Theor Appl Genet (2007) 115:837–847
DOI 10.1007/s00122-007-0612-3

ORIGINAL PAPER

Linkage disequilibrium in synthetic varieties of perennial ryegrass

J. Auzanneau · C. Huyghe · B. Julier · P. Barre

Received: 8 October 2006 / Accepted: 8 July 2007 / Published online: 16 August 2007

Abstract Synthetic varieties obtained after three to four panmictic generations are variable, not structured and so can be used for association studies. The pattern of linkage disequilibrium (LD) decay determines whether a genome scan or a candidate gene approach can be used for an association study between genotype and phenotype. Our goal was to evaluate the effect of the number of parents used to build the synthetic varieties on the pattern of LD decay. LD was investigated in the gibberelic acid insensitive gene (GAI) region in three synthetic varieties of perennial ryegrass (*Lolium perenne* L.) chosen for their contrasted number of parents in the initial polycrosses. Results were compared with those obtained from a core collection. STS and SSR markers were used to evaluate variation, structuration and LD in each variety. As expected, the varieties variability increased with the number of parents almost up to the core collection variability. No structuration was observed in the varieties. Significant LDs were observed up to 1.6 Mb in a variety originated from six related parents and not above 174 kb in a variety originated from 336 parents. These results suggest that a candidate gene approach can be used when varieties have a large number of parents and a genome scan approach can be envisaged in specific regions when varieties have a low number of parents. Nevertheless, we strongly recommend to estimate the pattern of LD decay in the population and in the genomic region studied before performing an association study.

Communicated by F. van Eeueijk.

J. Auzanneau · C. Huyghe · B. Julier · P. Barre (✉)
INRA, Unité de Génétique et d'Amélioration des Plantes Fourragères, BP6, 86600 Lusignan, France
e-mail: philippe.barre@lusignan.inra.fr

Introduction

The pattern of linkage disequilibrium (LD) decay determines whether a genome scan or a candidate gene approach can be used in an association study between genotype and phenotype (Ardlie et al. 2002; Flint-Garcia et al. 2003; Rafalski 2002). When LD slowly declines as distance increases, a low marker density is sufficient to detect an association between genotype and phenotype. In that case, a whole genome scan study can be considered. This gives a low accuracy on the location of the mutation responsible for the phenotype but does not require any knowledge of the genes involved in the phenotype. On the contrary, when LD rapidly declines, a much greater density of markers is required to identify an association between genotype and phenotype. In that case, a whole genome scan study is time and money consuming. Therefore a candidate gene approach with markers targeted on a few genes only will be preferred. This gives a high accuracy for the location of the mutation responsible for the phenotype but needs some knowledge on the possible genes involved in the phenotype.

Association studies based on LD are performed using natural populations. In comparison to mapping populations used for QTL analyses, these populations have the advantage not to be limited by the possibility to make the desired crosses and by the low number of segregating alleles (Gaut and Long 2003; Gupta et al. 2005). But these populations are often structured which could lead to spurious associations between a marker and a phenotype (Pritchard et al. 2000b). To overcome this problem, different methods have been developed to take into account the population substructure (Mackay and Powell 2007; Price et al. 2006; Pritchard et al. 2000a, b; Yu et al. 2006). Nevertheless, using these methods, it is not possible to identify the genetic

bases of traits differing between the subgroups of the population. For this reason, it is more suitable to use unstructured populations containing all the alleles of interest. This should be the case of synthetic varieties in some outcrossing species. They are obtained after three to four panmictic bulking generations starting from a variable number of parents (between four to several hundreds). In these varieties the LD between two loci should only be due to a physical link and should decrease more rapidly when the number of unrelated parents increases. At present, no LD studies or association analyses have been carried out on populations constituted by plants from a single synthetic variety. The only study using synthetic varieties has been carried out by Skot et al. (2005) in perennial ryegrass but with a population constituted by plants from 26 natural populations and three synthetic varieties. This reconstituted population does not avoid the substructure.

The main goal of the present study was to evaluate the effect of the number of parents at the origin of a synthetic variety, and thus the variability on the pattern of LD decay. The species of interest was the perennial ryegrass (*Lolium perenne* L.) on which we are studying the genetic basis of leaf length shown to be an important factor of intake rate by grazing dairy cows (Barre et al. 2006). Our strategy was to compare the pattern of LD decay between three synthetic varieties of perennial ryegrass contrasted for the number of parents in the initial polycrosses. We chose two extreme varieties: one with a very high number of parents (Herbie) and one with a few related parents (Aberavon), and one variety more typical in perennial ryegrass (Brest). Furthermore, a core collection was added in order to compare the observed variability of synthetic varieties to the overall variability of *Lolium perenne* L. and to compare the observed LD decay of synthetic varieties to the one found in a core collection. The study was conducted in the region of the gibberelic acid insensitive gene (GAI) whose homologous genes in rice, wheat, barley and corn act on organ growth and flowering time (Ogawa et al. 2000; Peng et al. 1997, 1999). GAI was mapped on linkage group 3 in rice (Chardon et al. 2004) (http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=13699786&itemID=65&view=gbwithparts) and on linkage group 4 in perennial ryegrass (Yamada and Forster 2005).

The results of our study should be helpful for designing association studies with synthetic varieties.

Materials and methods

Plant materials

Three synthetics varieties, Aberavon (Abe), Brest (Bre) and Herbie (Her), were chosen for their contrasted number of parents in the initial polycrosses. The numbers of parents used to build the varieties are six related plants for Abe (Peter Wilkins, personal communication), ten unrelated plants for Bre (Pierre Baudoin, personal communication) and 336 plants for Her (Thieu Pustjens, personal communication). We studied the third multiplication generation of each variety to avoid possible contamination during the last multiplication generation prior to commercialisation. Seeds were provided by Pierre Baudoin from Jouffray Drillaud, France, except for the seeds of Aberavon provided by Peter Wilkins from IGER, UK. For each variety 47 plants were studied. Moreover, we built a population with 47 ecotypes (one plant per ecotype) chosen in a Eurasian core collection in order to maximise the number of geographical origins (François Balfourier, personal communication; Table 1).

DNA extraction

DNA of Abe, Bre and Her was extracted from 50 mg of young leaf using Cheung procedure (Cheung et al. 1993). This method performed in a 96-well format is fast but the quantity of DNA extracted was just sufficient for this study. DNA of the core collection (Cc) was extracted from 1 g of young leaf using a cetyltrimethyl ammonium bromide (CTAB) protocol (Saghai-Maroof et al. 1984; Weising et al. 1991). This extraction is slow but it allowed to obtain enough DNA for other studies. The DNA quantity and quality of each sample was assessed by running 10 µl of 20th and 40th DNA diluted solution on a 1% agarose gel.

SSR markers

We chose three perennial ryegrass SSRs on the same linkage group (LG) as GAI (Fig. 1) and three on different LGs (Table 2). PCR reactions were set up in 25 µl volumes in 96-well PCR plates. Each PCR reaction contained 40 ng of template DNA, 0.2 µM of forward primer with the M13 queue (Boutin-Ganache et al. 2001), 0.1 µM of reverse primer, 0.1 µM of M13 primer labelled with a fluorochrome (700 or 800 nm), 1.5 mM for LPSSR023, LPSSR082, B1C9 and PRG or 2 mM for LP165 and B1B6 of $MgCl_2$, 0.2 mM dNTP, 0.5 U Platinium Taq DNA polymerase (Invitrogen) and 1× PCR buffer (Invitrogen). Amplifications were performed using a thermal cycler (PTC100, MJResearch). For PRG and LP165 markers, the PCR program was: 4 min at 94°C, followed by 31 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 2 min, and a final extension of 6 min at 72°C. For the other markers, a touchdown program was performed: 3 min at 94°C, followed by 10–13 cycles of (94°C for 30 s, annealing temperature minus 1°C per cycle for 30 s from 65 to 55°C for B1B6 and B1C9, from 67 to 55°C for LpSSR023, and from 62 to 50°C for LpSSR082 and 72°C for 1 min), followed by

Table 1 The origins of *Lolium perenne* L. accessions used in this study

Accession	Species	Country	Latitude	Longitude
280001	*L. perenne*	Cyprus	33.4°N	34.02°E
310008	*L. perenne*	Iran	38.15°N	48.18°E
130003	*L. perenne*	Greece	38°N	20.8°E
020118	*L. perenne*	Spain (Sierra Nevada)	38°N	3.45°O
270016	*L. perenne*	Turkey	39.46°N	30.32°E
040130	*L. perenne*	Italy (Sardinia)	40.81°N	9.32°E
270018	*L. perenne*	Turkey	41.01°N	28.58°E
110013	*L. perenne*	Yugoslavia (Bosnia)	42.2°N	20.73°E
911410	*L. perenne*	France (Corsica)	42.5°N	8.75°E
020089	*L. perenne*	Spain	43.4°N	6°O
020100	*L. perenne*	Spain	43.43°N	2.8°O
200002	*L. perenne*	Switzerland	46.33°N	6.95°E
080024	*L. perenne*	Rumania	46.46°N	22.48°E
160001	*L. perenne*	Hungary	46°N	19°E
140005	*L. perenne*	Austria	47.15°N	11.34°E
911262	*L. perenne*	France	47.5°N	4°E
200003	*L. perenne*	Switzerland	47.5°N	9.85°E
910313	*L. perenne*	France	47.67°N	3.48°O
120027	*L. perenne*	Germany	47.7°N	10.7°E
080017	*L. perenne*	Rumania	47.74°N	24.55°E
160005	*L. perenne*	Hungary	47.75°N	17.6°E
160003	*L. perenne*	Hungary	47°N	19°E
140001	*L. perenne*	Austria	48.23°N	14.02°E
180003	*L. perenne*	Belgium	51.08°N	3.45°E
190011	*L. perenne*	The Netherlands	51.43°N	4.05°E
060049	*L. perenne*	Great Britain	51.46°N	1.15°O
120001	*L. perenne*	Germany	51.5°N	9.75°E
060030	*L. perenne*	Great Britain	52.37°N	0.31°O
190009	*L. perenne*	The Netherlands	52.65°N	6.28°E
220005	*L. perenne*	Denmark	54.73°N	11.82°E
100003	*L. perenne*	Russia	54°N	28°E
220006	*L. perenne*	Denmark	55°N	9.78°E
100002	*L. perenne*	Russia	56°N	23°E
230001	*L. perenne*	Sweden	57.58°N	18.55°E
230002	*L. perenne*	Sweden	57.58°N	18.68°E
090001	*L. perenne*	Norway	59°N	7°E
090004	*L. perenne*	Norway	60°N	9°E
210012	*L. perenne*	Bulgaria	Not available	
050087	*L. perenne*	Irland	Not available	
050117	*L. perenne*	Irland	Not available	
070012	*L. perenne*	Poland	Not available	
150002	*L. perenne*	Czech Republic	Not available	
210014	*L. perenne*	Bulgaria	Not available	
300005	*L. perenne*	Liban	Not available	
300009	*L. perenne*	Liban	Not available	
110008	*L. perenne*	Yugoslavia (Macedonia)	Not available	
070022	*L. perenne*	Poland	Not available	

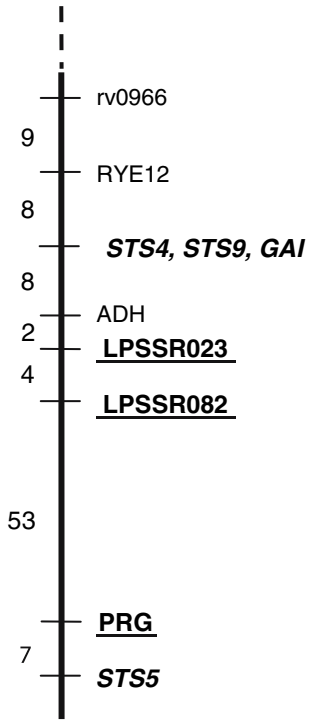

Fig. 1 Partial genetic map of *Lolium perenne* L. linkage group 4 from VrnA and 8490 mapping populations (Jensen et al. 2005a, b). Map distances are in cM. The markers used are in *bold*. SSRs are *underlined* and STSs are *italicized*

Table 2 SSRs positions and references

SSR names	Linkage groups	References
B1B6	1	Lauvergeat et al. (2005)
B1C9	3	Lauvergeat et al. (2005)
LPSSR082	4	Jensen et al. (2005b)
LPSSR023	4	Jensen et al. (2005b)
PRG	4	Kubik et al. (2001)
LP165	7	Kubik et al. (2001)

31 cycles of (94°C for 30 s, 55°C for B1B6, B1C9 and LpSSR23 or 50°C for LpSSR082 for 1 min and 72°C for 1 min), and a final extension of 5 min at 72°C. The DNA fragments were denatured and size-fractioned using polyacrylamide gels run on a *Li-Cor IR²* DNA analyser sequencer (Sciencetech). Fragment size data were recorded using Saga Generation 2 software and manually checked.

STS markers

A set of four perennial ryegrass STS primer pairs was developed to identify sequence polymorphism (Table 3). One STS (GAI) was developed in GAI gene and the other three (STS4, STS5 and STS9) on three BACs around the BAC including GAI (AC091775, AC082645 and AC118133). Genes where STSs were developed, were chosen using a software program (Barre et al. 2005) which indicates the genes with homologous sequences available in at least two grass species. Each primer pair was designed on an alignment between rice and at least one other grass species and chosen to include an intron excepted for GAI which has no intron on rice. PCR reactions were set up in 50 µl volumes in 96-well PCR plates. Each PCR reaction was performed with 40 ng of template DNA, 0.4 µM of primer, 2 mM $MgSO_4$, 0.2 mM dNTP, 1 U Platinium Taq DNA polymerase High Fidelity (Invitrogen) and 1× PCR buffer (Invitrogen). The amplifications were performed using a PTC100 thermal cycler with the following program: 10 min at 94°C, followed by 35 cycles of 94°C for 1 min, Ta (Table 3) for 1 min and 68°C for 2 min, a final extension of 10 min at 68°C.

PCR products quality was assessed by running 5 µl of each sample on a 1–2% agarose gel. DNA of each sample was purified using QIAquick Multiwell PCR Purification Kit (QIAGEN).

For all STS markers, each purified PCR product was sent to Millegen, Toulouse, France for direct sequencing of PCR products with STS forward primer. The direct sequencing of the PCR product allowed us to obtain the genotype of each SNP on the readable sequence (Table 3) but not the phase between SNPs.

For STS4 and STS5, 3 µl of each purified DNA were used for cloning STS sequences with the pGEM-T II kit (Promega). For each marker, one clone per genotype was sent to Millegen for sequencing with T7 and SP6 primers in order to obtain the entire cloned sequence. Clone sequences allowed us to obtain the phase between SNPs for STS4 and STS5 in each individual. These data were used for the calculation of within genes gametic LD.

Gene mapping

For each STS, length polymorphism between the parents of two mapping populations was investigated. STS9 was mapped using the 8490 mapping population (Jensen et al. 2005b). STS4 and STS5 were mapped using the VrnA mapping population (Jensen et al. 2005a). Joinmap 3.0 software was used to build maps. Because the genomic sequence of perennial ryegrass is not available, we had to estimate the distances (in bp) between genetically completely linked STSs in perennial ryegrass (STS4, STS9 and GAI) by the corresponding distances in the rice pseudomolecule (Version 3, 30 December, 2004). We did that although it is known that physical distances could differ greatly between species. The distances in bp between not completely linked markers (SSRs and STSs) belonging to LG4 were estimated using a map size of 772 cM (Jensen et al. 2005b) and a ryegrass genome size of 2.7 Gb (Bennett and Leitch 2005) which led to 1 cM corresponding to

Table 3 STSs description

STS	Source Genbank BAC accession number	Primer sequence	Expected size in rice (bp)	Observed size (bp)	PCR products sequencing Readable size in bp (intron/exon)	Ta (°C)
STS4	Rice AC091775	5′-AAGGTYGATTTGGAGAATAC-3′ 5′-TCCACTGCTTTTAAAAGAACT-3′	1,003	13,224	799 (761/38)	55
STS5	Rice AC082645	5′-ATGTGGGACTACCGCATGG-3′ 5′-TTCATCTGGGTGAGACGCA-3′	1,038	874	782 (423/359)	58
STS9	Rice AC118133	5′-GGGTGGTAYTGTGGACCT-3′ 5′-AAARCCTCCTTTGCATCCRTA-3′	291	300	251 (153/98)	52.7
GAI	Rice AC087797	5′-GACYTGGAGCCSTTCATGCT-3′ 5′-GTACACCTCSGACATGACCT-3′	348	370	232 (0/232)	58.7
Total			2,680	2,866	2,064 (1,337/727)	

3.5 Mb. These estimations are very rough knowing that the genetic size of a map varies between crosses and that the correspondence between physical and genetic distances varies along the genome. Furthermore, an arbitrary distance of 500 Mb was assigned between unlinked markers, i.e., on different linkage groups. In conclusion, the distances between markers used in this study are very rough estimates. It is fine for population comparisons in this study since the same distances were used for all populations. But one should be very careful when comparing the pattern of LD decay with another study.

Data analysis

SSR analysis

GENETIX software (Belkhir et al. 2004) was used to calculate number of alleles, expected heterozygosity on the basis of Nei's unbiased estimate (Nei 1978) and fixation index Fis (Weir and Cockerham 1984). GENEPOP software with default parameters (Raymond and Rousset 1995) was used to test deviations from Hardy Weinberg Equilibrium (HWE). Genetic relationships among individual genotypes for each population were analysed with a model based clustering approach using software package STRUCTURE (Pritchard et al. 2000a). The number of subpopulations (K) was set from one to eight, and each was run ten times. Each run started with 50,000 burn-ins followed by 250,000 iterations employing an admixture model.

STS analysis

The identity of amplified sequences was checked and the location of *Lolium perenne* introns was determined with a Blastn on NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). The sequences of each marker in each population were aligned and manually checked using Staden package 1.6.0 (July 2005; http://www.staden.sourceforge.net/). Particular care was taken with sequences from PCR products in which heterozygous SNPs were manually scored on the chromatograms. The data were formatted using Tritipol (http://www.bioweb.ensam.inra.fr/tritipol/) and a local program Refseqv5 (Fabienne Granier, INRA, Versailles, France) in order to get FASTA formatted sequences of the same length.

From the sequences obtained by direct sequencing of PCR products, we calculated the total number of SNPs per population, the number of rare SNPs (with one allele frequency above 90%) and the number of SNPs present in all populations using Excel software. Insertions–deletions (indels) polymorphism and rare SNPs were excluded (Jung et al. 2004; Tenaillon et al. 2002). Among the remaining SNPs, only SNPs present in all populations were used for further analyses: (1) calculation of expected heterozygosity, fixation index (Fis) and deviations from HWE using the same methods as SSRs and (2) tests of genotypic LD between markers (see below).

LD analysis

Gametic LD was analysed within both STS4 (1,322 bp) and STS5 (874 bp) separately using the clone sequences (one clone per genotype for 47 genotypes) for each population. These two STSs were chosen for gametic LD analysis because of their longest size in comparison to GAI and STS9. Indels polymorphism and rare SNP were excluded. Within STS4, only the 16 SNPs shared by all populations were used, leading to 120 pairs of SNPs. Within STS5, because with clone sequences only one SNP was found in Abe, the LD analyses were performed on Bre, Her and Cc only, using the 14 SNPs shared by these three populations, leading to 91 pairs of SNPs. The r^2 values of gametic LD according to (Hill and Robertson 1968) were determined using DNAsp software (Rozas et al. 2003). A Fisher exact test was also applied to test for LD between pairs of SNP using the same software.

Genotypic LD between markers, including both SNPs and SSR markers, was analysed using the data from direct

sequencing of PCR products for each population. Only the not rare SNPs present in all populations were included. Genotypic LD between markers were tested using Fisher exact test in GENEPOP software (Raymond and Rousset 1995).

Results

Variability increases with the number of parents used to build the varieties

SSRs analysis

A description of the SSRs including number of alleles (A), expected heterozygosity (H) and fixation index Fis is presented in Table 4. All SSRs were polymorphic in the five populations except B1B6 in Abe. Polymorphism varied among *loci* and populations. For A, there was a gradient of polymorphism among varieties with Abe the least polymorphic followed by Bre and then by Her. This gradient was similar to the one obtained from the number of parents used to build the varieties. Abe also had the lowest H, but Bre had a similar H as Her. Cc exhibited the highest number of alleles.

For all SSR markers, the three populations corresponding to the third generation of multiplication (Abe, Bre and Her) were at panmictic equilibrium (total Fis not significant, Table 4) indicating panmictic crosses.

For all varieties no structuration within populations was observed using STRUCTURE software (data not shown).

STSs analysis

All four developed STS gave a clear amplification and presented a single-locus profile. The observed sizes were close to those of rice (Table 3). As expected from rice observations, STS4 and STS9 mapped close to GAI but unexpectedly STS5 was mapped at the end of the linkage group 4 close to the PRG SSR marker (Fig. 1).

Numbers of SNPs per population, for all STS markers, are presented in Table 5. For all the populations as a whole, the level of polymorphism was similar for introns and exons with a slightly lower mean number of base pairs between consecutive SNPs in introns than in exons when all the SNPs were taken into account. All SNPs in exons were synonymous except two. Regarding the average number of base pairs between consecutive SNPs, Abe appeared to be less polymorphic than Bre which was again less polymorphic than Her. For this parameter, Her and the Cc were similar. Fifty eight SNPs (28%) were not present in the Cc but most of them were rare alleles (with one allele frequency above 90%). Among the frequent SNPs, 19 were shared by all populations with 5 in STS4, 2 in STS5, 2 in STS9 and 10 in GAI. These 19 common SNPs were used in further analyses: genotypic linkage disequilibrium between markers and expected heterozygosity.

All the populations had similar expected heterozygosity (about 0.4) except Abe which showed a lower value (Table 6). These differences were mainly due to differences in allele frequencies and not to the number of alleles which rarely exceeded two.

The pattern of LD decay depends on the number of parents used to build the varieties

Gametic LD at short distances

The r^2 values of gametic LD as a function of the distance for each population are presented in Figs. 2 and 3 for STS4 and STS5, respectively. For Abe, LD analyses were not

Table 4 Number of alleles (A), expected heterozygosity (H) and fixation index Fis, with averages across populations, in the studied perennial ryegrass populations (except Fis for the Core collection) at the six SSR *loci*

Populations											
SSRs markers	Aberavon			Brest			Herbie			Core collection	
	A	H	Fis	A	H	Fis	A	H	Fis	A	H
LPSSR082	5	0.60	0.02	14	0.91	−0.03	18	0.90	−0.02	24	0.95
B1B6	1	MD	MD	5	0.36	0.02	3	0.21	−0.09	6	0.52
B1C9	2	0.30	−0.05	4	0.26	−0.02	6	0.32	−0.12	14	0.65
LP165	3	MD[a]	MD[a]	3	0.64	0.05	5	0.56	0.35*	11	0.79
LPSSR023	7	0.73	0.18*	15	0.90	0.13	26	0.92	0.13*	33	0.97
PRG	2	0.51	0.36	4	0.71	0.11	4	0.72	−0.04	10	0.69
Total	20	0.42	0.14	45	0.63	0.05	62	0.60	0.06	98	0.76

MD Missing data

* Significant at $P < 0.05$ after Bonferroni correction, deficit or excess of heterozygous relative to HW expectations

[a] Too many null alleles (28%)

Table 5 Numbers of total SNPs and of not rare SNPs (no allele frequency above 90%) detected in intonic and exonic regions from direct sequencing of PCR products in the studied perennial ryegrass populations

Length (pb)	Total number of SNP			bp per SNP			Number of not rare SNP			bp per not rare SNP		
	Intron	Exon	Total	Intron	Exon	Total	Intron	Exon	Total	Intron	Exon	Total
	1,337	727	2,064	1,337	727	2,064	1,337	727	2,064	1,337	727	2,064
Aberavon	22	21	43	61	35	48	10	18	28	134	40	74
Brest	26	32	58	51	23	36	8	28	36	167	26	57
Herbie	83	38	121	16	19	17	29	33	62	46	22	33
Core collection	95	53	148	14	14	14	36	31	67	37	23	31
Total	138	68	206	10	11	10	41	34	75	33	21	28

bp per SNP represents the mean number of base pairs between two following SNPs

Table 6 Expected heterozygosity per STS calculated on the 19 SNPs from direct sequencing of PCR products shared in the four studied perennial ryegrass populations

	Aberavon	Brest	Herbie	Core collection
STS4	0.23	0.34	0.46	0.42
STS5	0.22	0.45	0.48	0.36
STS9	0.28	0.31	0.40	0.37
GAI	0.37	0.40	0.43	0.39
Global	0.31	0.41	0.46	0.44

A total per population is given

performed within STS5 since only one not rare and common SNP was found. The data showed two main results. Firstly, considering all the populations, the pattern of LD decline was dependent on the distance for STS5 but not for STS4 which had a higher level of LD. Secondly, for both STS4 and STS5, the LD was higher in Bre than in Her and in Cc which have similar pattern of LD decay. For example, in STS4, the percentages of r^2 values equal to 1 were above 80% for Bre and below 50% for Her and Cc. Furthermore, in STS4, Abe showed a similar pattern of LD decay as Bre.

Genotypic LD at long distances

The position of the LG4 SSR and STS loci are presented in Fig. 1. The distances between LG4 markers are given in Table 7. Linkage disequilibrium from genotypic data was tested for all pairs of SNP and SSR markers. Cumulative frequency distributions of obtained P value for all marker pairs sorted by distance are shown in Fig. 4 for every population. For each population, we used the distri-

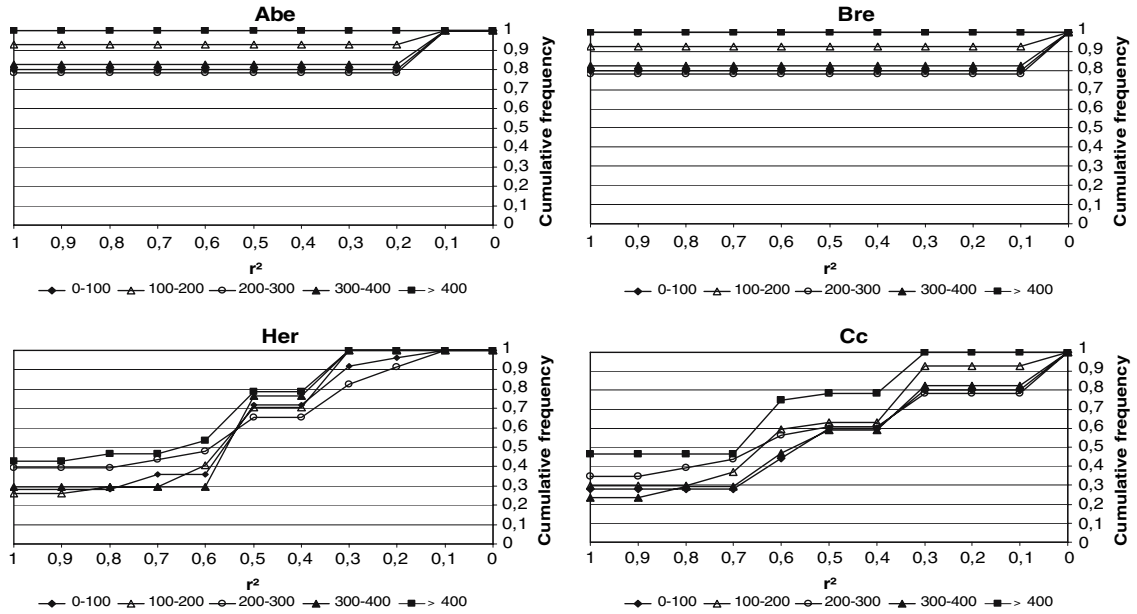

Fig. 2 Cumulative frequency distribution of r^2 values (gametic LD) sorted by between-marker distance for STS4 for each population: Aberavon (*Abe*), Brest (*Bre*), Herbie (*Her*) and the core collection (*Cc*)

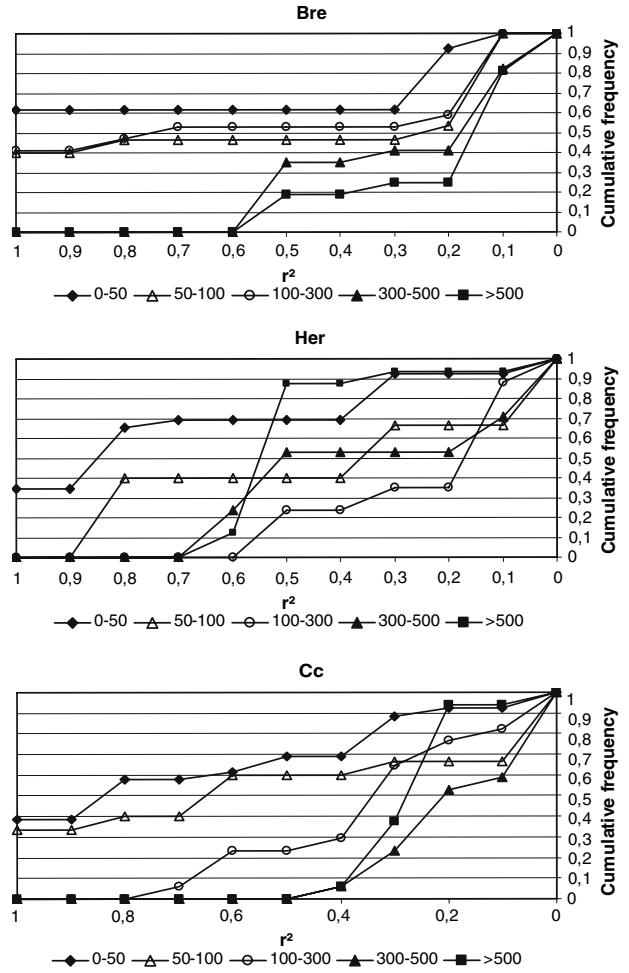

Fig. 3 Cumulative frequency distribution of r^2 values (gametic LD) sorted by between-marker distance for STS5 for each population: Brest (*Bre*), Herbie (*Her*) and the core collection (*Cc*)

bution of the *P* values obtained between unlinked markers in order to determine a *P* value threshold which delimitates the upper 5% of the distribution. The values obtained were 0.02 for Abe, 0.03 for Bre, 0.05 for Her and 0.04 for Cc. For these limits, proportions of significant tests at a short distance of 500 bp were 57, 54, 60 and 56% for Abe, Bre, Her and Cc, respectively. These results did not show any difference between populations but Abe was the only population with no *P* value higher than 0.3. At a distance of 174–175 kb, the proportions of significant tests were 50, 10, 4 and 4% for Abe, Bre, Her and Cc, respectively. At a distance of 1.6–1.8 Mb, the proportions of significant tests were 17, 16, 6 and 3% for Abe, Bre, Her and Cc, respectively. These results show that for long distances, i.e., between 174 kb and 1.8 Mb, Abe had the strongest LD followed by Bre. For distances above 14 Mb (4 cM) no *P* values different from the expectation without linkage were observed for all the populations.

Table 7 Estimated distances between LG4's markers

Marker pairs	Estimated distance
STS4–GAI	174 kb
GAI–STS9	1.6 Mb
STS4–STS9	1.8 Mb
LPSSR082–LPSSR023	14 Mb
STS5–PRG	24.5 Mb
STS4–LPSSR023	35 Mb
GAI–LPSSR023	35 Mb
STS9–LPSSR023	35 Mb
STS4–LPSSR082	49 Mb
GAI–LPSSR082	49 Mb
STS9–LPSSR082	49 Mb
LPSSR082–PRG	185.5 Mb
LPSSR023–PRG	199.5 Mb
STS5–LPSSR082	210 Mb
STS5–LPSSR023	224 Mb
STS4–PRG	234.5 Mb
GAI–PRG	234.5 Mb
STS9–PRG	234.5 Mb
STS4–STS5	259 Mb
GAI–STS5	259 Mb
STS5–STS9	259 Mb

Discussion

The present study is the first evaluation of LD pattern within synthetic varieties. It was conducted using only 47 plants per population. This sample size was low but found to be sufficient in other LD studies (Harmegnies et al. 2006; Liu and Burke 2006). The core collection was built using one individual per accession because the goal was not to study the structuration of perennial ryegrass but to evaluate the overall diversity. This type of population is often used for this purpose (Ghariani et al. 2003; Liu and Burke 2006).

Variability increases with the number of parents used to build the varieties

Variability analyses using SSRs and STS sequences revealed differences between varieties with Abe being the least variable, Her the most variable and Bre in between. In particular, the very low diversity found in STS5 for Abe could be explained by the narrowness of its genetic basis. This gradient of variability among these three varieties was in complete accordance with what we expected from the initial numbers of genotypes used to build the varieties (6 related genotypes for Abe, 10 unrelated genotypes for Bre and 336 genotypes from 4 varieties for Her). This relationship, in perennial ryegrass varieties, between the number of

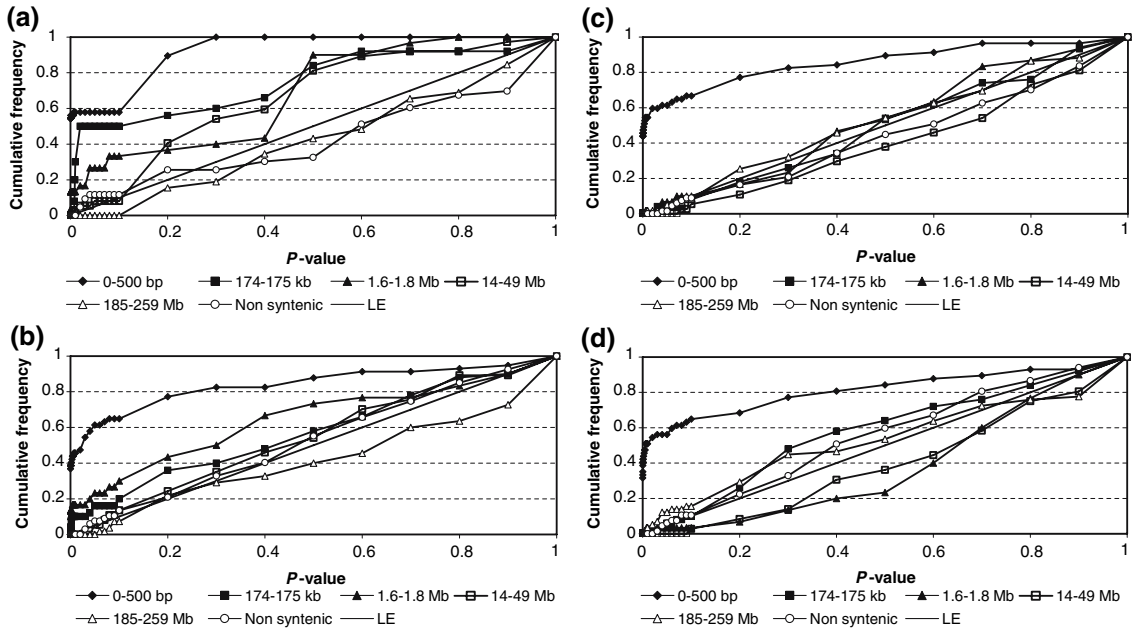

Fig. 4 Cumulative frequency distribution of *P* value of the Fisher exact test for linkage disequilibrium for marker pairs sorted by between-marker distance for Aberavon (**a**) Brest (**b**) Herbie (**c**) and core collection (**d**). *Black diagonal line* represents expected statistical significance (*P* value) assuming linkage equilibrium (LE)

parents and the observed variability has already been observed by Guthridge et al. (2001) using AFLP. Nevertheless, this relationship is not perfect and could be modified since the relatedness of the parents also matters. Variability observed in our study was comparable to the one presented by Kubik et al. (2001) on seven perennial ryegrass varieties, where A ranged from three to nine for LP165 and from two to seven for PRG. Cc was the most polymorphic but the difference with the variety Herbie was narrow and mainly due to the presence of rare alleles.

On all SSR markers, Abe, Bre and Her did not significantly deviate from HWE. This was already observed in seven perennial ryegrass varieties by Kubik et al. (2001). Nevertheless, in both studies, deviations from HWE were observed for several *loci* and varieties and mainly with a bias towards a deficit of heterozygous genotypes. This was also found with isozyme markers in natural populations of perennial ryegrass (Balfourier et al. 1998). One reason could be the presence of undetected null alleles. Missing data in our study did not exceed four per *locus* and per population except for LP165 in Abe were 13 (28%) missing data were observed. But even four missing data per *locus* represents a frequency of null allele of 29% in a population and could increase significantly the level of homozygosity by coding heterozygous plant with a null allele as a homozygous one. Furthermore, the absence of structuration observed for Abe, Bre and Her confirms panmictic crossing during the three generations of multiplication.

Map positions

As expected from rice observations, STS4, STS9 and GAI were completely genetically linked. But unexpectedly STS5 was mapped at the end of the linkage group 4. This result could be due to the amplification of another gene from the same gene family. In the rice genome, a highly homologous (more than 90% at the nucleotide level) gene of the one found in chromosome 3 (OSJNBb0033N16.10, putative alpha-coat protein) was found on chromosome 9. We amplified a single locus in perennial ryegrass since we got only one band and clear sequences even from PCR products. It is more likely that this locus is homologous to the one on rice chromosome 3 than to the one on chromosome 9 since it mapped on LG 4 of perennial ryegrass which corresponds mainly to chromosome 3 of rice and some parts of chromosome 1 and 11 (Alm et al. 2003). It is unlikely that the unexpected map position comes from a mapping problem since in the same mapping population we found STS4 at the expected position close to GAI, and STS5 not linked to STS4. The most likely explanation would be a chromosome rearrangement leading to a rupture of the synteny.

The pattern of LD decay depends on the number of parents used to build the varieties

In Cc, LD was detected with a high proportion of r^2 values higher than 0.1 up to 800 bp and no significant LD was

detected over 174 kb. Although it is not the same chromosomal region, this result is in accordance with a recent study on perennial ryegrass natural populations revealing that LD decays to a r^2 value of 0.1 over 2–3 kb (Skot et al. 2005). The comparison between STS4 and STS5 showed differences in the patterns of LD decay. We did not observe a decrease of LD up to 800 bp in STS4 but a strong decrease of LD up to 300 bp was observed in STS5 which is not surprising in an out-breeding species (Remington et al. 2001). Differences of patterns of LD decay among loci have already been noted in maize (Remington et al. 2001). The level of LD observed in Cc was similar to the one observed in the variety Herbie (of high variability) and was lower than in the varieties Abe and Bre.

An important result of our study is that Abe had a stronger LD than Bre which had a stronger LD than Her. This result suggested that the pattern of LD decay was linked to the genetic variability of the varieties resulting from the number and the relatedness of parents used to build the varieties. This link was expected since the decay of LD with the distance is influenced by the effective population size (Rafalski 2002). The number of parents used to create a variety and their relatedness determine the initial LD which decreases with the number of multiplication generations according to the formula $D_n = (1 - c)^n D_0$ with D a measure of LD, n generation and c distance in cM between the two considered loci (Gallais 2003). Considering this formula, a complete LD of $D_0 = 0.25$ between two loci separated by $c = 10$ cM decreases only to 0.18 in three generations. Even in Abe, LD was significant at a maximal distance of 0.5 cM (1.6 Mb) indicating that the initial LD between the parents of the variety was spread over a shorter distance than what we expected. This could be because we did not reveal all the alleles or because new mutations appeared.

Breeding prospects

The high level of diversity, the absence of structuration and the rapid decrease of LD in synthetic varieties built on many parents (above 10) made promising the use of such varieties in association studies with a candidate gene approach for the polymorphic traits. Similarly, this approach could also be valuable in populations of recurrent selection allowing a marker-assisted selection. The main constraint is the availability of relevant candidate genes.

In synthetic varieties with a low level of variability such as Abe but polymorphic for a trait of interest, the existence of a LD over 0.5 cM (17% of the tests were significant) should permit a genome scan approach. The genotyping of five unlinked SNPs every 0.5 cM should be performed in order not to miss any possible association. The development of such a set of markers is expensive and time consuming and should be facilitated by exploiting synteny with rice (completely sequenced) and with the development of a physical map in perennial ryegrass. For instance, one could easily develop this approach on a specific region of interest.

Another possibility, avoiding the possible lack of polymorphism in the varieties, could be association studies on created synthetic populations starting from a few well-known genotypes with a high level of LD. The interesting regions could be identified on the first generations of multiplication, and then specific regions could be studied on a more advanced generation of multiplication in order to have a more accurate estimation of the location of the genes involved in the studied trait.

Acknowledgments We thank F. Balfourier, P. Wilkins and P. Baudouin for their plant material. We thank M. Barillot, J. F. Bourcier, D. Cadier, P. Cormenier, C. Gibelin, J. Jousse and C. Largeaud for their technical help. We thank I. Cameleyre and S. Flajoulot for their advices in molecular biology and J. Ronfort and D. Manicacci for their help on LD analyses. Finally, we thank the region Poitou-Charentes for financial support.

References

Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z, Rognli OA (2003) A linkage map of meadow fescue (*Festuca pratensis* Huds.) and comparative mapping with other Poaceae species. Theor Appl Genet 108:25–40

Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 3:299–309

Balfourier F, Charmet G, Ravel C (1998) Genetic differentiation within and between natural populations of perennial and annual ryegrass (*Lolium perenne* and *L. rigidum*). Heredity 81:100–110

Barre P, Darrieutort G, Auzanneau J, Julier B (2005) Development and use of a tool for automated alignments of genes in the rice BAC's GenBank card against other species. Fourth symposium of Molecular Breeding of Forage and Turf, a satellite workshop of the XXth International Grassland Congress, Aberystwyth, Wales

Barre P, Emile JC, Betin M, Surault F, Ghesquière M, Hazard L (2006) Morphological characteristics of perennial ryegrass leaves that influence short-term intake in dairy cows. Agron J 98:978–985

Belkhir K, Borsa P, Chikhi L (2004) GENETIX, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier, France

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann Bot 95:45–90

Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. Biotechniques 31:24–28

Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. Genetics 168:2169–2185

Cheung WY, Hubert N, Landry BS (1993) A simple and rapid DNA microextraction method for plant, animal and insect suitable for RAPD and other PCR analyses. PCR Methods Appl 3:69–70

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Gallais A (2003) Quantitative genetics and breeding methods in autopolyploid plants. INRA, Paris, pp 66–68

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506

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

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Guthridge K, Dupal M, Kolliker R, Jones E, Smith K, Forster J (2001) AFLP analysis of genetic diversity within and between populations of perennial ryegrass (*Lolium perenne* L.). Euphytica 122:191–201

Harmegnies N, Farnir F, Davin F, Buys N, Georges M, Coppieters W (2006) Measuring the extent of linkage disequilibrium in commercial pig populations. Anim Genet 37:225–231

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. TAG Theor Appl Genet 38:226–231

Jensen LB, Andersen J, Frei U, Xing Y, Taylor C, Holm P, Lubberstedt T (2005a) 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

Jensen LB, Muylle H, Arens P, Andersen CH, Holm PB, Ghesquière M, Julier B, Lubberstedt T, Nielsen KK, De Riek J, Roldan-Ruiz I, Roulund N, Taylor C, Vosman B, Barre P (2005b) Development and mapping of a public reference set of SSR markers in *Lolium perenne* L. Mol Ecol Notes 5:951–957

Jung M, Ching A, Bhattramakki D, Dolan M, Tingey S, Morgante M, Rafalski A (2004) Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 *locus* in elite maize germplasm. Theor Appl Genet 109:681–689

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

Lauvergeat V, Barre P, Bonnet M, Ghesquière M (2005) Sixty simple sequence repeats (SSR) markers for use in the *Festuca/Lolium* complex of grasses. Mol Ecol Notes 5:401–405

Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics 173:321–330

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12(2):57–63. doi:10.1016/j.tplants.2006.12.001

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Ogawa M, Kusano T, Katsumi M, Sano H (2000) Rice gibberellin-insensitive gene homolog, OsGAI encodes a nuclear-localized protein capable of gene activation at transcriptional level. Gene 245:21–29

Peng JR, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP (1997) The *Arabidopsis* GAI gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev 11:3194–3205

Peng JR, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400:256–261

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909

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

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. Am J Human Genet 67:170–181

Rafalski A (2002) Application of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 54:94–100

Raymond M, Rousset F (1995) GENEPOP (Version1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Saghai-Maroof MA, Soliman K, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. PNAS 81:8014–8018

Skot L, Humphreys MO, Armstead I, Heywood S, Skot KP, Sanderson R, Thomas ID, Chorlton KH, Sackville Hamilton NR (2005) An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.) Mol Breed 15:233–245

Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley J, Gaut BS (2002) Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp mays L.). Genetics 162:1401–1413

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Weising K, Beyermann B, Ramser J, Kahl G (1991) Plant DNA fingerprinting with radioactive and digoxignated oligonucleotide probes complementary to simple repetitive DNA sequences. Electrophoresis 12:159–169

Yamada T, Forster JW (2005) QTL analysis and trait dissection in ryegrass (*Lolium* spp.). Fourth symposium of Molecular Breeding of Forrage and Turf, a satellite workshop of the XXth International Grassland Congress, Aberystwyth, Wales

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208