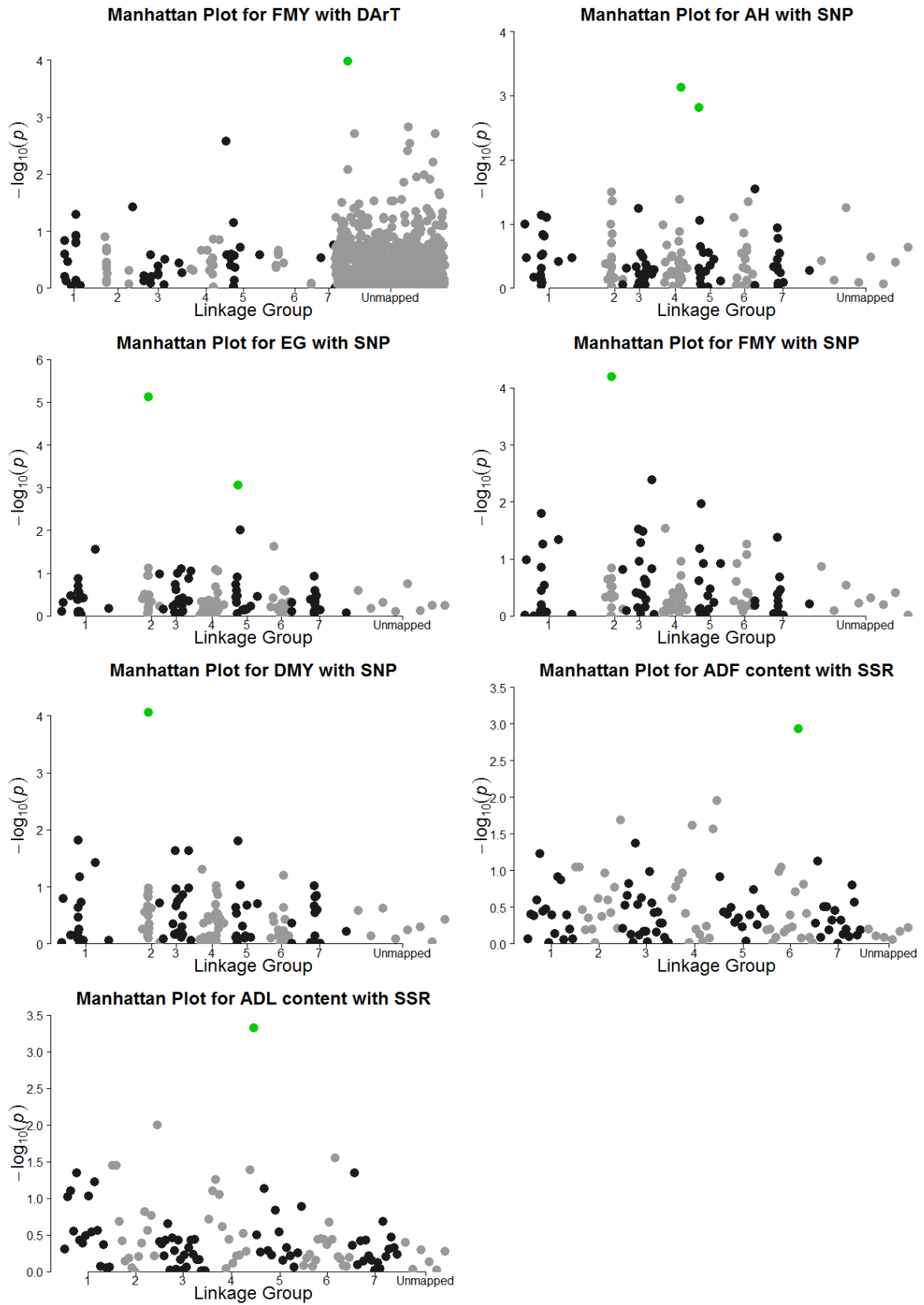
Figure 4.1 Manhattan Plot for the traits containing significant markers
The identified markers are highlighted with green color.
Markers with unknown position are grouped into 'unmapped'
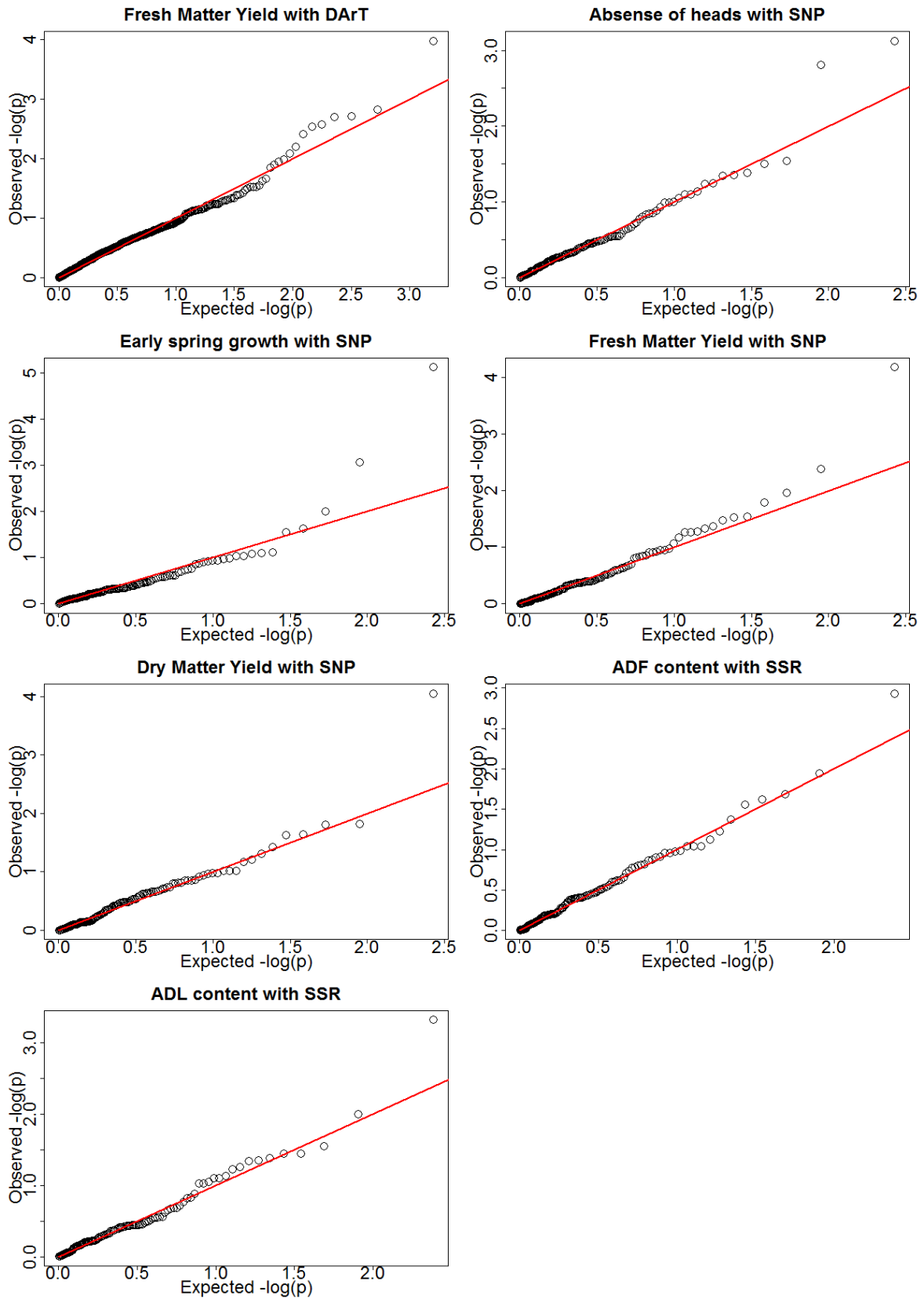
Figure 4.2 Quantile-quantile plots of traits containing significant markers
Red auxiliary line: y = x.

Table 4.5 Results of association mapping

Type	Marker Name	Trait	LG	Position	Effect	Variance %[a]
DArT	loPt.351327	FMY	-	-	-13.51	32.96%
SNP	loPt.351645	Absense of heads	4	54.331	4.90	31.11%
SNP	loPt.321347	Absense of heads	5	24.234	5.53	24.24%
SNP	loPt.211235	Early spring growth	2	81.555	2.17	41.60%
SNP	loPt.431327	Early spring growth	5	28.488	-1.71	25.38%
SNP	loPt.121639	FMY	2	81.555	18.03	35.13%
SNP	loPt.121639	DMY	2	81.555	3.16	37.39%
SSR	loPt.145664	ADF content	6	-	6.00	17.70%
SSR	loPt.132745	ADL content	4	-	0.78	27.07%

[a] variance explained by markers was estimated by simple regression on phenotypic value.

Table 4.6 Results of genomic selection for simulation scenario 1 and scenario 2

Size of TS[a]	3	6	9	12	15	18	21	24	27	30
Scenario 1: FMY										
DArTs	0.16	0.24	0.28	0.32	0.34	0.36	0.37	0.39	0.40	0.40
SNPs	0.09	0.11	0.17	0.19	0.21	0.25	0.27	0.29	0.32	0.32
SSRs	0.20	0.33	0.39	0.43	0.45	0.47	0.47	0.48	0.48	0.48
Scenario 1: DMY										
DArTs	0.15	0.23	0.29	0.34	0.37	0.40	0.41	0.44	0.45	0.46
SNPs	0.10	0.13	0.2	0.22	0.23	0.27	0.29	0.32	0.34	0.33
SSRs	0.19	0.31	0.39	0.43	0.46	0.49	0.5	0.52	0.51	0.53

Scenario 2 (size of TS fixed to 30)												
Traits	ADF	ADL	Ash	NDF	Protein	OMD	DNDF	AH	EG	SD	RC	WH
DArTs	0.19	0.34	0.30	0.12	0.02	0.12	0.18	0.32	0.17	0.01	0.35	0.37
SNPs	0.10	0.37	0.26	0.03	0.07	0.13	0.24	0.04	0.00	-0.02	0.38	0.04
SSRs	0.38	0.18	0.20	0.31	0.13	0.25	0.18	0.08	0.12	0.11	0.43	0.21

[a] TS: Training Set.
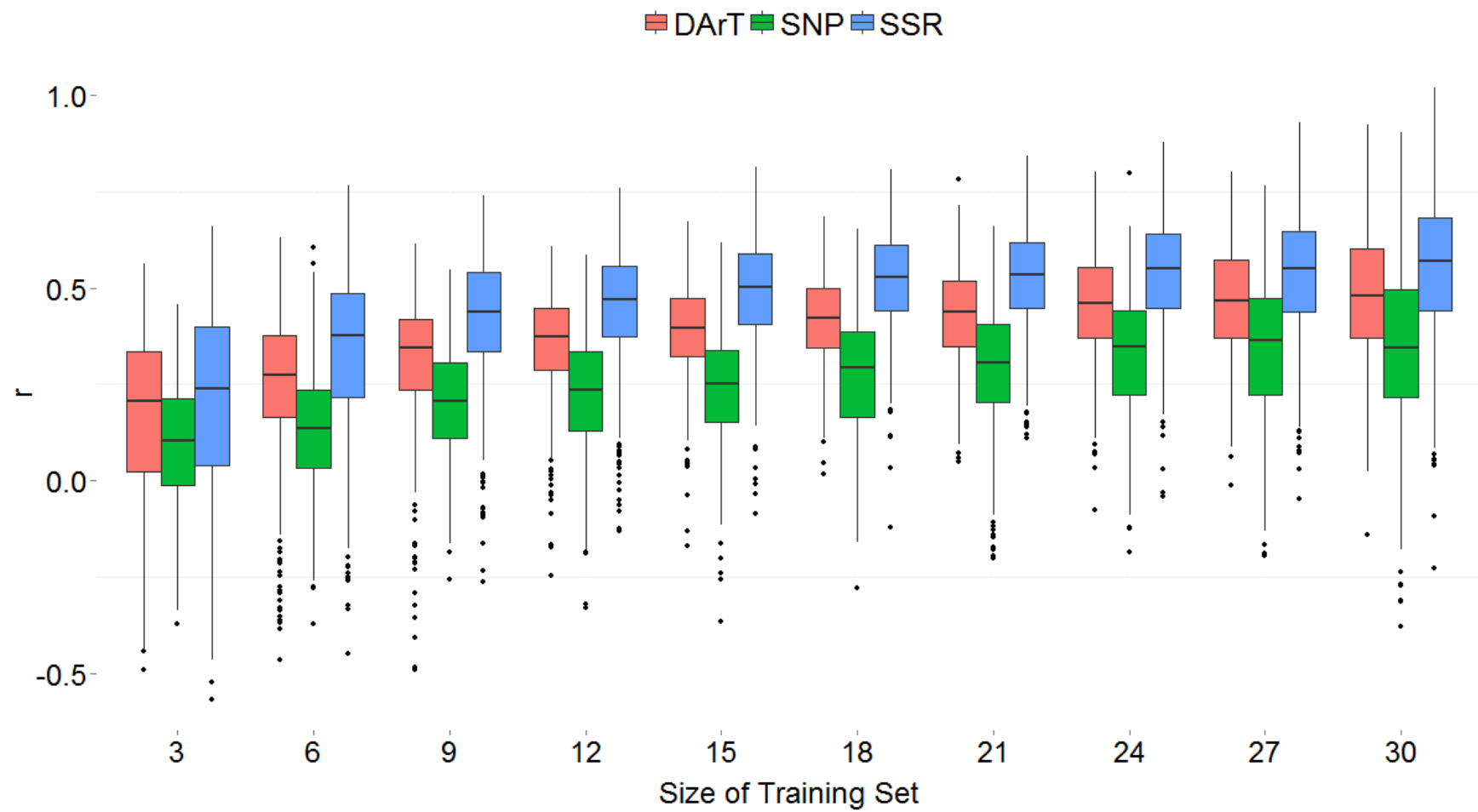
Figure 4.3 Boxplot for the cross validation results for fresh matter yield
The size of the training set ranged from 3 to 30 with an interval of 3

Figure 4.4 Boxplot for the cross validation results for dry matter yield
The size of the training set ranged from 3 to 30 with an interval of 3
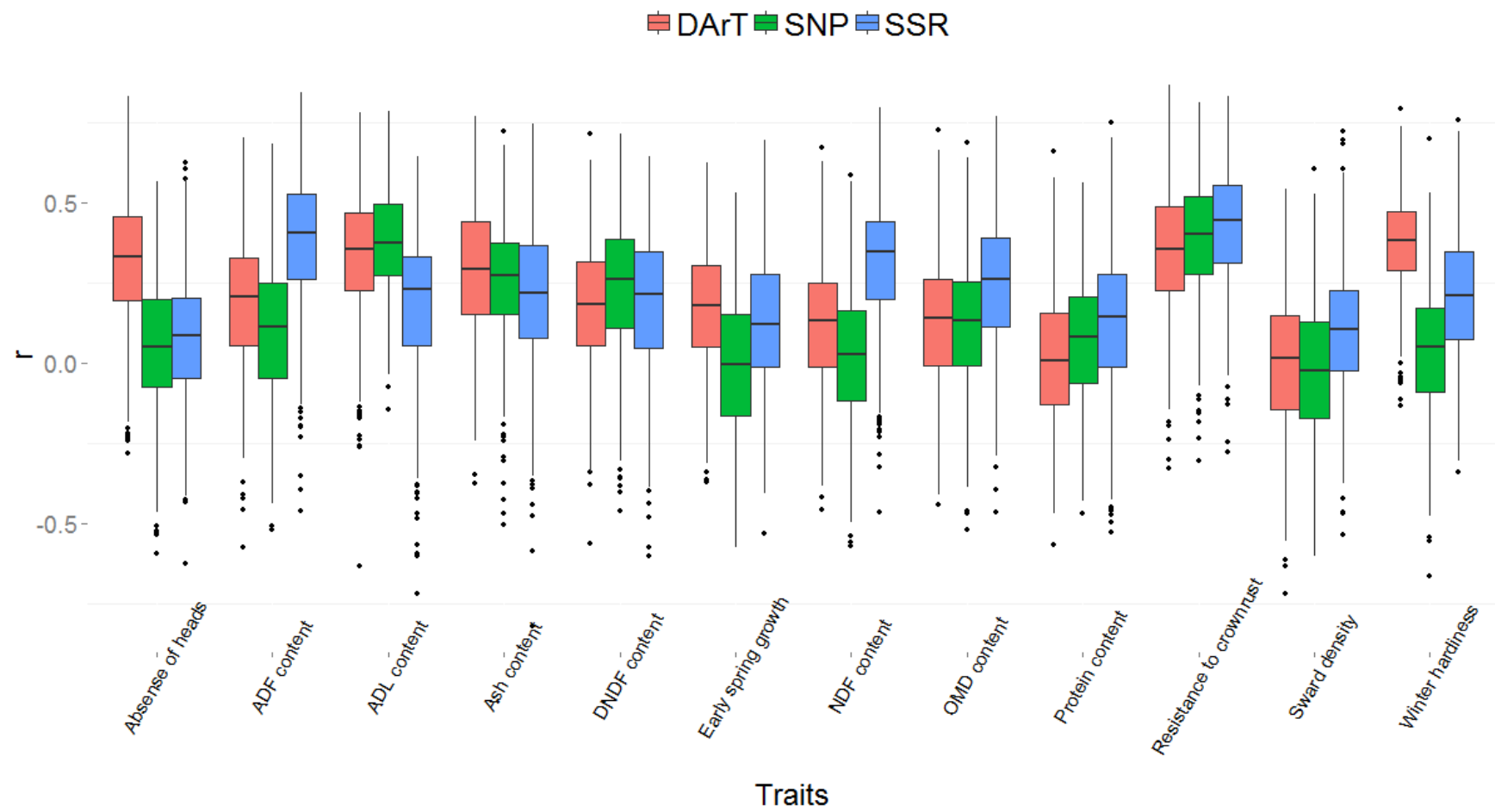
Figure 4.5 Boxplot for the cross validation results for traits in scenario 2
The size of the training set is fixed to 30.

4.4 Discussion

4.4.1 Phenotypic analysis

14 traits in total were recorded for each of the 46 accessions. The effect of nitrogen levels usually affected the phenotypic variance to a sign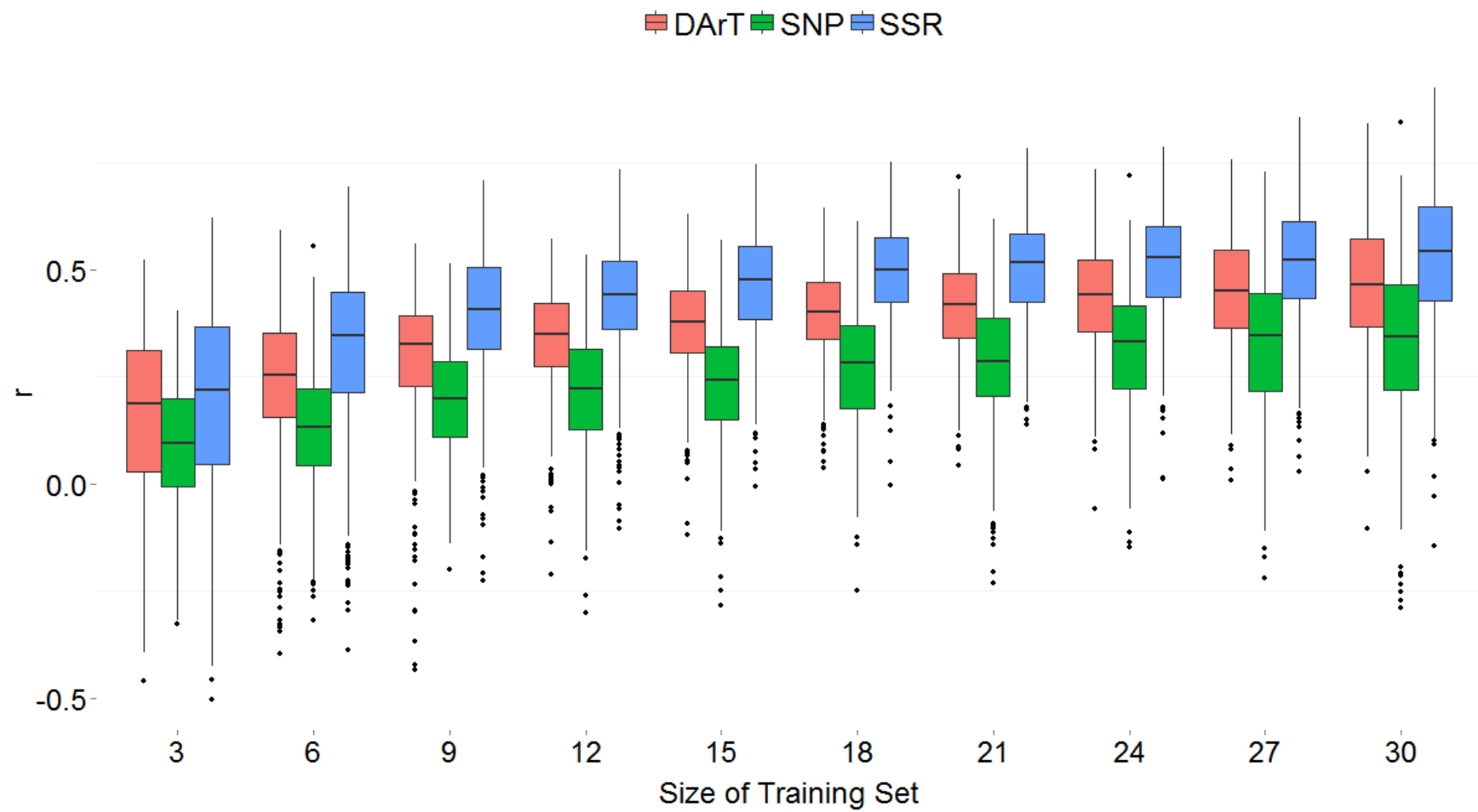ificant extent (Tab. 4.1, Tab. 4.2), which is in agreement with the study of Rasmussen et al. (2008) who reported significant influence of nitrogen supply on ADF, NDF, ash, protein, OMD, DMY. Due to the lack of replication, the significance of interaction cannot be confirmed for most of the traits. In FMY, DMY as well as WH, however, the nitrogen by genotype interaction was found to be smaller compared to the genotypic variance for almost one order of magnitude (Tab.4.2). Moreover, the broad sense heritability or repeatability was moderate (0.60) to high (0.88); large phenotypic variation was revealed for most of the traits (Tab. 4.3), which are favorable for the association mapping and genomic selection.

Spearman's rank correlations mainly uncovered two groups of traits: NIRS traits and non NIRS traits. Significant correlations were commonly observed within but not among groups (Tab. 4.4).

Within the NIRS group, based on correlations, two subgroups could be identified: one with ADL, NDF, ADL content and one containing protein, OMD, DNDF and ash content. The most prominent correlation coefficient were found between NDF and ADF (r = 0.98), OMD and DNDF (0.91), OMD and ADF (-0.94), NDF (-0.94). As measurements of cell wall components, high correlations between NDF, ADF and ADL were reported in grasses (Jancik et al. 2008) and maize (Cardinal et al. 2003). As important indexes to access the nutritional value, a close positive correlation between OMD and DNDF was also suggested (Nousiainen et al. 2004). The negative correlation between DNDF and NDF, ADF and ADL was revealed by Koukolová et al. (2004) and Jančík et al. (2011). The correlation estimates observed in the present experiment are consistent with the results from these studies.

Within the non NIRS group, both WH and RC showed significant positive correlations between FMY and DMY indicating the favorable influence of high WH and RC on biomass yield. Although EG was visually scored at early spring, it provided a good prediction of FMY and

DMY. The prominent correlation between FMY and DMY (r = 0.92) was also found by Conaghan et al. (2008) who reported correlation of 0.84 between these two traits.

4.4.2 Association mapping

Association mapping was conducted with 14 traits of 46 perennial ryegrass accessions and 800 DArT, 134 SNP and 45 SSR marker loci (121 alleles). A total of nine markers/alleles were identified for FMY, DMY, AH, EG, ADF, ADL (Tab. 4.5). Except for loPt.351327, the location or the LGs of the identified markers was known and could be used to compare them to related study on perennial ryegrass. Although the applications of genome wide association studies are rarely implemented in perennial ryegrass, a high number of markers linked to QTL for various traits via linkage mapping were reported (Shinozuka et al. 2012). Among them, Anhalt et al. (2009) uncovered two markers for FMY and three markers for DMY at LG2; two markers at LG6 were identified to be responsible for ADF content (Xiong et al. 2006). However, due to the differences in marker types and linkage maps, it is unknown whether the markers discovered in the present study on the same linkage groups were linked to the same QTL reported previously. It is worthwhile to stress that, apart from FMY, DMY, the rest of the traits where associated markers were found was only tested at one location implying the possible restriction of the validity of these markers in multiple environments.

In spite of the large number of traits investigated, the number of markers detected is rather limited. In a review on association studies, Al-Maskri et al. (2012) summarized the factors that determine the power of association studies. They include LD in the mapping material; type of gene action of the trait; size of the population and the design as well as the accuracy of the field trial. In the current study, the size of the population is very small (n = 46), the marker number is low considering the low LD commonly observed in perennial ryegrass (Skøt et al. 2005), and the phenotyping might be not accurate enough due to the lack of repetition for most of the traits. These factors might explain the low detection power of the association study.

A major obstacle in association studies are possible false positive associations of detected markers rendered by population structure and genotype relatedness. To handle this, mixed

linear models fitting genetic relatedness matrix were commonly applied (Yu et al. 2006). Endelman et al. (2012) proposed using A matrix to estimate the realized additive relationship matrix. Therefore we applied A matrix in the models and QQ plots were used to inspect the efficacy of population control. In QQ plot, efficient population control should be manifested by small deviations from the y = x line for the markers which is not linked to QTL. From this perspective, fitting the A matrix to the model constrained the false positive rate efficiently because the majority of the none-significant markers did not deviate from the auxiliary line (Fig. 4.2).

4.4.3 Genomic selection

We investigated two scenarios for genomic selection. In the first scenario, the highest prediction accuracy was always observed at the largest training set size but the upwards trend tended to reach a plateau and the highest average r value was always found in SSRs. However, a higher prediction accuracy of SSRs over SNPs and DArTs is generally unexpected. Although SSRs (allelic frequency) are more informative than DArTs (presence/absence) for the bulked sampled genotypes, the coverage of the genome is rather limited for traits influenced by many QTL, especially in species with low LD like perennial ryegrass (Skøt et al. 2005). On the other hand, SNPs covered more loci than SSRs and provided more accurate frequencies estimation (see Chapter 2.) but it did not outperform SSRs. Therefore it is suspected that the relatively high prediction of SSRs in Scenario 1 is caused by over estimation originated from small sample size with large variation between populations as reported before in maize (Zhao et al. 2012).

In scenario 2, a fixed training set size was utilized and performance of markers varied among these traits. But none of the r values outperformed that in scenario 1, and even for scenario 1 with suspected over estimation, the r value was not high enough to make accurate prediction. It is known that the number of markers and number of genotypes are of fundamental importance for the prediction (Hayes et al. 2013), but both factors were rather limited in the present study. Apart from limited sample size and marker numbers, the lack of repetition and lower heritability might be also accountable for the limitation of prediction

accuracy (Wang et al. 2014).

4.5   Conclusion and further perspectives

With the available marker data and phenotypic records, association mapping and genomic selection were applied. The results were not very promising but this should be mainly attributed to the fact that the design of the experiment is not optimized for association mapping and genomic selection. However, considering that genotyping and marker developments are increasingly accessible and affordable, the potential of these methodologies should not the ignored. It would be interesting to implement them in larger genotype and marker sets and to investigate whether these methods could be efficiently used for improving complex traits in breeding practice.

Chapter 5

General Discussion and Conclusion

The improvement of yield in *Lolium perenne* in the past years is not comparable to that of other crops despite of its economic importance. This might be in part attributed to the incomplete utilization of heterosis by the conventional breeding schemes. Hybrid breeding contributes to the impressive enhancement of yield in many crops, therefore this strategy was also proposed for perennial ryegrass (Vogeland and Pedersen 1993). Due to many technical problems like availability of CMS accessions, the hybrid performance of perennial ryegrass was only tested by semi-hybrids and the results were not very promising. On the other hand, molecular markers provide a powerful tool to facilitate the breeding work and their application has been already incorporated into many breeding programs. Therefore it would be interesting to inspect the use of molecular markers in breeding of perennial ryegrass, especially in hybrid breeding. In the present study, we first compiled a large germplasm set of 297 *Lolium perenne* accessions containing breeding material, ecotypes, landraces, varieties from various breeding companies and the IPK Genebank to represent much of the possible diversity of the European *Lolium perenne* germplasm pool. This germplasm set was genotyped with DArT, SNP, and SSR markers to investigate the genetic diversity as well as genetic distances among accessions. A subset of this germplasm set was used to produce hybrids and both parental material and hybrids were tested in the field to examine heterosis performance. Correlation studies, association mapping and genomic selection were conducted to inspect different possibilities to use molecular markers in breeding of perennial ryegrass.

5.1   Genetic diversity

As perennial ryegrass maintains a high level of diversity within the population (Bolaric et al. 2005b), individual genotypes might not be representative for the whole population. Thereby a bulk sampling strategy was employed and a special set of material was used to check the

proper bulk size for the genotyping experiments. Two main questions are whether bulk samples provided enough distinctiveness between accessions and if the proper number of individuals per bulk sample was selected. In the phenogram (Fig 2.2), these special accessions were clearly clustered, permitting to distinguish different accessions via bulk samples. This is consistent with Guthridge et al. (2001) who compared the discrimination between multiple individual samples and bulked sample with AFLP markers. In PCoA of this special set, it was also observed that with over 24 individuals per bulk the results were constant but with only few individuals the bulk sample results could still be rather variable (Fig. 2.3). Guthridge et al. (2001) proposed 20-30 individuals per sample for a reliable identification of accessions or cultivars for bulk sampling. The present study confirms this recommendation. Although bulk samples containing higher number of individuals could consistently provide adequate distinction between accessions, minor alleles might not be detected. For SSRs, the average number of detected alleles per locus was 8.2, which is lower than 9.9 (Wang et al. 2009), 13.3 (Brazauskas et al. 2011) and 19.4 (Kubik et al. 2001) that have been reported in other studies. Therefore it seems that if the detection of minor alleles is of major interest, individual genotypes should be used or when analyzing SSR also weak bands have to be considered.

Genetic distances within the germplasm set revealed high levels of diversity. The JD for DArT markers ranged from 0.00 to 0.73; the MRD for SNPs ranged from 0.03 to 0.52; the MRD for SSR markers ranged from 0.26 to 0.76 (Fig. 2.1). Nei´s genetic diversity for DArTs, SNPs and SSRs was 0.26, 0.32 and 0.45, respectively. The genetic diversity detected in the current study, however, was not higher than in related studies. Hu et al. (2011) reported genetic diversity of 0.28 within 75 perennial ryegrass accessions with dominant ISSR markers; Brazauskas et al. (2011) observed genetic diversity of 0.63 among 37 perennial ryegrass accessions with SSR markers. The number of accessions used in this study was much larger than in previous studies, but genetic diversity revealed was not necessarily higher. This contradiction might be explained by different source of diversity in different studies. It is common to observe higher diversity within accession rather than among them in perennial ryegrass (Guthridge et al. 2001; Bolaric et al. 2005b; Elazreg et al. 2011) and it is known that if the within population variability is high the variability between accession will be reduced because the common alleles between accessions are likely to increase (Guthridge et al. 2001). Bulked samples only revealed among accessions variability but the related studies were usually

based on individual genotypes. Therefore, these high diversity values for these studies could result from large within accession variation.

Clustering analyses were conducted to investigate the population structure in the germplasm set. STRUCTURE analysis suggested three subgroups. However, apart from the largest group containing 250 accessions, only 15 accessions could be assigned to other groups with adequate probability (Fig. 2.6); In PCo-based clustering, though four and three subgroups were revealed by DArT and SNP markers, respectively, the majority of the germplasm was assigned to one subgroup (Fig. 2.7, Fig. 2.8); in AMOVA based on passport data, none of geographical origin, contributor, ploidy level or breeding status could explain more than 10% of the variation (Tab. 2.4). These results indicated a lack of structure in the collected germplasm set which might be attributed to intensive usage of natural resources (Bolaric et al. 2005b; Brazauskas et al. 2011) and lack of establishment of heterotic pools in breeding practice (Brummer 1999).

5.2 Heterosis and genetic distance

Heterosis and correlation studies were conducted only for FMY and DMY. The total yield for two consecutive growth years was summed up for both traits. FMY and DMY were found to be highly correlated which is in accordance with Conaghan et al. (2008) who observed a correlation of 0.84 and proposed indirect selection solely based on FMY.

The maximum BPH estimated in the present study was 14.2% for FMY and 10.2% for DMY. Positive BPH and MPH were observed for the majority of hybrids. At current time, heterosis in *Lolium perenne* has been only examined for semi-hybrids due to the limited sources of CMS and difficulties in producing SI hybrids. By intercrossing 6 adapted perennial ryegrass varieties, Foster (1971a) reported PBH of 25% and 31% of two hybrids under spaced-plant condition. Under competitive sward condition, highest PBH was found to be 17% (Foster 1971b). However, the majority of semi-hybrids produced in these studies showed negative BPH. A higher ratio of positive BPH was obtained in our study, but highest BPH was not comparable to the previous results. The higher ratio of BPH could be attributed to the

exploitation of CMS system in the hybrid production because it eliminates the possible intercrossing within populations. The maximum BPH in the present study might be influenced by limited CMS accessions harboring constrained GCA. This might be manifested by GCA of the tetraploid CMS accessions, where both small value and small variance was acquired (Tab. 3.6).

It is noteworthy that the accessions used in the field trials are only a portion of all the genotyped material, and there is still a large amount of genetic resources that have not been utilized and their performance in producing population hybrids was unknown. The germplasm tested in the field as either pollinators or CMS accessions was highlighted with red and black color in Fig. 5.1. The green spots represent material that has been genotyped but was not used in the heterosis study. It can be seen that many diverse accessions revealed by molecular markers were not examined for their potential heterotic effect.
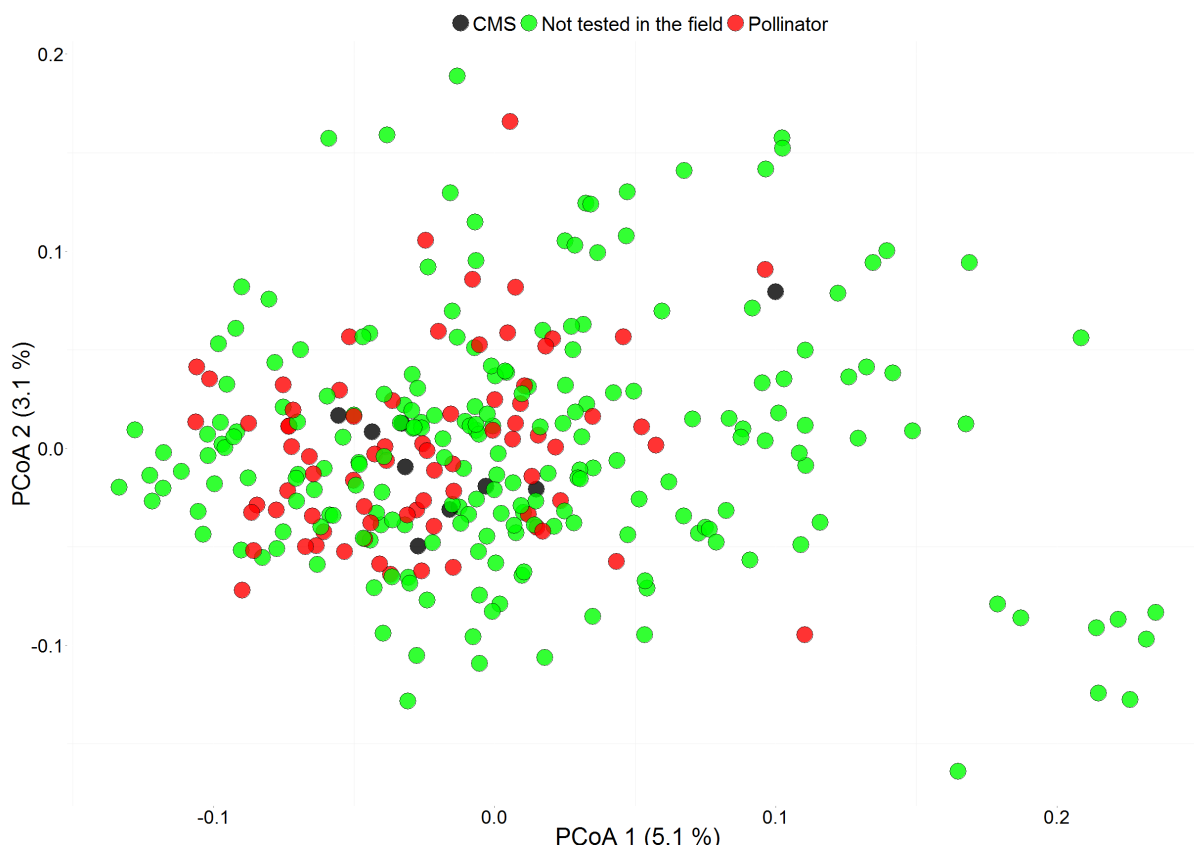

Figure 5.1 PCoA of DArTs on set I (297 accessions) showing the parental material tested in the field

The GD was positively correlated with hybrid performance and MPH but the correlation

coefficients were low and not of practical value (Tab. 3.9). The graphical representation of the relationships revealed that significant correlations observed in the 2011 sowings were mainly caused by three hybrids with high genetic distance and high BPH (Fig. 3.4). For the rest of the material, however, no clear correlations could be found. A higher yield resulted from higher genetic distances was reported by Kölliker et al. (2005) for the composition of synthetics. In other species, however, the correlation between GD and MPH was not consistently observed (Joshi et al. 2001; Jaikishan et al. 2010; Wu et al. 2013) despite of their theoretical relationship (Melchinger 1999). The possible reason for the low correlation might be attributed to the presence of epistasis and the fact that the markers used to estimate the GD are not linked to QTL for the trait (Reif et al. 2012).

Compared to the average genetic gain of around 4% per decade in DMY (Humphreys 1999), the BPH obtained in the study seemed to be interesting. However, the production of the hybrid seeds via CMS system is much more costly than the production of synthetics or open pollination varieties. In this perspective, a gain of a maximum of 14% might be not very promising. Heterotic patterns are needed in order to proceed with hybrid breeding. Brummer (1999) suggested that one should initially look for heterotic patterns from geographically distinct ecotypes. Although AMOVA showed little variance explained by geographic regions or contributors at the whole germplasm level (Tab. 2.4), some separation of material could be revealed in the PCoA. For example, many accessions from NPZ appeared to be distinct from others (Fig. 5.2). This separation might be caused by narrower genetic base because the majority of NPZ material was either clones or S1 synthetics constructed by intercrossing between only several clones. Increased inbreeding exploits the massive diversity within each accession and enlarges the variation between accessions; thereby, a clearer segregation could be observed. In addition, several accessions with Northern or Southern European origins appeared to be distinct from others as well (Fig. 5.3) which might correspond to Brummer (1999) who proposed geographical origin could be used to build heterotic pools in *Lolium perenne*. If the heterotic effect of this material could be confirmed in the field, it would be a good starting point for the formation of heterotic patterns.
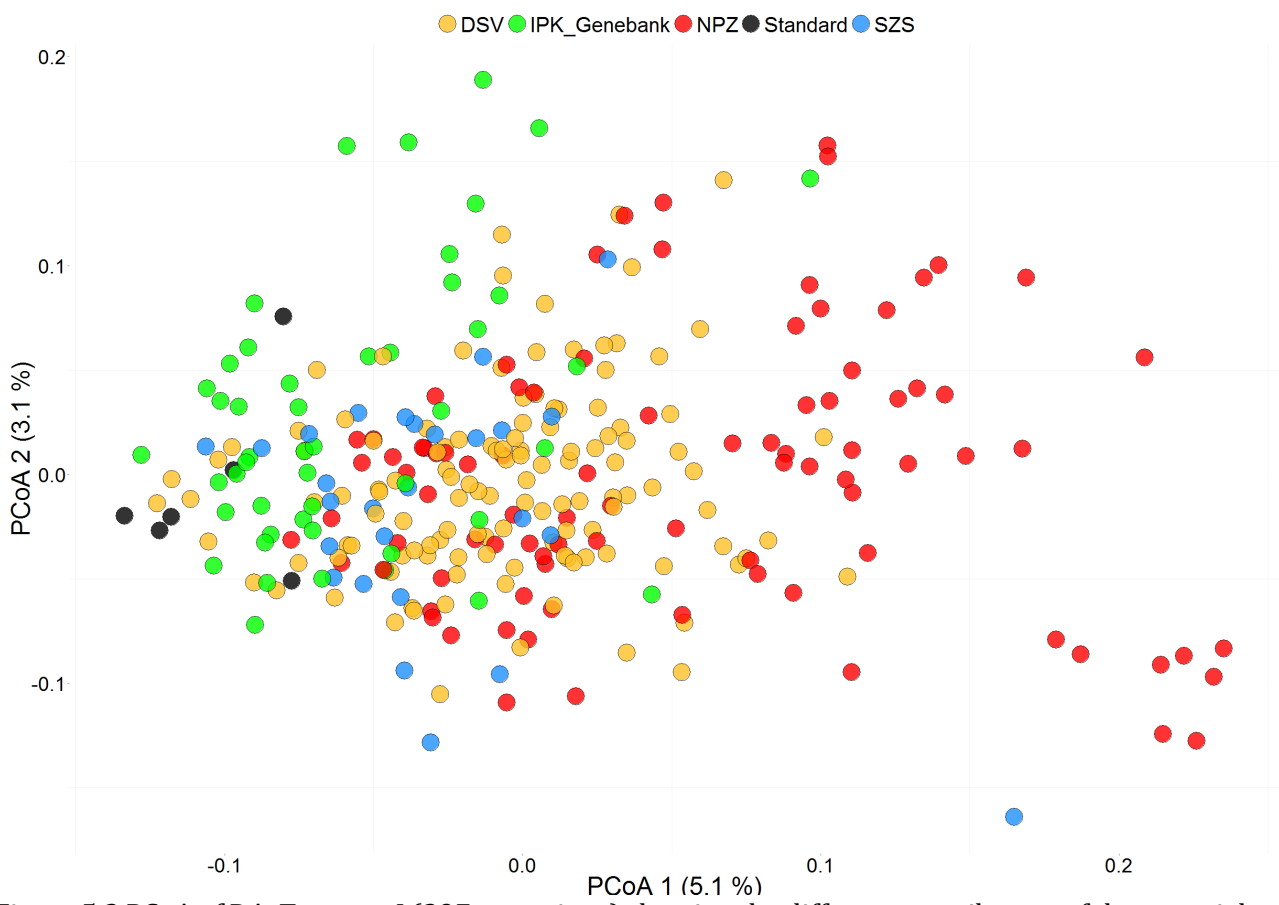
Figure 5.2 PCoA of DArTs on set I (297 accessions) showing the different contributors of the material
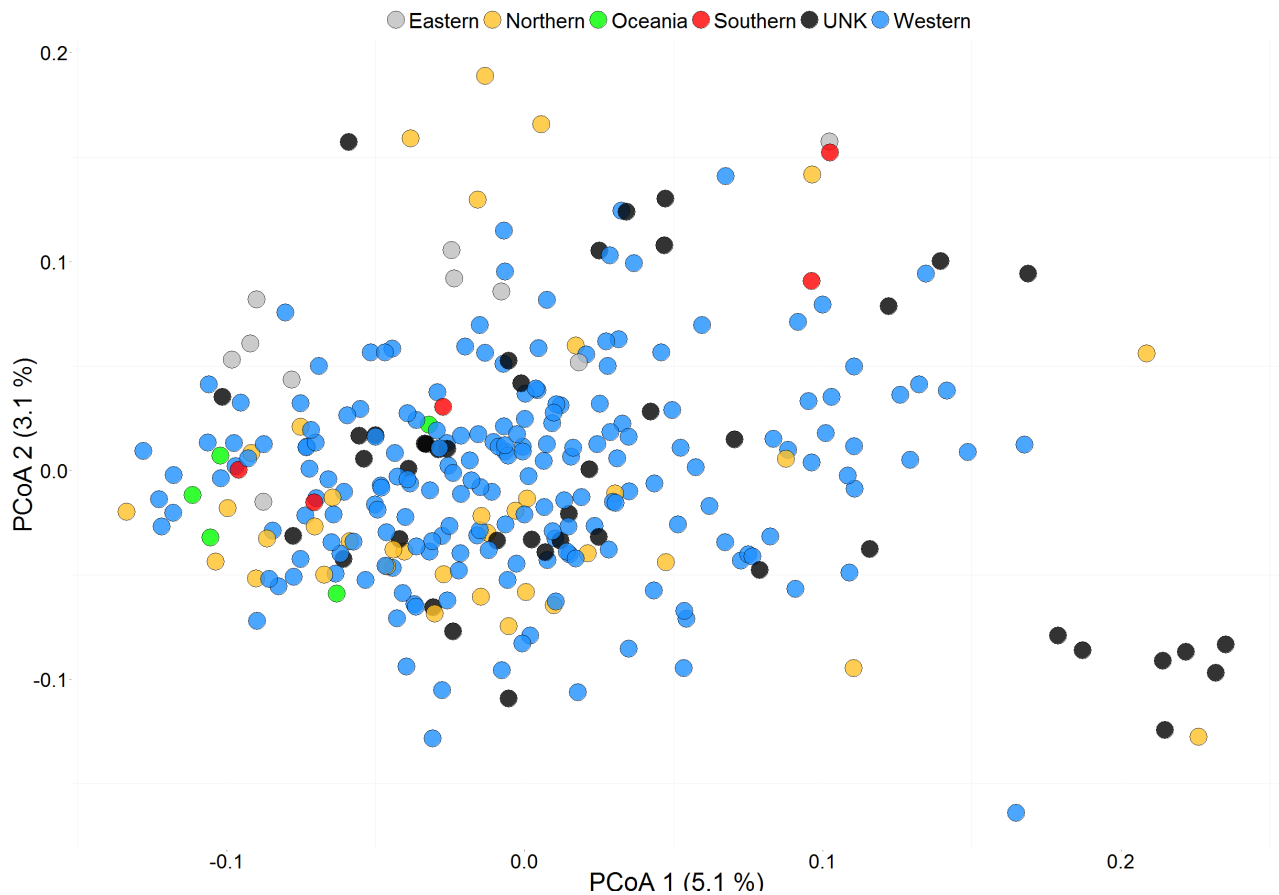
Figure 5.3 PCoA of DArTs on set I (297 accessions) showing the different geographical origin of the material

5.3   Association mapping and genomic selection

The objective of this project was hybrid breeding and the potential application of markers in assisting construction of population hybrids. However, with the genotype and phenotype data of the parental accessions, it is also possible to conduct association mapping and genomic selection. In the association study, a total of nine markers were identified for FMY, DMY, AH, EG, ADF, ADL. Some of the identified markers were on the same LG as previously reported with linkage mapping of bi-parental populations. For instance, Anhalt et al. (2009) uncovered two markers for FMY and three markers for DMY at LG2; Xiong et al. (2006) identified two markers at LG6 for ADF content. However, it is not possible to confirm whether these markers were linked to the same QTL due to different genetic maps and marker types among studies. Association mapping generally requires a large population size and a high marker density which is not attainable for the current dataset. This might explain the low

number of markers discovered in this study even though a lot of traits were analyzed.

In genomic selection, generally a low prediction accuracy was obtained. For FMY and DMY, however, a moderate prediction accuracy of up to 0.53 was found with SSRs. This accuracy has to be confirmed by further studies because the LD is small in perennial ryegrass which requires a high density of markers in genomic selection. With SSR markers from only 45 marker loci, this prerequisite is not fulfilled.

Both association mapping and genomic selection have been successfully used in many other species to find QTL and increase the selection efficiency, but they were rarely used in *Lolium perenne*. However, some proposals towards the potential implementation of both methods in perennial ryegrass have been given (Hayes et al. 2013). Therefore it is expected that they could be also further investigated and benefit the understanding and the breeding of perennial ryegrass.

5.4   Marker comparison

The genotyping of the bulked accessions was conducted with DArT, SNP and SSR markers which allowed the comparison of the three marker types. Contradictory relationships with different marker types in diversity studies were reported before (Jones et al. 1997; Posselt and Barre 2006; Simko et al. 2012). In this study we generally observed a high consistency between marker types, which is manifested by the moderate to high correlations among their resultant GD (Chapter 2), a similar ability to distinguish set III (Fig. 2.2), a similar shape of the distribution of the GD (Fig. 2.1, Fig. 3.1) and similar results obtained from clustering analysis (Fig. 2.7, Fig. 2.8). For the diversity study, DArT markers showed higher discriminative power, repeatability and consistency and therefore appeared to be superior to the other marker types. In spite of these advantages, it is noteworthy to stress that all the marker types are suitable for diversity study of perennial ryegrass with bulk samples. In addition, when considering the correlation to MPH, DArT markers did not clearly outperform the other two marker types.

5.5 Conclusion and further perspectives

The genotyping based on bulk samples provided adequate distinction and consistency in analyzing the diversity of obligate outbreeding perennial ryegrass accessions. Based on this method, large genetic diversity was found within the large *Lolium perenne* germplasm set. A clear population structure was not identified, which might correspond to the lower heterosis (maximum BPH of 14%) observed. With this level of heterosis, hybrid breeding might not be of economic interest due to the higher cost in production of hybrid seeds. Further investigations in the hybrid breeding should be aimed at the identification and maintenance of heterotic groups. With the knowledge that larger variation was harbored within each accession, some inbreeding might help to further increase the variation between accessions. Moreover, although the prediction of heterosis based on GD has to be further confirmed, the application of molecular markers should be very helpful in detecting heterotic patterns.

Chapter 6
Summary

Perennial ryegrass (*Lolium perenne* L.) is an outcrossing diploid species (2n = 2x = 14) and has high evolutionary lineage with rice, wheat and barley. It is an important forage grass in temperate regions but also serves as turf grass. Owing to its high yield potential and good resistance to both biotic and abiotic stresses, it is also suggested as alternative or supplement for biogas production. For this purpose, further enhancement of the biomass yield would be of fundamental importance. Novel breeding schemes are thereby required because the genetic gain per year in biomass yield is rather limited under the present breeding strategies.

In comparison to mass selection or the development of synthetic cultivars that are broadly used at present, hybrid breeding could provide better utilization of heterosis. The discovery of Cytoplasmic Male Sterile (CMS) in perennial ryegrass and the increasing accessibility of molecular markers could pave the path towards hybrid breeding.

To assess genetic diversity as well as population structures, a large germplasm set was genotyped by 1384 DArT, 182 SNP and 48 SSR (with 393 alleles) markers. This germplasm set (set I) consists of 297 *Lolium perenne* accessions with different breeding status (varieties, breeding material, ecotypes), ploidy levels (2x / 4x), geographical origin as well as donors (breeding companies, IPK Genebank). Due to the large genetic variation within each accession which is commonly observed for perennial ryegrass, a bulk sampling strategy with 30 individual plants per bulk, instead of individual plants, was used. In addition to set I, set II was sampled to compare the discriminative ability and the repeatability of different marker types. It consists of other *Lolium* species (two samples of *Lolium multiflorum* and one sample of *Lolium x hybridum*), material from the *Lolium perenne* VrnA mapping population (eight samples) and seven replicated samples taken from set I. To confirm the feasibility of bulk sampling strategy, set III containing 37 samples was developed based on four accessions from set I. The difference between samples from sets I, II and III lies in the variable size per bulk (1, 12, 24, 36, 48, 100 individuals instead of fixed 30 in set I and set II) used in set III. For all the aforementioned samples, Modified Roger`s distances (MRD) were calculated for SNP and SSR markers due to their co-dominant nature and Jaccard distances (JD) were estimated for

dominant DArT markers.

Parallel to the genotypic analysis, field experiments were conducted to inspect the heterosis by hybrids produced with CMS. A subset of set I consisting of CMS females and pollinators was used to produce these hybrids. The hybrids and their corresponding parents were assigned into two independent experiments sown in 2010 and 2011. In the 2010 sowings, 31 diploids and 10 tetraploids parents with 55 F1 hybrids derived from them were planted. In the 2011 sowings, 48 diploid parents and their resultant 48 F1 hybrids were tested. For each sowing, plants were grown in sward conditions at five locations and with two different nitrogen levels (optimal fertilization and 60% of that). Fresh matter yield (FMY) and dry matter yield (DMY) were measured for two successive years (2011/2012 for 2010 sowings and 2012/2013 for 2011 sowings) and the total yield was summed up. The better-parent heterosis (BPH) and mid-parent heterosis (MPH) were calculated from the yield performance of hybrids and their parents. Apart from FMY and DMY, 12 other traits including ADF content, ADL content, ash content, NDF content, protein content, OMD content, DNDF content, absence of heads (AH), early spring growth (EG), standing ability (SD), resistance to crown rust (RC), winter hardiness (WH) were also recorded, but phenotyping of these traits was mainly carried out at only one location.

Phenograms of set III based on the genotypic data revealed four clear clusters corresponding to the four repeated accessions independent of the marker type, indicating a sufficient discriminative power for all three marker types. In the Principal Coordinate Analysis (PCoA) for set III we further confirmed the success of bulk sampling strategy, which is 30 individuals/bulk, used in the experiment because if less than 12 individuals were bulked, a certain variation could still be observed between samples from the same accession; if the bulk size is larger than 24, consistent results could always be obtained.

The perennial ryegrass germplasm set I presented large genetic variation. For allelic polymorphism, of the 1384 DArT loci, 1380 were found to be polymorphic; all 182 SNP loci were polymorphic; the number of alleles for SSRs ranged from 2 to 23 with an average number of 8.2 alleles per locus. For the genetic distance estimates, the JD for DArT markers ranged from 0.00 to 0.73 with a mean distance of 0.45; for SNPs, the MRDs were between 0.03 and 0.52 with an average of 0.34; for SSRs, the MRDs ranged from 0.26 to 0.76 with a mean value of 0.54. Nei´s genetic diversity was 0.26, 0.32 and 0.54 for DArT, SNP and SSR

markers, respectively. Despite of high diversity, there was no clear population structure identified based on the results from AMOVA, STRUCTURE clustering and PCo-based clustering.

In the field trials, the genotypic variation for FMY and DMY was significant (p = 0.01) in both sowings and for both ploidy levels. Moderate to high broad-sense heritability (FMY: 0.67, 0.88, 0.70; DMY: 0.62, 0.84, 0.68 for diploids from the 2010, the 2011 sowings and tetraploids from the 2010 sowings, respectively) was observed for both traits. High correlation was found between FMY and DMY (r = 0.73, 0.90 for diploids from the 2010 and 2011 sowings, r = 0.78 for tetraploids from the 2010 sowings). MPH and BPH were observed in the majority of the F1 hybrids: for FMY in the 2010 sowings, diploid hybrids on average showed MPH of 13.88% (9.41% - 19.58%) and BPH of 3.23% (-2.31% - 14.24%), tetraploid hybrids displayed an average MPH of 6.85% (-1.16% - 14.31%) and BPH of 4.94% (-4.20% - 13.91%). In the 2011 sowings, diploid hybrids exhibited an average MPH of 3.24% (-3.99% - 10.40%) and BPH of -1.10% (-8.82% - 7.84%). The heterosis patterns for DMY were similar to that of FMY.

The correlations between genetic distances estimated by the molecular markers and MPH as well as F1 hybrid per se performance were positive in most cases. However, most of correlation coefficients were lower than 0.5. With this accuracy, the prediction of heterosis or hybrid performance solely based on GD might not be applicable in practice. Moreover, in the 2011 sowings, several correlations were significant at p = 0.05. However, graphical presentation of these correlations revealed that these significances were mainly caused by three hybrids possessing high genetic diversity and high MPH, but for hybrids with lower genetic distances no correlation could be observed. The enhancement of MPH caused by high genetic diversity has to be confirmed with further studies with more hybrids generated from highly diverse parents.

With the genotypic and the phenotypic data of 46 diploid parental accessions in the 2011 sowings, association mapping (AM) and genomic selection (GS) were also conducted. Based on the standard of minor allelic frequencies > 0.1 and missing value rate < 0.3 within the 46 accessions, 800 DArT, 134 SNP and 45 SSR (with 121 alleles) markers were utilized for both analyses.

In AM, a total of nine markers were identified for different traits: one DArT with unknown position and one SNP marker on LG 2 were detected for FMY; the same SNP marker was also

found to be accountable for DMY; two SNP markers on LG4 and LG5 were identified for AH; two SNP markers on LG2 and LG5 were detected for EG; one SSR allele on LG6 and one SSR allele on LG 4 were identified for ADF and ADL content, respectively.

In GS, two scenarios were simulated: scenario 1) variable training set for model training and variable validation set for model performance testing; scenario 2) fixed training set and validation set. Scenario 1 was applied to FMY and DMY, where the phenotypic data were tested at five locations to observe the performance of GS with different training size and determine the optimized condition for scenario 2, which was then applied to the other agronomic traits. For each training set and validation set combination, 500 cross validations were simulated. As results, in scenario 1 the prediction accuracy measured as average correlation coefficient over 500 cross validations increased along with the increase of the training set. At size of 18 - 21 the increase tended to reach a plateau but the maximum r values were usually observed with training set size of 30. Therefore in scenario 2 the training set was fixed to 30. The resultant average r values were lower than that for FMY and DMY. In addition, none of the marker types could consistently outperform the others.

In conclusion, the genotyping based on bulk samples provided adequate distinction and consistency in analyzing the diversity of perennial ryegrass accessions. Based on this method, large genetic diversity was found within the large *Lolium perenne* germplasm set. No clear population structure was identified, which might correspond to the relatively low heterosis (maximum BPH of 14%) observed. With this level of heterosis, hybrid breeding might not be of economic interest due to the higher cost in production of hybrid seeds. Further investigations on the hybrid breeding should be aimed at identification and maintenance of heterotic groups. Application of molecular markers might serve as efficient tool in assisting this process.

Bibliography

Abdel-Mawgood AL (2012) DNA based techniques for studying genetic diversity. Genet Divers Microorg Tech, Croat 95–122.

Aguirre AA, Studer B, Frei U, Lübberstedt T (2012) Prospects for hybrid breeding in bioenergy grasses. BioEnergy Res 5:10–19.

Al-Maskri AY, Sajjad M, Khan SH (2012) Association mapping: a step forward to discovering new alleles for crop improvement. Int J Agric Biol 14:153–160.

Altpeter F, Xu J, Ahmed S, Gentransfer AG (2000) Generation of large numbers of independently transformed fertile perennial ryegrass (*Lolium perenne* L .) plants of forage- and turf-type cultivars. 6:519–528.

Anhalt UCM, Pat JSH-H, Piepho HP, et al. (2009) Quantitative trait loci mapping for biomass yield traits in a *Lolium* inbred line derived F2 population. Euphytica 170:99–107.

Armstead IP, Turner LB, Marshall AH, et al. (2008) Identifying genetic components controlling fertility in the outcrossing grass species perennial ryegrass (*Lolium perenne*) by quantitative trait loci analysis and comparative genetics. New Phytol 178:559–571.

Barrett BA, Turner MA, Lyons TB, et al. (2010) Evaluation of semi-hybrid perennial ryegrass populations. Proc. New Zeal. Grassl. Assoc. pp 11–16

Berg KD, Glaser CL, Thompson RE, et al. (2000) Detection of microsatellite instability by fluorescence multiplex polymerase chain reaction. J Mol Diagnostics 2:20–28.

Bert PF, Charmet G, Sourdille P, et al. (1999) A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers. Theor Appl Genet 99:445–452.

Bolaric S, Barth S, Melchinger a. E, Posselt UK (2005a) Genetic diversity in European perennial ryegrass cultivars investigated with RAPD markers. Plant Breed 124:161–166.

Bolaric S, Barth S, Melchinger AE, Posselt UK (2005b) Molecular genetic diversity within and among German ecotypes in comparison to European perennial ryegrass cultivars. Plant Breed 124:257–262.

Boller B, Posselt UK, Veronesi F (2010) Fodder crops and amenity grasses. Springer

Börner A, Khlestkina EK, Pshenichnikova TA, et al. (2012) Genetics and genomics of plant genetic resources. J. Stress Physiol. Biochem. 8(3).

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314.

Brazauskas G, Lenk I, Pedersen MG, et al. (2011) Genetic variation, population structure, and linkage disequilibrium in European elite germplasm of perennial ryegrass. Plant Sci 181:412–420.

Brummer EC (1999) Capturing heterosis in forage crop cultivar development. Crop Sci 39:943–954.

Burton GW (1948) performance of various mixtures of hybrid and parent inbred pearl millet, *Pennisetum glaucum* (L.) R. BR. J. Am. Soc. Agron.

Calsyn E, Ghesquiere A, Baert J, Riek J de (2005) Study of genetic diveristy between and within ryegrass populations of ECPGR collection by means of AFLP markers. Rep. 8th Meet. ECPGR Work. Gr. forages. Rome, pp 122–131

Cardinal A, Lee M, Moore K (2003) Genetic mapping and analysis of quantitative trait loci affecting fiber and lignin content in maize. Theor Appl Genet 106:866–874.

Casler MD (2001) Breeding forage crops for increased nutritional value. Adv Agron 71:51–107.

Casler MD, Brummer EC (2008) Theoretical Expected Genetic Gains for Among-and-Within-Family Selection Methods in Perennial Forage Crops. Crop Sci 48:890-902.

Casler MD, Pedersen JF, Eizenga GC, Stratton SD (1996) Germplasm and cultivar development. Cool forage grasses 413–469.

Cheng-Xiang A, Yu X-M, Shen G-N, et al. (2012) Allele frequency analysis of Chinese chestnut (Castanea mollissima) populations using fluorescent simple sequence repeats (SSR) analysis. Afr J Biotechnol 11:13767–13774.

Cogan NOI, Smith KF, Yamada T, et al. (2005) QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 110:364–380.

Conaghan P, Casler MD, O'Kiely P, Dowley LJ (2008) Efficiency of indirect selection for dry matter yield based on fresh matter yield in perennial ryegrass sward plots. Crop Sci 48:127–133.

Cornish MA, Hayward MD, Lawrence MJ (1979) Self-incompatibility in ryegrass. Heredity (Edinb) 43:129–136.

Cruz VM V, Kilian A, Dierig DA (2013) Development of DArT Marker Platforms and Genetic Diversity Assessment of the U.S. Collection of the New Oilseed Crop Lesquerella and Related Species. PLoS One 8:e64062.

Devey FD, Hayward MD, Kearsey MJ, et al. (1989) Genetic Analysis of Production Characters in *Lolium*. Plant Breed 103:63–72.

Dewey DR (1980) Some applications and misapplications of induced polyploidy to plant breeding. Polyploidy - Biol. Relev. Springer, pp 445–470

Duvick DN (1992) Genetic contributions to advances in yield of US maize. Maydica (Italy)

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361.

Elazreg H, Chtourou-Ghorbel N, Ghariani S, et al. (2011) Studying genetic diversity of the Tunisian *Lolium perenne* and *Festuca arundinacea* with AFLP markers. J Food, Agric Environ 9:409–415.

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255.

Foster CA (1973) Interpopulational and intervarietal F1 hybrids in *Lolium perenne*: performance in field sward conditions. J Agric Sci 80:463–477.

Foster CA (1971a) Interpopulational and intervarietal hybridization in *Lolium perenne* breeding: heterosis under non-competitive conditions. J Agric Sci 76:107–130.

Foster CA (1971b) Interpopulational and intervarietal F1 hybrids in *Lolium perenne*: Heterosis under simulated-sward conditions. J Agric Sci 76:401–409.

Fraley C, Raftery AE, others (2007) Model-based methods of classification: Using the mclust software in chemometrics. J Stat Softw 18:1–13.

Frandsen KJ (1986) Variability and Inheritance of Digestibility in Perennial Ryegrass (*Lolium perenne*), Meadow Fescue (*Festuca pratensis*) and Cocksfoot (*Dactylis glomerata*) II. F1 and F2 Progeny. Acta Agric Scand 36:241–263.

Garcia AAF, Benchimol LL, Barbosa AMM, et al. (2004) Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. Genet Mol Biol 27:579–588.

Gaue I, Baudis H (2006) Male sterility in grasses of the genus *Lolium*. U.S. Patent Application No. 11/520,186.

Ghariani S, Trifi-Farah N, Chakroun M, et al. (2003) Genetic diversity in Tunisian perennial ryegrass revealed by ISSR markers. Genet Resour Crop Evol 50:809–815.

Gilbert JE, Lewis R V, Wilkinson MJ, Caligari PDS (1999) Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theor Appl Genet 98:1125–1131.

Guthridge KM, Dupal MP, Kölliker R, et al. (2001) AFLP analysis of genetic diversity within and between populations of perennial ryegrass (*Lolium perenne* L.). Euphytica 122:191–201.

Hannaway DB, Fransen S, Cropper JB, et al. (1999) Perennial ryegrass (*Lolium perenne* L.).

Hayes BJ, Cogan NOI, Pembleton LW, et al. (2013) Prospects for genomic selection in forage plant species. Plant Breed 132:133–143.

Hayes BJ, Goddard ME, others (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Hayward MD, Forster JW, Jones JG, et al. (1998) Genetic analysis of *Lolium*. I. Identification of linkage groups and the establishment of a genetic map. Plant Breed 117:451–455.

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49:1–12.

Hill J, Becker HC, Tigerstedt PMA (1998) Quantitative and ecological aspects of plant breeding. Wiley Online Library

Hu T, Li H, Li D, et al. (2011) Assessing genetic diversity of perennial ryegrass (*Lolium perenne* L.) from four continents by inter-simple sequence repeat (ISSR) markers. African J Biotechnol 10:19365–19374.

Huang X, Feng Q, Qian Q, et al. (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076.

Humphreys M, Feuerstein U, Vandewalle M, Baert J (2010) Ryegrasses. Fodd. Crop. amenity grasses. Springer, pp 211–260

Humphreys MO (1999) The contribution of conventional plant breeding to forage crop improvement 2000. Proc. XVIII Int. Grassl. Congr. Winnipeg, Manitoba Saskatoon, (Vol. 3), (Eds J. G. Buchanan-Smith, L. D. Bailey P. McCaughey). pp 71–77

Humphreys MO (2005) Genetic improvement of forage crops – past, present and future. J Agric Sci 143:441-448.

Islam M, Studer B, Møller IM, et al. (2014) Genetics and biology of cytoplasmic male sterility and its applications in forage and turf grass breeding. Plant Breed 133:299–312.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. . Bulletin de la Socièté Vaudense des Sciences Naturelles 44: 223-270.

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25–e25.

Jaikishan I, Rajendrakumar P, Ramesha MS, et al. (2010) Prediction of heterosis for grain yield in rice using "key"informative EST-SSR markers. Plant Breed 129:108–111.

James KE, Schneider H, Ansell SW, et al. (2008) Diversity arrays technology (DArT) for pan-genomic evolutionary studies of non-model organisms. PLoS One 3:e1682.

Jancik F, Homolka P, Cermak B, et al. (2008) Determination of indigestible neutral detergent fibre contents of grasses and its prediction from chemical composition. Czech J. Anim. Sci. (Czech Republic)

Jančík F, Koukolová V, Homolka P, Haman J (2011) Comparison of analyses to predict ruminal fibre degradability and indigestible fibre in temperate grass silages. S Afr J Anim Sci 41:297–308.

Jensen LB, Andersen JR, Frei U, et al. (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.

Jonah PM, Bello LL, Lucky O, et al. (2011) Review : The Importance of Molecular Markers in Plant Breeding Programmes. Glob. J. Sci. Front. Res. 11(5).

Jones CJ, Edwards KJ, Castaglione S, et al. (1997) Reproducibility testing of RAPD , AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3:381–390.

Jones ES, Mahoney NL, Hayward MD, et al. (2002) An enhanced molecular marker based genetic map of perennial ryegrass (*Lolium perenne*) reveals comparative relationships with other Poaceae genomes. Genome 45:282–295.

Jones ES, Sullivan H, Bhattramakki D, Smith JSC (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet 115:361–371.

Jones N, Ougham H, Thomas H, Pašakinskiene I (2009) Markers and mapping revisited: finding your gene. New Phytol 183:935–966.

Joshi SP, Bhave SG, Chowdari K V, et al. (2001) Use of DNA markers in prediction of hybrid performance and heterosis for a three-line hybrid system in rice. Biochem Genet 39:179–200.

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol 125:1198–1205.

King J, Thomas A, James C, et al. (2013) A DArT marker genetic map of perennial ryegrass (*Lolium perenne* L.) integrated with detailed comparative mapping information; comparison with existing DArT marker genetic maps of *Lolium perenne*, *L. multiflorum* and *Festuca pratensis*. BMC Genomics 14:437.

Kirkpatrick BW, Cowan CM, Dentine MR (1991) Differential amplification of alleles: potential for misclassification with PCR genotyping. Anim Biotechnol 2:1–14.

Kölliker R, Boller B, Widmer F (2005) Marker assisted polycross breeding to increase diversity and yield in perennial ryegrass (*Lolium perenne* L.). Euphytica 146:55–65.

Kopecky D, Bartos J, Lukaszewski A, et al. (2009) Development and mapping of DArT markers within the *Festuca - Lolium* complex. BMC Genomics 10:473.

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29.

Koukolová V, Weisbjerg MR, Hvelplund T, et al. (2004) Prediction of NDF degradation characteristics of grass and grass/clover forages based on laboratory methods. J Anim Feed Sci 13:691–708.

Kraft T, Säll T (1999) An evaluation of the use of pooled samples in studies of genetic variation. Heredity (Edinb) 82:488–494.

Kubik C, Sawkins M, Meyer WA, Gaut BS (2001) Genetic diversity in seven perennial ryegrass (*Lolium perenne* L.) cultivars based on SSR markers. Crop Sci 41:1565–1572.

Lamkey KR, Edwards JW (1999) Quantitative genetics of heterosis. Genet Exploit heterosis Crop ASA, CSSA, SSSA, Madison, WI 31–48.

Lewandowski I, Scurlock JMO, Lindvall E, Christou M (2003) The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. Biomass and Bioenergy 25:335–361.

Li J, Lühmann A-K, Weißleder K, Stich B (2011) Genome-wide distribution of genetic diversity and linkage disequilibrium in elite sugar beet germplasm. BMC Genomics 12:484.

Luetke Entrup E (2008) Forage and Amenity grasses. In: G. Roebbeleben (ed.) The development of plant breeding in Germany (1908 - 2008). Vor Pflanzenzüchtg 75:415–425.

Marsden JE, Schwager SJ, May B (1987) Single-locus inheritance in the tetraploid treefrog *Hyla versicolor* with an analysis of expected progeny ratios in tetraploid organisms. Genetics 116:299–311.

Melchinger AE (1999) Genetic diversity and heterosis. The genetics and exploitation of heterosis in crops. 10

Michelmore R, Paran I, Kesseli R V (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci 88:9828–9832.

Montilla-Bascón G, Sánchez-Martín J, Rispail N, et al. (2013) Genetic diversity and population structure among oat cultivars and landraces. Plant Mol Biol Report 31:1305–1314.

Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 147:969–977.

Nair RM (2004) Developing tetraploid perennial ryegrass (*Lolium perenne* L.) populations. New Zeal J Agric Res 47:45–49.

Nei M (1987) Molecular evolutionary genetics. Columbia University Press

Nestmann S, Sretenovic Rajicic T, Dehmer KJ, et al. (2011) Plant species diversity and composition of experimental grasslands affect genetic differentiation of *Lolium perenne* populations. Mol Ecol 20:2188–203.

Nestmann, S., Sretenovic Rajicic, T., Dehmer, K. J., Fischer, M., Schumacher, J., & Roscher, C. (2011). Plant species diversity and composition of experimental grasslands affect genetic differentiation of *Lolium perenne* populations. Molecular Ecology, 20(10), 2188–2203.

Nousiainen J, Ahvenjärvi S, Rinne M, et al. (2004) Prediction of indigestible cell wall fraction of grass silage by near infrared reflectance spectroscopy. Anim Feed Sci Technol 115:295–311.

O'Donovan M, Dillon P (1999) Measurement of grassland management practice on commercial dairy farms. Final Rep. Proj.

Ofori A, Becker HC (2008) Breeding of Brassica rapa for Biogas Production: Heterosis and Combining Ability of Biomass Yield. BioEnergy Res 1:98–104.

Pesek J, Baker RJ (1971) Comparison of predicted and observed responses to selection for yield in wheat. Can J Plant Sci 51:187–192.

Poehlman JM (1994) Breeding field crops. Iowa State University Press

Poland J, Endelman J, Dawson J, et al. (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103–113.

Posselt U, Barre P (2006) Comparative analysis of genetic similarity between perennial ryegrass genotypes investigated with AFLPs, ISSRs, RAPDs and SSRs. Czech J Genet Plant Breed 3:87–94.

Posselt UK (2010a) Breeding methods in cross-pollinated species. Fodd. Crop. amenity grasses. Springer, pp 39–87

Posselt UK (2010b) Identification of heterotic patterns in perennial ryegrass. Sustain. Use Genet. Divers. Forage Turf Breed. Springer, pp 569–572

Posselt UK, Monjardino P, Câmara A da, et al. (2001) Constraints in the selection of parents for synthetic cultivars. Breed. Stress Toler. Fodd. Crop. amenity grasses. Proc. 23rd Meet. Fodd. Crop. Amenity Grasses Sect. EUCARPIA, Azores, Port. 1-4 Oct. 2000. pp 34–39

Potter LR (1987) Effect of crown rust on regrowth, competitive ability and nutritional quality of perennial and Italian ryegrasses. Plant Pathol 36:455–461.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Core Team (2013) R: A Language and Environment for Statistical Computing.

Rasmussen S, Parsons AJ, Fraser K, et al. (2008) Metabolic profiles of *Lolium perenne* are differentially affected by nitrogen supply, carbohydrate content, and fungal endophyte infection. Plant Physiol 146:1440–1453.

Reif JC, Hahn V, Melchinger AE (2012) Genetic basis of heterosis and prediction of hybrid performance. Helia 35:1–8.

Reif JC, Melchinger AE, Frisch M (2005) Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Sci 45:1–7.

Rhodes I (1971) The relationship between productivity and some components of canopy structure in ryegrass (*Lolium* spp.). J Agric Sci 77:283–292.

Roldàn-Ruiz I, Dendauw J, Van Bockstaele E, et al. (2000) AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). Mol Breed 6:125–134.

Roldán-Ruiz I, Van Euwijk FA, Gilliland TJ, et al. (2001) A comparative study of molecular and morphological methods of describing relationships between perennial ryegrass (*Lolium perenne* L.) varieties. Theor Appl Genet 103:1138–1150.

Ruge B, Linz A, Gaue I, et al. (2003) Molecular characterization of cytoplasmic male sterility in *Lolium perenne*. Vor Pflanzenzüchtg 59:121–127.

Safari H, Shirvani H, Jafari AA, et al. (2014) The study of genetic variation for *Lolium perenne* using ISSR molecular markers. Int J Biosci 4:75–81.

Salces BM, Ahring BK, Uellendahl HW (2013) Catch crops as an alternative biomass feedstock for biogas plants. Int. Anaerob. Dig. Symp. BiogasWorld 2013. pp 92–98

Searchinger T, Heimlich R, Houghton RA, et al. (2008) Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319:1238–1240.

Shinozuka H, Cogan NOI, Spangenberg GC, Forster JW (2012) Quantitative Trait Locus (QTL) meta-analysis and comparative genomics for candidate gene prediction in perennial ryegrass (*Lolium perenne* L.). BMC Genet 13:101.

Simko I, Eujayl I, van Hintum TJL (2012) Empirical evaluation of DArT, SNP, and SSR marker-systems for genotyping, clustering, and assigning sugar beet hybrid varieties into populations. Plant Sci 184:54–62.

Skøt L, Humphreys J, Humphreys MO, et al. (2007) Association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne* (L.). Genetics 177:535–547.

Skøt L, Humphreys MO, Armstead I, et al. (2005) An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). Mol Breed 15:233–245.

Studer B, Kölliker R, Muylle H, et al. (2010) EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (Lolium spp.). BMC Plant Biol 10:177.

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.

Tomaszewski C, Byrne SL, Foito A, et al. (2012) Genetic linkage mapping in an F2 perennial ryegrass population using DArT markers. Plant Breed 131:345–349.

Utz H. (2011) PLABSTAT: A computer program for the statistical analysis of plant breeding experiments. Inst. Plant Breeding, Seed Sci. Popul. Genet. Univ. Hohenheim, Ger.

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299.

Vogeland KP, Pedersen JF (1993) Breeding systems for cross-pollinated perennial grasses. Plant Breed Rev 11:251-274.

Wang J, Dobrowolski MP, Cogan NOI, et al. (2009) Assignment of individual genotypes to specific forage cultivars of perennial ryegrass based on SSR markers. Crop Sci 49:49–58.

Wang Y, Mette MF, Miedaner T, et al. (2014) The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. BMC Genomics 15:556.

Warburton ML, Reif JC, Frisch M, et al. (2008) Genetic diversity in CIMMYT nontemperate maize germplasm: landraces, open pollinated varieties, and inbred lines. Crop Sci 48:617–624.

Weiland P (2007) Country updates on biogas, Germany. Present. 7th Meet. IEA Bioenergy Task 37 Work. Berlin. Available www. iea-biogas. net

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer New York

Wilkins PW (1989) Genotype/harvesting frequency and genotype/nitrogen level interactions for annual dry matter yield in *Lolium perenne* in relation to breeding. Euphytica 41:207–214.

Wilkins PW (1991) Breeding perennial ryegrass for agriculture. Euphytica 52:201–214.

Wilkins PW, Humphreys MO (2003) Progress in breeding perennial forage grasses for temperate agriculture. J Agric Sci 140:129–150.

Wilkins PW, Lovatt JA (2004) Recent gains from forage grass breeding. IGER Innov 8:18–21.

Wit F (1974) Cytoplasmic male sterility in ryegrasses (*Lolium* spp.) detected after intergeneric hybridization. Euphytica 23:31–38.

Wright S (1978) Evolution and genetics of populations. Vol. IV. The University of Chicago Press, Chicago

Wu J-W, Hu C-Y, Shahid MQ, et al. (2013) Analysis on genetic diversification and heterosis in autotetraploid rice. Springerplus 2:439.

Xiong Y, Fei S, Brummer EC, et al. (2006) QTL analyses of fiber components and crude protein in an annual × perennial ryegrass interspecific hybrid population. Mol Breed 18:327–340.

Yamada T, Jones ES, Cogan NOI, et al. (2004) QTL analysis of morphological, developmental, and winter hardiness-associated traits in perennial ryegrass. Crop Sci 44:925–935.

Yu J, Pressoir G, Briggs WH, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208.

Yu X, Bai G, Luo N, et al. (2011) Association of simple sequence repeat (SSR) markers with submergence tolerance in diverse populations of perennial ryegrass. Plant Sci 180:391–398.

Zhao Y, Gowda M, Liu W, et al. (2012) Accuracy of genomic selection in European maize elite breeding populations. Theor Appl Genet 124:769–776.

Appendix

Passport data of germplasm set I containing 297 *Lolium perenne* L. accessions

	Geographic Origin	Breeding Status	Contributor	Sampling	Ploidy
Alligator	Western	variety	Standard	Population	4x
Aubisque	Western	variety	Standard	Population	4x
Argoal	Western	variety	Standard	Population	2x
Delphin	Northern	variety	Standard	Population	4x
Lipresso	Western	variety	Standard	Population	2x
Fennema	Western	variety	Standard	Population	2x
2040391	Western	Breeding Material	DSV	Population	2x
2040832	Western	Breeding Material	DSV	Population	2x
2040392	Western	Breeding Material	DSV	Population	2x
2040472	Western	Breeding Material	DSV	Population	2x
2040371	Western	Breeding Material	DSV	Population	2x
2030038	Western	Breeding Material	DSV	Population	2x
2030270	Western	Breeding Material	DSV	Population	2x
6010	Western	Breeding Material	DSV	Population	2x
6015	Western	Breeding Material	DSV	Population	2x
6017	Western	Breeding Material	DSV	Population	2x
6018	Western	Breeding Material	DSV	Population	2x
6023	Western	Breeding Material	DSV	Population	2x
6027	Western	Breeding Material	DSV	Population	2x
6030	Western	Breeding Material	DSV	Population	2x
6032	Western	Breeding Material	DSV	Population	2x
6033	Western	Breeding Material	DSV	Population	2x
6034	Western	Breeding Material	DSV	Population	2x
6045	Western	Breeding Material	DSV	Population	2x
6046	Western	Breeding Material	DSV	Population	2x
6048	Western	Breeding Material	DSV	Population	2x
6050	Western	Breeding Material	DSV	Population	2x
6051	Western	Breeding Material	DSV	Population	2x
6052	Western	Breeding Material	DSV	Population	2x
6055	Western	Breeding Material	DSV	Population	4x
6057	Western	Breeding Material	DSV	Population	4x
6058	Western	Breeding Material	DSV	Population	4x
6062	Western	Breeding Material	DSV	Population	4x
6064	Western	Breeding Material	DSV	Population	4x
6066	Western	Breeding Material	DSV	Population	4x
6068	Western	Breeding Material	DSV	Population	4x
6070	Western	Breeding Material	DSV	Population	4x
6071	Western	Breeding Material	DSV	Population	4x
6072	Western	Breeding Material	DSV	Population	4x
6073	Western	Breeding Material	DSV	Population	4x
ASTURION	Western	variety	DSV	Population	2x
ASTORGA	Western	variety	DSV	Population	2x
ZLp_96-024	Western	Unknown	DSV	Population	2x

ASMIR	Western	variety	DSV	Population	4x
LIMBOS	Western	variety	DSV	Population	4x
OCTAVIO	Western	variety	DSV	Population	2x
VAUDAIRE	Western	variety	DSV	Population	2x
SURES	Western	variety	DSV	Population	4x
ASTONENERGY	Northern	variety	DSV	Population	4x
Lp_9928D	Oceania	Unknown	DSV	Population	2x
BANGUET_LE	Oceania	variety	DSV	Population	2x
QUARTET_LE	Oceania	variety	DSV	Population	2x
KRC_6625	Oceania	Unknown	DSV	Population	2x
KRC_6626	Oceania	Unknown	DSV	Population	2x
SLp092046	Northern	ecotype	DSV	Population	2x
SLp092052	Northern	ecotype	DSV	Population	2x
SLp092044	Northern	ecotype	DSV	Population	2x
SLp092048	Northern	ecotype	DSV	Population	2x
SLp092039	Northern	ecotype	DSV	Population	2x
SLp092003	Northern	ecotype	DSV	Population	2x
SLp092050	Northern	ecotype	DSV	Population	2x
SLp092017	Northern	ecotype	DSV	Population	2x
SLp_080901	Northern	ecotype	DSV	Population	2x
CMS_236_A	UNK	Breeding Material	NPZ	Population	2x
CMS_213_A	UNK	Breeding Material	NPZ	Population	2x
CMS_237_A	UNK	Breeding Material	NPZ	Population	2x
Inoval	Western	variety	NPZ	Population	2x
Urspr.klon_CMS_206	UNK	Breeding Material	NPZ	Clone	2x
Urspr.klon_CMS_213	UNK	Breeding Material	NPZ	Clone	2x
Urspr.klon_CMS_215	Western	Breeding Material	NPZ	Clone	2x
Urspr.klon_CMS_216	Western	Breeding Material	NPZ	Clone	2x
Urspr.klon_CMS_236	UNK	Breeding Material	NPZ	Clone	2x
Urspr.klon_CMS_237	Western	Breeding Material	NPZ	Clone	2x
Urspr.klon_CMS_238	Western	Breeding Material	NPZ	Clone	2x
Urspr.klon_CMS_240	UNK	Breeding Material	NPZ	Clone	2x
Urspr.klon_CMS_246	UNK	Breeding Material	NPZ	Clone	2x
Maint._MSF1_403	Western	Breeding Material	NPZ	Population	2x
Maint._MSF1_459	Western	Breeding Material	NPZ	Population	2x
Maint._MSF1_460	UNK	Breeding Material	NPZ	Population	2x
Best._MSF1_451	UNK	Breeding Material	NPZ	Population	2x
Best._MSF1_461	UNK	Breeding Material	NPZ	Population	2x
Best._MSF1_500	Western	Breeding Material	NPZ	Population	2x
Best._MSF1_502	Northern	Breeding Material	NPZ	Population	2x
Best._MSF1_503	Western	Breeding Material	NPZ	Population	2x
Best._MSF1_504	Western	Breeding Material	NPZ	Population	2x
Best._MSF1_507	Western	Breeding Material	NPZ	Population	2x
Best._MSF1_508	Western	Breeding Material	NPZ	Population	2x
Best._MSF1_509	Northern	Breeding Material	NPZ	Population	2x
Best._MSF1_510	UNK	Breeding Material	NPZ	Population	2x
Best._MSF1_511	Western	Breeding Material	NPZ	Population	2x
PC_08_2805	Western	Breeding Material	NPZ	Population	2x
PC_08_2808	UNK	Breeding Material	NPZ	Population	2x

PC_08_2902	Southern	ecotype	NPZ	Population	2x
PC_08_3006	Eastern	ecotype	NPZ	Population	2x
PC_08_3008	Southern	ecotype	NPZ	Population	2x
PC_09_2902	UNK	Breeding Material	NPZ	Population	2x
PC_09_2903	Northern	Breeding Material	NPZ	Population	2x
PC_09_2908	UNK	Breeding Material	NPZ	Population	2x
Aberavon	Northern	variety	NPZ	Population	2x
Arakan	Western	variety	NPZ	Population	2x
Urspr.klon_CMS_217	UNK	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_221	Western	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_225	UNK	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_228	Western	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_229	UNK	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_230	Northern	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_231	Western	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_232	UNK	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_233	Western	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_234	Northern	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_235	Northern	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_239	Western	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_241	Western	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_242	UNK	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_243	UNK	Breeding Material	NPZ	Clone	4x
Urspr.klon_CMS_244	UNK	Breeding Material	NPZ	Clone	4x
Maint._MSF1_384	UNK	Breeding Material	NPZ	Population	4x
Maint._MSF1_423	Western	Breeding Material	NPZ	Population	4x
Best._MSF1_455	UNK	Breeding Material	NPZ	Population	4x
Best._MSF1_464	UNK	Breeding Material	NPZ	Population	4x
Best._MSF1_470	UNK	Breeding Material	NPZ	Population	4x
Best._MSF1_471	UNK	Breeding Material	NPZ	Population	4x
Best._MSF1_481	UNK	Breeding Material	NPZ	Population	4x
PC_08_4801	UNK	Breeding Material	NPZ	Population	4x
PC_09_4803	UNK	Breeding Material	NPZ	Population	4x
PC_09_4901	UNK	Breeding Material	NPZ	Population	4x
PC_09_4805	UNK	Breeding Material	NPZ	Population	4x
PC_09_4902	UNK	Breeding Material	NPZ	Population	4x
PC_09_4903	UNK	Breeding Material	NPZ	Population	4x
PC_09_4808	Northern	Breeding Material	NPZ	Population	4x
PC_09_4904	Western	Breeding Material	NPZ	Population	4x
PC_09_4906	Western	Breeding Material	NPZ	Population	4x
Syn_06_4501	Western	Breeding Material	NPZ	Population	4x
Syn_06_4503	UNK	Breeding Material	NPZ	Population	4x
Syn_06_4505	UNK	Breeding Material	NPZ	Population	4x
Syn_06_4702	UNK	Breeding Material	NPZ	Population	4x
Syn_08_4001	UNK	Breeding Material	NPZ	Population	4x
Syn_08_4103	UNK	Breeding Material	NPZ	Population	4x
Syn_08_4408	Northern	Breeding Material	NPZ	Population	4x
Syn_08_4507	Northern	Breeding Material	NPZ	Population	4x
Syn_08_4605	Western	Breeding Material	NPZ	Population	4x

Syn_08_4606	Western	Breeding Material	NPZ	Population	4x
Syn_08_4607	UNK	Breeding Material	NPZ	Population	4x
NPZ_FNR_1_(2010)	UNK	Breeding Material	NPZ	Population	4x
2060005	Western	Breeding Material	DSV	Population	2x
2060030	Western	Breeding Material	DSV	Population	2x
2060166	Western	Breeding Material	DSV	Population	2x
2060452	Western	Breeding Material	DSV	Population	2x
2060912	Western	Breeding Material	DSV	Population	2x
2060956	Western	Breeding Material	DSV	Population	2x
2062148	Western	Breeding Material	DSV	Population	2x
2062153	Western	Breeding Material	DSV	Population	2x
2090502	Western	Breeding Material	DSV	Population	2x
2090505	Western	Breeding Material	DSV	Population	2x
2090516	Western	Breeding Material	DSV	Population	2x
2060033	Western	Breeding Material	DSV	Population	2x
2060049	Western	Breeding Material	DSV	Population	2x
2060072	Western	Breeding Material	DSV	Population	2x
2060118	Western	Breeding Material	DSV	Population	2x
2060123	Western	Breeding Material	DSV	Population	2x
2060356	Western	Breeding Material	DSV	Population	2x
2060480	Western	Breeding Material	DSV	Population	2x
2060497	Western	Breeding Material	DSV	Population	2x
2060927	Western	Breeding Material	DSV	Population	2x
2090503	Western	Breeding Material	DSV	Population	2x
2090504	Western	Breeding Material	DSV	Population	2x
2083007	Western	Breeding Material	DSV	Population	2x
2083010	Western	Breeding Material	DSV	Population	2x
2060286	Western	Breeding Material	DSV	Population	2x
2060328	Western	Breeding Material	DSV	Population	2x
2060756	Western	Breeding Material	DSV	Population	2x
2060900	Western	Breeding Material	DSV	Population	2x
2060903	Western	Breeding Material	DSV	Population	2x
2062031	Western	Breeding Material	DSV	Population	2x
2062060	Western	Breeding Material	DSV	Population	2x
2062166	Western	Breeding Material	DSV	Population	2x
91623	Western	Breeding Material	DSV	Population	2x
2030872	Western	Breeding Material	DSV	Population	2x
2030323	Western	Breeding Material	DSV	Population	2x
2030350	Western	Breeding Material	DSV	Population	2x
2030367	Western	Breeding Material	DSV	Population	2x
2020548	Western	Breeding Material	DSV	Population	2x
2030830	Western	Breeding Material	DSV	Population	2x
2030926	Western	Breeding Material	DSV	Population	2x
2030337	Western	Breeding Material	DSV	Population	2x
2030117	Western	Breeding Material	DSV	Population	2x
2020795	Western	Breeding Material	DSV	Population	2x
106232	Western	Breeding Material	DSV	Population	2x
2040121	Western	Breeding Material	DSV	Population	2x
2030377	Western	Breeding Material	DSV	Population	2x

2040371	Western	Breeding Material	DSV	Population	2x
S4	Western	variety	SZS	Population	2x
S11	Western	Breeding Material	SZS	Population	2x
S12	Western	Breeding Material	SZS	Population	2x
S14	Western	Breeding Material	SZS	Population	2x
S15	Western	Breeding Material	SZS	Population	2x
S21	Western	Breeding Material	SZS	Population	2x
S22	Western	Breeding Material	SZS	Population	2x
S23	Western	Breeding Material	SZS	Population	2x
S26	Western	Breeding Material	SZS	Population	2x
S35	Western	Breeding Material	SZS	Population	2x
S40	Western	Breeding Material	SZS	Population	2x
40544	Western	Breeding Material	SZS	Population	2x
40575	Western	Breeding Material	SZS	Population	2x
40603	Western	Breeding Material	SZS	Population	2x
40634	Western	Breeding Material	SZS	Population	2x
40664	Western	Breeding Material	SZS	Population	2x
40695	Western	Breeding Material	SZS	Population	2x
GR3109	Northern	variety	IPK_Genebank	Population	2x
GR5041	Western	ecotype	IPK_Genebank	Population	2x
GR8422	Western	ecotype	IPK_Genebank	Population	2x
GR8428	Northern	Breeding Material	IPK_Genebank	Population	2x
GR3091	Northern	variety	IPK_Genebank	Population	2x
GR3525	Western	ecotype	IPK_Genebank	Population	2x
GR7867	Western	variety	IPK_Genebank	Population	2x
GR8419	Western	variety	IPK_Genebank	Population	2x
GR3107	Northern	variety	IPK_Genebank	Population	2x
GR3231	Northern	variety	IPK_Genebank	Population	2x
GR3368	Eastern	ecotype	IPK_Genebank	Population	2x
GR3511	Eastern	ecotype	IPK_Genebank	Population	2x
GR2704	Northern	variety	IPK_Genebank	Population	2x
GR2725	Northern	variety	IPK_Genebank	Population	2x
GR2910	Northern	Breeding Material	IPK_Genebank	Population	2x
GR3084	Western	variety	IPK_Genebank	Population	2x
GR3236	Western	variety	IPK_Genebank	Population	2x
GR3243	Western	variety	IPK_Genebank	Population	2x
GR3352	Western	ecotype	IPK_Genebank	Population	2x
GR2859	Western	ecotype	IPK_Genebank	Population	2x
GR2915	Northern	landrace	IPK_Genebank	Population	2x
GR2929	Eastern	Breeding Material	IPK_Genebank	Population	2x
GR3122	Eastern	landrace	IPK_Genebank	Population	2x
GR3467	Western	ecotype	IPK_Genebank	Population	2x
GR5646	UNK	ecotype	IPK_Genebank	Population	2x
GR6882	Southern	ecotype	IPK_Genebank	Population	4x
GR8420	Western	ecotype	IPK_Genebank	Population	2x
GR3142	Eastern	variety	IPK_Genebank	Population	2x
GR3172	Eastern	variety	IPK_Genebank	Population	4x
GR3373	Eastern	ecotype	IPK_Genebank	Population	2x
GR3550	Eastern	ecotype	IPK_Genebank	Population	2x

GR5015	Northern	variety	IPK_Genebank	Population	2x
GR5100	Western	ecotype	IPK_Genebank	Population	2x
GR5112	Western	ecotype	IPK_Genebank	Population	2x
GR5113	Western	ecotype	IPK_Genebank	Population	2x
GR7398	Southern	ecotype	IPK_Genebank	Population	2x
GR7420	Southern	ecotype	IPK_Genebank	Population	2x
GR7672	Northern	variety	IPK_Genebank	Population	2x
GR7804	Northern	ecotype	IPK_Genebank	Population	2x
GR8340	Northern	variety	IPK_Genebank	Population	2x
GR8502	Northern	variety	IPK_Genebank	Population	2x
GR8605	Western	ecotype	IPK_Genebank	Population	2x
GR8611	Western	ecotype	IPK_Genebank	Population	2x
GR8808	Western	variety	IPK_Genebank	Population	2x
GR8826	Eastern	ecotype	IPK_Genebank	Population	2x
GR9013	Northern	Breeding Material	IPK_Genebank	Population	2x
GR6598	UNK	ecotype	IPK_Genebank	Population	2x
GR9047?	Western	variety	IPK_Genebank	Population	2x
SH_2/74	Western	ecotype	NPZ	Population	2x
SH_3/109	Western	ecotype	NPZ	Population	2x
SH_3/128	Western	ecotype	NPZ	Population	2x
SH_4/155	Western	ecotype	NPZ	Population	2x
SH_5/181	Western	ecotype	NPZ	Population	2x
SH_5/187	Western	ecotype	NPZ	Population	2x
SH_6/222	Western	ecotype	NPZ	Population	2x
SH_7/282	Western	ecotype	NPZ	Population	2x
Barelan	Western	variety	NPZ	Population	4x
107802	Western	Breeding Material	DSV	Population	2x
2020271	Western	Breeding Material	DSV	Population	2x
2020340	Western	Breeding Material	DSV	Population	2x
2020432	Western	Breeding Material	DSV	Population	2x
2020505	Western	Breeding Material	DSV	Population	2x
2020511	Western	Breeding Material	DSV	Population	2x
2020662	Western	Breeding Material	DSV	Population	2x
2030083	Western	Breeding Material	DSV	Population	2x
2030176	Western	Breeding Material	DSV	Population	2x
2030331	Western	Breeding Material	DSV	Population	2x
2030359	Western	Breeding Material	DSV	Population	2x
2030406	Western	Breeding Material	DSV	Population	2x
2030442	Western	Breeding Material	DSV	Population	2x
2030472	Western	Breeding Material	DSV	Population	2x
2030529	Western	Breeding Material	DSV	Population	2x
2030611	Western	Breeding Material	DSV	Population	2x
2030721	Western	Breeding Material	DSV	Population	2x
2030783	Western	Breeding Material	DSV	Population	2x
2030867	Western	Breeding Material	DSV	Population	2x
2040102	Western	Breeding Material	DSV	Population	2x
2040352	Western	Breeding Material	DSV	Population	2x
2040606	Western	Breeding Material	DSV	Population	2x
40725	Western	Breeding Material	SZS	Population	2x

40756	Western	Breeding Material	SZS	Population	2x
40787	Western	Breeding Material	SZS	Population	2x
40817	Western	Breeding Material	SZS	Population	2x
40848	Western	Breeding Material	SZS	Population	2x
2011-12_1	Western	Breeding Material	SZS	Population	2x
2011-13_1	Western	variety	SZS	Population	2x
2011-14_1	Western	Breeding Material	SZS	Population	2x
2011-15_1	Western	Breeding Material	SZS	Population	2x
2011-16_1	Northern	variety	SZS	Population	2x

Acknowledgement