# Development of ryegrass allele-specific markers for rust resistance in *Lolium perenne*

T. LÜBBERSTEDT, B. SCHEJBEL, U. FREI, B. WOLLENWEBER and Y. XING

## Keywords

ryegrass, GRASP, crown rust, molecular markers, SNPs

## Introduction

In the years to come, European livestock production will face a number of challenges to remain competitive at the international market. At the same time European agricultural policies emphasize the need for improved quality, environmentally sustainable production methods, traceability of the feed and food supply, and diversity. On-farm forage is vital in underpinning European livestock production of the future. To fulfill this strategy, improved production of forage grasses is essential. In contrast to high performing cereals and other „cash crops", forage grasses allow for economic and consumer-friendly feed production under low input conditions (fertilizers, pesticides). Traditionally, forage grass breeding objectives have concentrated on improving sward yield and persistency (JOHNSON and BEYER 1974). However, in the future, breeding programs must pay more attention to specific nutritional needs of grazing ruminants and the impact of livestock farming on the environment (SPANGENBERG et al. 2001).

The genus *Lolium* contains the most important forage grass species in Europe, *L. perenne* and *L. multiflorum*. Together the two species cover 23% of the grassland area (52 million ha) in Europe with the perennial *L. perenne* being the most prevalent grass species. About 45 000 t of ryegrass seed is used each year at a cost of about 160 million Euro with Europe being a net exporter of *Lolium* seeds. Progress in conventional *Lolium* breeding is hampered by the application of simple but inefficient population breeding schemes and by the slow phenotypic evaluation process. Taking into account the high degree of diversity of the breeding materials used by ryegrass breeders, molecular tools for early selection would dramatically accelerate progress in breeding especially of perennials. In the past few years the amount of basic biological information has increased dramatically on the function of hundreds of plant genes, largely through studies on the molecular genetics of the model plant *Arabidopsis thaliana.* Current research in the field of plant genomics has focused on analyses of mutant alleles, created by insertional mutagenesis, and sequencing the entire genomes of selected plant species. This has provided extensive knowledge on gene functions. Several genomics projects act at the level of genes, i.e., gene function is deter-mined by comparison of „wildtype" with „loss of function" alleles. However, plant breeding primarily acts at the level of minor allelic variation among wildtype alleles of a given gene. Thus, allele and SNP identification is of crucial importance to effective and competitive plant breeding but very little information has been generated on the naturally occurring allelic sequence variation in crop plants (THORNSBERRY et al. 2001).

It constitutes a complementary approach to map based cloning and insertional mutagenesis and with the technologies currently available for most of our crop plants it may be the most realistic and readily applicable strategy for improvement of forage grasses to determine allele functions for relevant genes. Development of gene-derived and allele-specific markers associated with relevant characters is the goal in the EU project GRASP (http://www-grasp-euv.dk; LÜBBERSTEDT et al. 2003). GRASP is a EU framework V project involving eight public institutions and one commercial partner. The major target of GRASP is the development of „functional" gene-derived and allele-specific single nucleotide polymorphism markers associated with relevant traits in *Lolium perenne* such as forage quality, nitrogen use efficiency, disease resistance, and abiotic stress tolerance.

Crown rust (*Puccinia coronata* Corda f.sp. *lolii*) is one of the most devastating pathogens of perennial ryegrass (*Lolium perenne* L.), an important forage and turf grass species world-wide (HOPKINS 2000; SIMONS 1970). Resistance to crown rust involves both major and minor resistance genes (KIMBENG 1999). Two independent dominant resistance genes named *Cr1* and *Cr2* have been identified in three $BC_1$ populations (LELLBACH 2001), but not mapped so far. Several QTL for rust resistance have been identified so far (DUMSDAY et al. 2003; MUYLLE et al. 2005; RODERICK et al. 2003; SIM et al. 2005). Markers linked to resistance loci, as identified in QTL studies, can be employed in pyramiding strategies in breeding programs to reduce the risk of resistance breakdown (CRUTE 1998). Ideally these markers would be derived from the respective resistance genes to provide a high level of informativeness (ANDERSEN and LÜBBERSTEDT 2003). Meanwhile more than 40 resistance genes have been isolated in plants (QUINT, 2006), and used to find additional resistance gene analogues (RGAs) in several species (e.g., MADSEN et al. 2003). Resistance genes can be divided into five classes of which the largest class includes NBS-LRR (Nucleotide binding site - leucine rich repeat) genes (DANGL and JONES 2001). The NBS-LRR genes contain regions with highly conserved amino acid sequences such as P-loop, kinase2, GLPL, and MHD (The names refer to the conserved amino acid sequences).

**Autoren:** Thomas LÜBBERSTEDT, Britt SCHEJBEL, Ursula FREI, Bernd WOLLENWEBER, Yongzhong XING, Danish Institute of Agricultural Sciences (DIAS), Department of Genetics and Biotechnology, Research Centre Flakkebjerg, DK-4200 SLAGELSE, thomas.luebberstedt@agrsci.dk

These motifs have been used to isolate NBS-LRR genes based on sequence homology from several plant species including maize, rice, barley, soybean, and lettuce (e.g, Madsen et al. 2003). In Italian ryegrass (*Lolium multiflorum*) 62 RGAs have been cloned by the use of degenerate primers (Ikeda 2005). Futhermore, one RGA has been published and mapped to LG1 in a perennial ryegrass population (MUYLLE et al. 2005). Additional EST-derived RGAs have been mapped, but no sequence information released (FAVILLE et al. 2004). In this study, we sequenced about 1kb regions of 11 expressed disease resistance candidate genes and 3 laccases from 20 genotypes (Lolium Test Set, LTS) employed in the EU project GRASP (LÜBBERSTEDT et al. 2003).

The objectives were to

(1) compare the nucleotide diversity within and between different gene classes,

(2) study effects of natural selection on gene mutations,

(3) determine the amount and structure of LD within these genes, and

(4) to discuss the prospects of candidate-gene based association mapping in ryegrass.

## Materials and methods

A total of 20 genotypes of perennial ryegrass (*Lolium perenne* L.) originating from various sources were included in this study (= Lolium Test Set, LTS). These genotypes represent a wide range of genetic diversity within ryegrass (POSSELT et al. submitted). Three laccase genes and 11 potential disease resistance genes were selected from the annotation of EST sequences generated within the project DAFGRI (http://www.dafgri.dk), which included homologues of nucleotide binding site and leucine rich repeat (NBS-LRR) like, pathogenesis related (PR), Mitogen-activated protein kinase (MAPK), enhanced disease resistance (EDR), and plastid pyruvate kinase A (PKpA) protein coding genes. On the basis of candidate mRNA sequences, 14 pairs of primers were designed to amplify about 1 kb genomic fragments from the 20 genotypes for each of the 14 genes. A touch down PCR program was used beginning with 5 min at 94°C, followed by 12 cycles of 30 s at 94°C, 60 s at annealing temperature 67 °C, 60 s at 72°C with the annealing temperature decreasing by 1°C per cycle, followed by 29 cycles of 30 s at 94°C, 60 s at 55°C, 60 s at 72°C and 10 min at 72°C. All 14 primer pairs ran with the same PCR program on a MJ Research thermocycler (Applied Biosystems, Califonia) in 25 µl reaction mixtures containing 20 ng DNA, 0.2 µM primer, 0.2 mM dNTPs, 0.4 u BD Advantage 2 polymerase, and 2.5 µl 10×BD advantage 2 PCR buffer.

The PCR products were purified from agarose gel using QiaQuick spin columns (Qiagen, Valencia, USA) according to manufacturer instructions. Purified fragments were cloned into vector pCR$^{â}$2.1-TOPO (TOPO TA cloning kit, Invitrogen, Califonia). Five clones per gene for each genotype were randomly picked for plasmid DNA extraction. Purified plasmid DNA was used for allele sequencing on the MegaBACE1000 (Amersham Bioscience, Califonia). Sequence chromatogram files from the same genotype were assembled into contigs by using SEQMAN (DNA star, Madison, WI), and consensus sequences were edited manually to resolve discrepancies. Consensus sequences for all the 20 genotypes were aligned by using CLUSTAL alignment. Polymorphisms which appeared only in one genotype were rechecked in chromatogram files.

When calculating the number of haplotypes, all polypmorphic sites including Indels and segregating sites with 2 and more variants were taken into consideration. Direct comparison of mRNA sequence and its corresponding genomic DNA sequences was used to determine exon and intron regions. Alignment data for each candidate gene were used for nucleotide diversity and linkage disequilibrium (LD) analysis. DnaSP version 4 (ROZAS et al. 2003) was used for the following analyses. All calculations were based on 40 alleles from the 20 heterozygous diploid genotypes. If one genotype was homozygous in the sequenced region, its allele sequence was presented twice in the alignment in order to determine the allele frequency for the 20 genotypes.

LD was estimated by using standardized disequilibrium coefficients (D') (HEDRICK 1987) and squared allele-frequency correlations ($r^2$) (WEIR 1996) for pairs of SNP loci. Sites with alignment gaps or polymorphic sites segregating for three or four nucleotides were completely excluded from the analysis. Fisher's exact test (1935) was used to determine the statistical significance of LD. Decay of LD with distance in base pairs (bp) between sites within the same gene was evaluated by nonlinear regression in Statistica (HILL and LEWICKI 2006).

## Results

About 1kb long sequences have been obtained from all 14 genes and 20 heterozygous LTS genotypes investigated. The 14 genes will be subdivided in the following into NBS-LRR resistance candidate genes (NBS genes: six different NBS genes), Non-NBS-LRR resistance candidate genes (NNL genes: five different NNL genes), and Laccases (LAC genes).

The average number of haplotype-alleles within the collection of LTS genotypes was 20.4, 11.4, and 14.0 for NBS, NNL, and LAC genes, respectively. For the SSRs, the everage number of different alleles was 10.7. The number of haplotypes varied between 12-27, 9-15, and 7-21 for NBS, NNL, and LAC genes, respectively. The majority of alleles was present at low frequency (<5%): 16.6, 8.0, and 10.6 for NBS, NNL, and LAC genes, respectively. In contrast, only few haplotype alleles were detected at frequencies above 20% with average absolute frequencies of 0.6, 1.2, and 1.4 for NBS, NNL, and LAC genes, respectively.

The level of homozygosity found in the 20 LTS genotypes as determined by the haplotype and SSR alleles was much higher as expected under an assumption of random mating. Homozygosities ranged from 20-60% (expected: 4-19%), 35-75% (expected: 17-33%), 50-69% (expected: 10-32%), and 10-93% (expected: 6-44%), for NBS, NNL, LAC, and SSR alleles, respectively.

Generally, Indel and SNP polymorphisms were more frequent in non-coding as compared to coding regions, and synonymous SNP mutations were more fre-

quent than non-synonymous SNP mutations in coding regions. However, for two NBS genes higher SNP frequencies were found in non-coding as compared to non-coding regions. Moreover, in most NBS genes more non-synonymous than synonymous SNP polymorphisms were found. The SNP density was highly variable between genes and ranged from 16 SNPs in 1030 bp to up to 277 SNPs in 1036 bp within the LTS.

The intragenic LD between pairs of SNP polymorphisms as expressed by $r^2$ ranged from 0.07-0.19, 021-0.54, and 0.17-0.63 for individual NBS, NNL, and LAC genes, respectively. Percentages for significant LD between SNP pairs ranged from 8-33%, 31-58%, and 22-79%. When plotting $r^2$ values against the distance between SNP pairs, $r^2$ values fell below 0.2 within distances of 15-220 bp, 300-1600 bp, and 250-195000 bp for individual NBS, NNL, and LAC genes, respectively.

The minimum number of SNPs required to discrimate all haplotypes within the LTS was below 12 for each of the 14 genes investigated. For the majority of these genes, as few as 3-6 SNP polymorphisms were sufficient for complete differen-tiation of all haplotypes found within the collection of LTS genotypes.

## Conclusions

Substantial sequence variation for three classes of expressed resistance candidate genes was found in a collection of 20 unrelated diploid perennial ryegrass genotypes, employed in the EU project GRASP. For most genes, both high SNP densities and low intragenic LD were found. The on average highest SNP densities and and lowest LD were found for NBS-LRR resistance gene candidates. This is in agreement with the biological role of these genes, which are key players in pathogen recognition. Often multiple alleles at respective resistance loci are involved in gene-by-gene interactions with pathogen genes.

For two laccases, LD was extending over larger regions, supporting earlier findings from other species that the degree of LD is strongly genome region-specific. However, the otherwise generally low LD substantially decaying within genes is in agreement with good prospects for candidate-gene based association studies in ryegrass, which should allow detection of short sequences or even sequence motifs associated with the traits of interest, comparable to the study of THORNSBERRY et al. (2001). In contrast, genome-wide association studies based on natural populations would require a very high density of markers with at least a few markers/1 kb.

Alternatively, artificial populations with high or moderate levels of LD can be employed in association studies, such as the synthetic populations studied within GRASP (LÜBBERSTEDT et al. 2003). For tracing and discrimination of allele sequences in this kind of experiments only 3-6 SNP markers per gene are sufficient. This information can be implemented into SNP marker assays allowing a high degree of multiplexing.

## Acknowledgments

This work was supported by the EU framework V project GRASP http://www.grasp-euv.dk/.

## References

ANDERSEN, J. R. and T. LÜBBERSTEDT, 2003: Functional markers in plants. Trends in Plant Science 8:554-560.

CRUTE, I.R., 1998: The elucidation and exploitation of gene-for-gene recognition. Plant Pathology 47:107-113.

DANGL, J.L. and J.D.G. JONES, 2001: Plant pathogens and integrated defence responses to infection. Nature 411:826-833.

DUMSDAY, J.L., K.F. SMITH, J.W. FORSTER and E.S. JONES, 2003: SSR-based genetic linkage analysis of resistance to crown rust (*Puccinia coronata* f.sp. lolii) in perennial ryegrass (*Lolium perenne*). Plant Pathology 52:628-637.

FAVILLE, M.J., A.C. VECCHIES, M. SCHREIBER, M.C. DRAYTON, L.J. HUGHES, E.S. JONES, K.M. GUTHRIDGE et al., 2004: Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). Theor. Appl. Genet.

FISHER, R.A., 1935: The logic of inductive inference. J.R. Stat.Soc. Ser. A. 98:39-54.

HOPKINS, A., 2000: GRASS Its production and utilization. Oxford: Blackwell Science Ltd.

HEDRICK, P.W., 1987: Gametic disequilibrium measures: proceed with caution. Genetics 117:331-341.

HILL, T. and P. LEWICKI, 2006: STATISTICS Methods and Applications. StatSoft, Tulsa, OK.

IKEDA, S., 2005: Isolation of disease resistance gene analogs from Italian Ryegrass (*Lolium multiflorum* Lam.). Grassland Science 51:63-70.

JOHNSON, I.J. and E.H. BEYER, 1974: Forage Crop Breeding. In: Heath M.E., MMetcalfe D.S., Barnes R.F. (eds): Forages. 3rd ed. Iowa State University Press, Ames, Iowa:114-125.

KIMBENG, C.A., 1999: Genetic basis of crown rust resistance in perennial ryegrass, breeding strategies, and genetic variation among pathogen populations: a review. Australian Journal of Experimental Agriculture 39, 3:361-378.

LELLBACH, H., 2001: Stand der Züchtungsforschung zur Kronen- und Schwartzrostresistenz bei Weidelgräsern. Paper presented at 52. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs, 2001, at BAL Gumpenstein.

LÜBBERSTEDT, T., B.S. ANDREASEN, P.B. HOLM, et al., 2003: Development of ryegrass allele-specific (GRASP) markers for sustainable grassland improvement - a new EU Framework V project. Czech Journal of Genetics and Plant Breeding 39:125-128.

MADSEN, L.H., N.C. COLLINS, M. RAKWALSKA, G. BACKES, N. SANDAL, L. KRUSELL, J. JENSEN et al., 2003: Barley disease resistance gene analogs of the NBS-LRR class: identification and mapping. Molecular Genetics and Genomics 269:150-161.

MUYLLE, H., J. BAERT, E. VAN BOCKSTAELE, J. PERTIJS and I. ROLDAN-RUIZ, 2005: Four QTLs determine crown rust (*Puccinia coronata* f. sp. *lolii*) resistance in a perennial ryegrass (*Lolium perenne*) population. Heredity 95:348-357.

QUINT, M. and T. LÜBBERSTEDT, 2006: Resistance Gene Analogs - rapid identification of R gene candidates. In: New Techniques on Characterization of Plant viruses, G.P. Rao (ed.) (in press).

RODERICK, H.W., M.O. HUMPHREYS, L. TURNER, I. ARMSTEAD and D. THOROGOOD, 2003: Isolate specific quantitative trait loci for resistance to crown rust in perennial ryegrass. Paper presented at International Symposium on Grass Breeding 2002, at Braunschweig, Germany.

ROZAS J, J.C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS 2003: DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

SIM, S.C., T. CHANG, J. CURLEY, K. DIESBURG, L. NELSON and G. JUNG, 2004: Genetic Dissection of Genes Controlling Resistance To Crown Rust (*Puccinia coronata* f.sp. *lolii*) In Ryegrass. PAGXIII.

SIMONS, M.D., 1970: Crown Rust of Oats and Grasses. Massachusetts, USA. The American Phytopathological Society monograph: no. 5.

SPANGENBERG G., R. KALLA, A. LIDGETT, T. SAWBRIDGE, E.K. ONG and U. JOHN 2001: Breeding forage plants in the genome era. In: Spangenberg G. (ed.): Molecular Breeding of Forage Crops. Kluwer Academic Publ., Netherlands: 1-39.

THORNSBERRY, J.M., M.M. GOODMAN, J. DOEBLEY, S. KRESOVICH, D. NIELSEN and E.S. BUCKLER, 2001: Dwarf8 polymorphisms associate with variation in flowering time. Nature Genet 28:286-289.

WEIR, B.S., 1996: Genetic Data Analysis II. Sinauer, Sunderland, MA, USA.