Plant Physiology Preview. Published on November 26, 2012, as DOI:10.1104/pp.112.207282

Running head: The Perennial Ryegrass GenomeZipper

Dr. Bruno Studer, Institute of Agricultural Sciences, Forage Crop Genetics, ETH Zurich, 8092 Zurich, Switzerland

Tel.: +41 (0)44 632 01 57
Email: bruno.studer@usys.ethz.ch

Journal research area: Genome Analysis

The Perennial Ryegrass GenomeZipper – Targeted Use of Genome Resources for Comparative Grass Genomics

Matthias Pfeifer[1], Mihaela Martis[1], Torben Asp[2], Klaus F. X. Mayer[1], Thomas Lübberstedt[3], Stephen Byrne[2], Ursula Frei[3], and Bruno Studer[4§]

[1]Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany

[2]Department of Molecular Biology and Genetics, Faculty of Science and Technology, Research Centre Flakkebjerg, Aarhus University, Forsøgsvej 1, 4200 Slagelse, Denmark

[3]Department of Agronomy, Iowa State University, 1204 Agronomy Hall, 50011 Ames, IA, USA

[4]Institute of Agricultural Sciences, Forage Crop Genetics, ETH Zurich, 8092 Zurich, Switzerland

[§]Corresponding author

Financial Source

This work has been funded by The Danish Council for Independent Research, Technology and Production Sciences (project 274-08-0300).

Abstract

Whole-genome sequences established for model and major crop species constitute a key resource for advanced genomic research. For outbreeding forage and turf grass species like ryegrasses (*Lolium* spp.), such resources are yet to be developed.

Here, we present a model of the perennial ryegrass (*Lolium perenne* L.) genome on the basis of conserved synteny to barley (*Hordeum vulgare* L.) and the model grass genome Brachypodium (*Brachypodium distachyon* L.), as well as rice (*Oryza sativa* L.) and sorghum [*Sorghum bicolor* (L.) Moench]. A transcriptome-based genetic linkage map of perennial ryegrass served as a scaffold to establish the chromosomal arrangement of syntenic genes from model grass species. This scaffold revealed a high degree of synteny and macro-collinearity, and was then utilised to anchor a collection of perennial ryegrass genes *in silico* to their predicted genome position. This resulted in the unambiguous assignment of 3,315 out of 8,876 previously unmapped genes to the respective chromosomes. In total, the GenomeZipper incorporates 4,035 conserved grass gene loci which were used for the first genome-wide sequence divergence analysis between perennial ryegrass, barley, Brachypodium, rice, and sorghum.

The perennial ryegrass GenomeZipper is an ordered, information-rich genome scaffold, facilitating map-based cloning and genome assembly in perennial ryegrass and closely related Poaceae species. It also represents a milestone in describing synteny between perennial ryegrass and fully sequenced model grass genomes, thereby increasing our understanding of genome organization and evolution in the most important temperate forage and turf grass species.

The majority of the most important food crops, such as wheat (*Triticum* ssp.), rice (*Oryza sativa* L.), maize (*Zea mays* L.), barley (*Hordeum vulgare* L.), and sorghum [*Sorghum bicolor* (L.) Moench] belong to the grass family. Significant investment in these crop species and the model grass species Brachypodium (*Brachypodium distachyon* L.) has led to the establishment of complete genome sequences for rice (International Rice Genome Sequencing Project, 2005), maize (Schnable et al., 2009), sorghum (Paterson et al., 2009), and Brachypodium (The International Brachypodium Initiative, 2010), constituting major resources for genetic and genomic applications. The complex genome organization and large genome size of both bread wheat [*Triticum aestivum*; 2n=6x=42, 17 Gbp (Bennett and Smith, 1991)] and barley [2n=2x=14, 5.1 Gbp (Doležel et al., 1998)] has delayed the development of a reference genome sequence. However, genome sequencing efforts are ongoing for both crops (http://www.barleygenome.org/; http://www.wheatgenome.org/). A novel approach incorporating cytogenetics, next generation sequencing (NGS) and bioinformatics to systematically exploit synteny with model grasses was recently used in barley to establish a genome-wide putative linear gene index of the barley genome. Notably, 21,766 barley genes were assigned to individual chromosome arms and assembled in a linear order (Mayer et al., 2011).

For forage and turf grass species, however, the development of tools and resources to conduct genomic research has so far lagged behind major cereal crop species and genome sequences have yet to be established. Thus, targeted use of grass genome sequence resources by comparative genomics provides a major opportunity for non-model species to efficiently explore genomic information for genetic and breeding applications. To date, most of these comparative studies between grass genomes have been focused on cereals such as barley, wheat, rice, maize, and sorghum (Salse and Feuillet, 2007). These species revealed significant macro-collinearity between their genomes and led to the construction of a consensus grass map based on 25 rice linkage blocks (Devos and Gale, 2000; Feuillet and Keller, 2002; Devos, 2005). More recent studies proposed that grass genomes have evolved from a five-chromosome ancestral genome into a twelve-chromosome intermediate (Salse et al., 2008), from which grass species evolved through a series of evolutionary shuffling events such as whole-genome or segmental duplications, diploidization, small-scale rearrangements and gene conversions (Salse et al., 2009). It has been estimated that the different grass species diverged from a common ancestor approximately 55 to 65 million years ago

(Mya) (Grass Phylogeny Working Group, 2001). The subfamily Pooideae evolved about 40 to 54 Mya (Gaut, 2002; Sandve et al., 2008; Massa et al., 2011) from a common ancestor shared with bamboo (subfamily Bambusoideae) and rice (subfamily Ehrhartoideae) (Grass Phylogeny Working Group, 2001). Within the subfamily Pooideae, the tribe Brachypodieae is more distantly related to the Poeae than the Triticeae, Aveneae, and Bromeae (Bouchenak-Khelladi et al., 2008). Moreover, the taxonomists place the genera *Lolium*, *Phleum*, and *Festuca* closer to *Avena* than to *Triticum* and *Hordeum* (Grass Phylogeny Working Group, 2001).

Perennial ryegrass (*Lolium perenne* L.) is a diploid (2n=2x=14) member of the subfamily Pooideae, belonging to the tribe Poeae, genus *Lolium*. First attempts to describe synteny between Poeae, Aveneae, and Triticeae species found that the genetic maps of perennial ryegrass and Triticeae cereals are conserved in terms of orthology and collinearity (Jones et al., 2002; Sim et al., 2005). A similarly high degree of synteny was found between meadow fescue (*Festuca pratensis* Huds.) and the Triticeae genomes by comparative mapping of 117 loci with known map positions (Alm et al., 2003). The 4AL/5AL translocation that is characteristic for some Triticeae species was absent in both perennial ryegrass and meadow fescue (Alm et al., 2003; Devos, 2005; Sim et al., 2005). Furthermore, complete orthology of meadow fescue linkage group (LG) 4 and rice chromosome 3 (thereafter referred to as Os3) led to the conclusion that meadow fescue has a more ancestral configuration of the genome than any of the Triticeae species (Alm et al., 2003). However, these early comparative mapping studies had two major limitations, namely the marker technology used and the resolution of the comparative maps. Firstly, the restriction fragment length polymorphism (RFLP) markers used, relied on cross-species hybridization of heterologous DNA probes that were preselected for the ability to provide a single, clear signal, thereby limiting the detection of whole or partial genome duplication events. It is also possible that different stringency parameters during hybridization determined, whether a probe detected single or duplicated loci on the map, which evidently lead to discrepancies between studies (Jones et al., 2002; Sim et al., 2005). Problems also arose due to difficulties in differentiating orthologs and paralogs in gene families, since comparative mapping by RFLPs often identified paralogous rather than orthologous sequences, thus leading to an underestimation of collinearity (Salse et al., 2008). Secondly, the resolution of comparative maps was limited by the number of markers and individuals used to construct the genetic linkage map. The

maps based on 109, 120, and 117 RFLP markers used by Jones et al. (2002), Sim et al. (2005), and Alm et al. (2003), respectively, only permitted identification of large-scale conserved chromosomal segments and sites of rearrangement.

In order to overcome these limitations, genetic linkage maps based on sequence-derived markers can be used to anchor the marker position in the genome of sequenced model grass species by means of bioinformatic tools. By avoiding mapping biases introduced by hybridization-based marker technologies such as RFLPs, these maps provide better options for grass comparative genomics. Until now, such genome-wide *in silico* comparative studies in forage and turf grass species have been limited by the availability of mapped marker sequences. However, a transcriptome-based genetic linkage map has recently become available for perennial ryegrass (Studer et al., 2012) with genome size of around 2.6 Gbp [1C = 2,623 Mbp (Kopecký et al., 2010)]. This map contains 838 DNA markers spanning 750 centiMorgan (cM) with an average marker distance of less than 0.9 cM, making it the most saturated genetic linkage map of perennial ryegrass to date. Of the 838 DNA markers mapped, 767 are expressed sequence tag (EST)-derived. This resource provides the resolution required for a detailed analysis of synteny between perennial ryegrass and reference grass genomes such as Brachypodium, rice, and sorghum. The GenomeZipper approach identifies syntenic regions in these reference genomes and arranges the syntenic blocks along a marker scaffold. It further integrates the genome information of sequenced grass species to construct a linear gene order model at the highest possible resolution, and to resolve species-specific local rearrangements. Such synteny models have proven successful to determine gene order and orientation on both single flow-sorted choromosomes (Mayer et al., 2009; Berkman et al., 2011; Vitulo et al., 2011; Wicker et al., 2011; Hernandez et al., 2012) and whole genomes (Mayer et al., 2011). Extending this concept to perennial ryegrass enables *in silico* prediction of the genome location of unmapped genes, and allows gene-based marker development in specific target regions to accelerate fine-mapping and map-based cloning of genes or QTL of interest.

The primary goal of this study was to establish a linear gene order model of the perennial ryegrass genome on the basis of conserved synteny to barley, Brachypodium, rice, and sorghum. In addition, this study aimed at (i) characterizing chromosomal rearrangements of syntenic genes between perennial ryegrass, barley, and sequenced grass species by means of sequence homology analysis, (ii) predicting

in silico the genomic location of genes by using the identified syntenic relationships and (iii) applying the GenomeZipper to integrate barley, rice, sorghum, and Brachypodium genes in a synteny model that facilitate the detailed study of comparative genomics, mechanisms of evolution, speciation, and domestication in forage and turf grasses.

RESULTS

Input data analysis

The genetic linkage map based on single nucleotide polymorphisms (SNPs) of expressed perennial ryegrass genes (Studer et al., 2012) served as a marker scaffold for the GenomeZipper. For this study, a total of 762 gene-derived markers (between 79 and 154 on each LG, 109 on average) have been located on the genetic linkage map. The length of the marker-containing sequences ranged from 193 to 3623 bp with an average of 889 bp and with a mean GC content of 49.6% (Table 1). The total map length was 750 cM, spanning from 63 cM on LG 3 to 151 cM on LG 2 (mean LG length of 107 cM), with an average marker distance below 0.9 cM. Another 8,876 perennial ryegrass EST contigs and singletons of unknown chromosomal origin (thereafter referred to as unigenes) were used for *in silico* mapping. These unigenes represent more than 5 Mbp of nucleotide sequence information of the perennial ryegrass transcriptome (Table 1).

Global synteny between perennial ryegrass and barley

A synteny-based draft of the barley genome containing 21,766 ordered genes (the barley GenomeZipper, Mayer et al., 2011), was used to investigate the syntenic relationship between perennial ryegrass and barley. For each of the seven perennial ryegrass LGs, the EST sequences of mapped DNA markers were compared against these barley genes (BLASTN, ≥85% alignment identity over at least 100 bp). Using a sliding window approach (250 kbp window size, 50 kbp window shift), the number of syntenic barley genes matched by marker sequences from each LG was calculated and illustrated as heatmaps (Figure 1). In total, 301 of 762 (40%) markers matched a barley full length (fl)-cDNA (Table 2). A close syntenic relationship between the perennial ryegrass LGs and the barley chromosomes (thereafter referred to as 1H to 7H) was discovered. Except for LG 4 and LG 5, marker sequences from each LG matched their counterparts on the corresponding barley chromosome, i.e. LG 1 on 1H, LG 2 on 2H, LG 3 on 3H, LG 6 on 6H, and LG 7 on 7H (Figure 1). Moreover, the perennial ryegrass LGs and the corresponding barley chromosomes were mostly collinear, indicating a highly conserved gene order between these two species. A large-scale chromosomal translocation on LG 4 with respect to Triticeae chromosomes 4 and 5 was found. More specifically, 46 marker sequences of LG 4

significantly matched 32 fl-cDNAs located on barley 4H and 12 fl-cDNAs located on 5H. These marker-associated fl-cDNAs were equally distributed covering more than 90% of 4H and spanning from AK364867 (locus 26 in the barley GenomeZipper) to AK253081.1 (locus 2540). On 5H, we identified three non-syntenic loci (AK370577, AK365621, and AK374623) as well as one large syntenic segment located in the distal end of the long arm. This segment ranged from AK250782.1 (locus 2380) to AK250137.1 (locus 2973) and covered almost 20% of the chromosome. In comparison to barley, the gene order of the translocated segment on perennial ryegrass LG 4 was inverted. In accordance with these results, 32 unique gene loci on 5H have been tagged using the LG 5 marker sequences. The matches were distributed over almost 70% of 5H spanning from locus 104 (AK248198.1) to 2295 (AK363788) in the barley GenomeZipper, highly collinear in their gene order.

Bridging grass genomic synteny from barley to perennial ryegrass

The high level of synteny between perennial ryegrass and barley allowed the development of a genome template by aggregating the available marker scaffold with gene and genome structure information of Brachypodium, rice, and sorghum. Although a global comparative map between perennial ryegrass and sequenced grass species was established, its resolution was limited by the number of mapped marker sequences. To overcome this limitation, we made use of the high degree of collinearity between perennial ryegrass and barley and conferred available high-resolution synteny information between barley and Brachypodium, rice, and sorghum as described in Mayer et al. (2011) into the perennial ryegrass genome structure (Supplemental Figure S1). For the barley chromosomes 1H, 2H, 3H, 6H, and 7H, synteny information was directly transferred to perennial ryegrass chromosomes L1, L2, L3, L6, and L7, respectively. For chromosome L4, the syntenic segments of 4H and 5H were combined. Correspondingly, the segment located at the terminal end of 5H was neglected for bridging of syntenic information to L5. As a result, 18 conserved segments in Brachypodium, 24 in rice, and 20 in sorghum were discovered and assigned to the corresponding chromosomes of the perennial ryegrass genome (Figure 2). Over all chromosomes, 3,926 syntenic genes of Brachypodium, 3,255 of rice, and 3,238 of sorghum were identified (Table 2).

In silico chromosome assignment of perennial ryegrass genes

The perennial ryegrass unigenes were first assigned to their putative chromosomal origin by comparing the ESTs to protein sequences of annotated genes of Brachypodium, rice, and sorghum (BLASTX, ≥70% alignment identity covering at least 30 amino acids) which correspond to syntenic genome segments. In total, 4,520 out of 8,876 (51%) unigenes matched a syntenic gene of Brachypodium, rice, and/or sorghum. Of these, 1,205 genes were discarded from further analysis due to contradictory chromosomal assignments via the barley bridge. Finally, 3,315 (73%) of tagged unigenes (between 408 and 558 on each chromosome) were assigned to chromosomes *in silico* (Table 2).

Construction of the perennial ryegrass GenomeZipper

To resolve the linear order of the chromosome-assigned unigenes, synteny information of Brachypodium, rice, and sorghum was integrated using a modified version of the GenomeZipper protocol (Mayer et al., 2011) with the marker scaffold as a basis (Table 2). Using this projection and positioning, the inferred gene order incorporated 4,035 loci, varying between 487 (L6) and 690 (L2) for the individual chromosomes (Table 3, Supplemental Data 1). In total, 2,758 Brachypodium, 2,270 rice, and 2,351 sorghum genes with orthologs present in perennial ryegrass were arranged. Furthermore, 2,438 barley fl-cDNA sequences were uniquely associated to either a marker sequence or a syntenic gene. Using a stringent best bidirectional hit (bbh) criterion to at least one element of the GenomeZipper scaffold, 2,865 of the 3,315 (86%) chromosome-assigned perennial ryegrass unigenes were integrated along the chromosome model. Using a less stringent anchoring criterion (first-best hit), all but two chromosome-assigned unigenes (3,313) were anchored in the Genome Zipper. Out of a total of 4,035 gene loci incorporated in the GenomeZipper, 3,405 (84%) loci linked a perennial ryegrass gene to its ortholog in Brachypodium, rice, and/or sorghum. Of these, 1,059 loci (31%) were supported by one syntenic gene, 718 loci (21%) by two, and 1,628 loci (48%) were supported by genes of all three model species. The number of loci exclusively tagged by Brachypodium genes (506) was considerably higher than the counterparts for rice (294) and sorghum (259) (Figure 3).

Genome-wide K_a/K_s analysis and estimation of divergence time

Sequence divergence analysis between perennial ryegrass, barley, Brachypodium, rice, and sorghum was performed based on the rate of non-synonymous (K_a) and synonymous (K_s) substitutions between orthologous gene pairs. Using the 8,876 unigenes, high-quality protein alignments to 3,301 barley, 3,789 Brachypodium, 3,434 rice, and 3,528 sorghum orthologs were generated based on a stringent bbh BLAST analysis. For each orthologous gene pair, the K_a/K_s ratio indicating purifying ($K_a/K_s < 1$) or positive ($K_a/K_s > 1$) selection, was calculated. Frequency distributions of K_s, K_a and K_a/K_s values are shown in Figure 4A and Supplemental Figure S2A and S2B. Overall, strong purifying selection acted on the majority of genes and mean K_a/K_s values of 0.16, 0.17, 0.16, and 0.15 were measured for barley, Brachypodium, rice, and sorghum, respectively. Furthermore, the average synonymous substitution rates of orthologous gene pairs was used to investigate the evolutionary relationship. Based on the mode K_s rates against barley (0.31), Brachypodium (0.33), rice (0.53), and sorghum (0.59) (Figure 4A), the divergence times of perennial ryegrass from barley and Brachypodium, rice, and sorghum was estimated to 22-30, 23-32, 37-52, and 42-58 Mya, respectively (Figure 4B).

The perennial ryegrass GenomeZipper – a high resolution template for comparative grass genomics

The syntenic relationship of perennial ryegrass to Brachypodium, rice, and sorghum is characterized by a high degree of synteny and macro-collinearity (Figure 2). For Brachypodium, the syntenic blocks defined as coloured bars, revealed a highly similar pattern when comparing Brachypodium to barley, *Aegilops tauschii* (the D genome donor of hexaploid wheat), and wheat (Figure 2A; The International Brachypodium Initiative, 2010). For rice, complete conservation was observed between perennial ryegrass chromosome L6 and rice chromosome Os2 at the current resolution. For L3, mainly represented by Os1, an additional segment of Os4 was identified that is also present on L2 and L4. Chromosome L1 is represented by an insertion of a segment of Os10 between two distinct segments of Os5. Similarly, L7 evolved from a nested insertion of Os8 into Os6, and L2 is composed of two distinct fragments from both Os4 and Os7. In contrast, L4 and L5 showed large-scale chromosomal rearrangements and are represented by Os3, Os4, Os10, Os11, and OS3, Os9, Os12, respectively

(Figure 2B). For sorghum chromosome 5 (Sb5), however, no orthologous gene relationships to perennial ryegrass was found (Figure 2C).

For a more detailed comparative analysis, 2,375 syntenic Brachypodium, 1,957 rice, and 2,008 sorghum genes that were anchored in both the perennial ryegrass and the barley GenomeZipper, were selected for further investigations (Table 4). To visualize collinear blocks, their physical positions (i.e. the anchored loci in the GenomeZipper) were plotted against each other (Figure 5, Supplemental Figure S3). The chromosomal origin of each gene is colour-coded and illustrates the syntenic relationship at higher genetic resolution. While the global collinearity described above was confirmed, distinctive small-scale chromosomal rearrangements were obtained that differentiate both species clearly.

DISCUSSION

The perennial ryegrass GenomeZipper provides a high-resolution scaffold of the perennial ryegrass genome and offers the opportunity for a more detailed analysis of the organization and evolution of forage and turf grass genomes. The integrative GenomeZipper approach has emerged as a key common standard for comparative genome analysis that enables the rapid development of a draft gene-augmented chromosomal template for large and complex grass genomes (Mayer et al., 2009; Berkman et al., 2011; Mayer et al., 2011; Vitulo et al., 2011; Hernandez et al., 2012). This is an important development for perennial ryegrass, since the size and complexity of its genome are major barriers towards developing a reference genome sequence. Such chromosomal templates are also instrumental for genome resequencing and large-scale marker development strategies that will ultimately enable implementation of genome-wide association studies and genomics-based breeding concepts.

From model to crop species using the GenomeZipper

A reference genome sequence holds the key for detailed cross-species comparative genome analyses. An excellent example of this is the rice genome, which has proven to be a valuable resource for comparative studies in many grass species at both the gene and genome level (Sorrells et al., 2003; Stein et al., 2007; Hackauf et al., 2009; Tamura et al., 2009). More recently, the genome sequence of Brachypodium, a member of the Pooideae subfamily, has become available and provides great promise to become a powerful model system to study the genomes of economically more important pooid grasses including wheat, barley, oat (*Avena sativa* L.), rye (*Secale cereale* L.), as well as forage and turf grass species (The International Brachypodium Initiative, 2010). Although Brachypodium is phylogenetically much closer to the tribes Triticeae, Aveneae, and Poeae than rice (Bouchenak-Khelladi et al., 2008; Massa et al., 2011), it still exhibits a different chromosome number, and major genome rearrangements are to be expected in comparison to Triticeae and Poeae genomes. Comparative analysis of both the marker scaffold and the GenomeZipper from perennial ryegrass to the virtual barely genome (Mayer et al., 2011) revealed an extensive conservation of gene order, thus identifying barley as a promising model for genome analysis in Poeae species. This is of particular interest for future comparative

genomics studies due to the ongoing efforts to sequence the barley genome, which is due for completion in the near future (Nils Stein, personal communication). Indeed, barley is – when compared to Brachypodium – phylogenetically closer to ryegrasses and shares the same number of chromosomes (Bouchenak-Khelladi et al., 2008). The structural conservation of the perennial ryegrass and the barley genome is surprisingly high, considering that their split dates back to 22-30 Mya, only shortly after the split with Brachypodium 23-32 Mya. Subsequent to the split from Brachypodium, a core Pooideae ancestor with new seven chromosomes evolved from a twelve-chromosome ancestral state and its genome size greatly expanded (Kellogg and Bennetzen, 2004; Catalán et al., 2012). Both these events occurred before the Triticeae-Poeae split (i.e. the barley-ryegrass split). These genome structure and genome size changes might have been triggered by genome stress or adaptive processes linked to the global climatic changes at the Eocene-Oligocene boundary around 26-34 Mya (Zachos et al., 2001; Liu et al., 2009), as been suggested by Sandve and Fjellheim (2010).

Similar findings of synteny between perennial ryegrass and Triticeae species have earlier been reported for LG 1, LG 3, and LG 5, but not for LG 2, LG 4, LG 6, and LG 7, each containing non-syntenic regions that indicated large-scale chromosomal rearrangements (Jones et al., 2002; Sim et al., 2005). However, it remained difficult to resolve, if the non-syntenic loci found in these previous studies indicated chromosomal rearrangements, or just reflected mapping errors and limitations of RFLP-based comparative mapping.

GenomeZipper-based comparative grass genomics provide novel insights into genome evolution of perennial ryegrass

The perennial ryegrass GenomeZipper provides the opportunity to compare the overall extent of collinearity between perennial ryegrass and Brachypodium, rice, and sorghum. Overall, 48%, 21%, and 31% of the loci anchored in the GenomeZipper matched syntenic genes in one, two, or all three reference genomes, respectively. The number of syntenic loci was similar in rice and sorghum (294 and 259, respectively) but considerably higher in Brachypodium (506). This is reflecting a closer phylogenetic relationship of perennial ryegrass to Brachypodium when compared to rice and sorghum, and consistent with data from barley (Mayer et al., 2011).

Moreover, species-specific chromosome duplications and rearrangements can now be explored to increase our understanding of genome evolution in perennial ryegrass. For

example, the reduction in chromosome number from a predicted ancestral condition with x = 12 occurred independently in Brachypodium and perennial ryegrass because none of the chromosome fusions are shared between the two species (The International Brachypodium Initiative, 2010). In contrast, the nested insertions of rice chromosomes that formed perennial ryegrass chromosomes L1, L2, and L7 are shared between the tribes Triticeae and Poeae, and have already been described as chromosomal rearrangements differentiating rice and the Triticeae (Gale and Devos, 1998; Luo et al., 2009). This is further supported by studies in oat (Deynze et al., 1995), meadow fescue (Alm et al., 2003), and ryegrass (Sim et al., 2005). Therefore, these chromosomal rearrangements appear to characterize the common lineage of the Triticeae, Aveneae, and Poeae tribes. The absence of the inverse translocation of Triticeae chromosomes 4 and 5 relative to perennial ryegrass L4 provides the opportunity to differentiate Poeae from Triticeae genomes (Alm et al., 2003; Devos, 2005; Sim et al., 2005). The present study confirmed the absence of the Triticeae-specific translocation in perennial ryegrass and further resolved its chromosomal break point. A detailed structural characterisation of this region is of particular interest for comparative genomics of physiological processes such as vernalisation, frost tolerance, drought tolerance, and winter survival, as several QTL for abiotic stress tolerance have been found in close proximity to the chromosome breaking point (Alm et al., 2011). The perennial ryegrass GenomeZipper provides the resolution to study co-location of these QTL with orthologous QTL and stress tolerance genes such as dehydrins, CBF transcription factors or vernalisation response genes of other grass species.

The described synteny will also prove highly useful for cytogenetic approaches such as fluorescent *in situ* hybridization (FISH) to physically locate genes and chromosome landmarks (Hasterok et al., 2006). For example, syntenic genes and bacterial artificial chromosomes (BACs) identified using the GenomeZipper can be used for precise and robust molecular karyotyping of ryegrasses or less characterised grass species. This will be helpful to study genome organization and evolution in Poaceae species in general and in forage and turf grasses in particular.

In silico mapping of perennial ryegrass genes

The perennial ryegrass GenomeZipper assigned 3,315 of 8,876 (37%) unigenes to chromosomes, out of which 2,865 (32%) were anchored *in silico* to genome positions.

Assuming that perennial ryegrass contains between 28,000 and 32,000 genes, in line with estimates for other diploid grass genomes (International Rice Genome Sequencing Project, 2005; The International Brachypodium Initiative, 2010; Massa et al., 2011; Mayer et al., 2011), approximately 10% of the perennial ryegrass genes were mapped by the GenomeZipper. This percentage can be significantly increased by combining more comprehensive transcriptome data with improved resolution of the marker scaffold and whole-genome sequencing data, as it was the case in barley (Mayer et al., 2011). Assuming that around 60% to 85% of grass genomes are collinear (Zhang et al., 2012) and a similar percentage of perennial ryegrass genes can be anchored, it might be possible to replace genetic linkage mapping by *in silico* prediction of gene position and order via the GenomeZipper. Physiological processes conserved in closely related grass species such as vernalisation (Asp et al., 2011) or self incompatibility (Shinozuka et al., 2009) might be good targets for comparative *in silico* gene prediction.

The GenomeZipper will also be of benefit for the generation of a complete reference genome sequence in perennial ryegrass. As the development of a fingerprinted BAC-based physical map (Swain et al., 2011) and *de novo* sequencing (Byrne et al., 2011) of the perennial ryegrass genome is underway, the GenomeZipper will prove useful for ordering an orientating genomic sequence. Due to the large and highly repetitive genome, current shotgun sequencing strategies producing short NGS read data yield very fragmented assemblies. However, genic regions of the genome tend to assemble quite well into larger contigs that can be directly incorporated into the perennial ryegrass GenomeZipper. This will be a very efficient approach to position these gene rich stretches of genomic sequence.

The perennial ryegrass GenomeZipper – a useful tool for fine-mapping and map-based cloning

The primary application of the GenomeZipper will be prediction of regional candidate genes for map-based cloning and QTL fine-mapping. While comparative approaches based on a single genome sequence frequently fail due to regional breakdown of synteny, the side-by-side integration of three different reference genomes complemented with the virtual barley genome, provide opportunities to overcome limitations imposed by species-specific local differences. This enabled us to anchor

and order genes even in regions, where single model genomes may contain structural rearrangements, gene loss, or translocations.

Moreover, the perennial ryegrass GenomeZipper provides gene order information from centromeric regions, which are notoriously difficult to address by recombination-based linkage mapping (King et al., 2007). Suppression of recombination in centromeric regions leading to clustering of genes has been described in perennial ryegrass and meadow fescue (King et al., 2002; Studer et al., 2012) and other grass species such as barley (Stein et al., 2007) and Brachypodium (Huo et al., 2011). However, these regions encompass substantial parts of chromosomes and genes (King et al., 2007), limiting the success of map-based cloning. While the GenomeZipper facilitates identification of candidate genes and development of functional markers in these regions, alternative strategies based on linkage disequilibrium (Ingvarsson and Street, 2010; Rafalski, 2010), substitution and deletion lines (Harper et al., 2011), TILLING (McCallum et al., 2000) or T-DNA insertion libraries (Krishnan et al., 2009) may be required to associate genes to specific functions.

Impacts of the perennial ryegrass GenomeZipper for forage and turf grass genomics

Apart from revising our understanding of the genomic relationship of perennial ryegrass to well described grass species, the GenomeZipper will be useful for a broad range of forage and turf grass species that are close relatives of perennial ryegrass within Poeae but – so far – not well characterized. For *Poa*, *Dactylis*, and *Phleum* species, for example, the perennial ryegrass GenomeZipper constitute a unique tool for efficient development of markers at any genome position that underlie trait variation in perennial ryegrass and/or other major grass species such as barley, Brachypodium, rice, and sorghum. As an example, multiple sequence alignments of genes conserved within Poaceae that have a well defined biological function can easily be generated by means of the GenomeZipper. Conserved regions within these sequence alignments can be identified and then used for primer design in order to isolate orthologs in the species of interest. This will greatly benefit linkage mapping-based QTL analysis or candidate gene-based association mapping in genetically more complex *Poa* and *Phleum* species, where linkage mapping is generally difficult (Barcaccia et al., 1998; Porceddu et al., 2002). For other forage and turf grass species

such as fescues (*Festuca* spp.), for which considerably more EST or genomic sequence resources are available, the present study provides the technological tools for the development of GenomeZippers in these species. In the future, we envision using next generation transcriptome sequencing in uncharacterized forage and turf grass species and aligning the assembled genes to the perennial ryegrass GenomeZipper *in silico*, thereby rapidly obtaining high resolution maps.

CONCLUSIONS

The GenomeZipper presented here is an ordered, information rich scaffold of the perennial ryegrass genome that can be used to unlock the genomes of the most important forage and turf grass species. It constitutes an important tool for the assignment of candidate genes to QTL regions, thereby accelerating map-based cloning. It will also assist functional studies and the assembly of the perennial ryegrass genome and other pooid grasses. Ultimately, GenomeZipper-based comparative genomics enables the maximum use of the significant investments that have been made in establishing genomic resources for model species, allowing us to accelerate research in orphan crop species.

MATERIALS AND METHODS

Perennial ryegrass marker scaffold and unigenes

The genetic linkage map of expressed perennial ryegrass genes (Studer et al., 2012) served as a scaffold for the GenomeZipper. In brief, the VrnA two-way pseudo-testcross mapping population consisting of 184 F_2 perennial ryegrass genotypes was used to map EST-derived SNPs. For *in silico* mapping, 8,876 non-redundant, unmapped unigenes were used (Asp et al., 2007).

Bridging grass genomic synteny from barley to perennial ryegrass

For building the synteny bridge between perennial ryegrass and Brachypodium, rice, and sorghum via the virtual barley genome, synteny comparison between the perennial ryegrass and the barley genome was performed first. Virtual barley chromosomes were created by concatenating the sequences of barley fl-cDNAs as ordered in the barley GenomeZipper (Mayer et al., 2011). Then, available perennial ryegrass marker sequences were compared chromosome-wise against barley using BLASTN search (E value ≤1e-5, identity ≥85%, match length ≥100 bp). Matches were visualized as heatmaps using a manual python script by counting the number of hits in a 250 kbp window that was moved in 50 kbp steps. Syntenic relationships between perennial ryegrass and Brachypodium, rice, and sorghum were derived by transferring available syntenic information between barley and the model grass species to the perennial ryegrass genome. For the chromosomes L1, L2, L3, L6, and L7, syntenic segments of the corresponding barley chromosomes 1H, 2H, 3H, 6H, and 7H were conferred upon the perennial ryegrass chromosomes as previously reported for barley (Mayer et al., 2011). For chromosome L4, the Brachypodium, rice, and sorghum genes that were anchored in the corresponding region of the barley GenomeZipper were extracted first. This region started from locus 2380, the anchoring position of the first tagged barley fl-cDNA of this syntenic segment on chromosome 5H (AK250782.1). To define syntenic segments in model grass genomes, the extracted genes were separated according to their chromosomal origin, permitting determination of the start and end positions of the corresponding region in each model grass genome. Similarly, the available syntenic information up to locus 2379 from the 5H has been transferred to chromosome L5.

Identification of syntenic genes and *in silico* mapping of perennial ryegrass genes

Nucleotide sequences of 8,876 unmapped perennial ryegrass unigenes were compared to protein sequences of Brachypodium, rice, and sorghum genes using BLASTX sequence analysis. Only matches with an E value ≤1e-5, ≥70% identity and match length ≥30 amino acids were considered. Tagged Brachypodium, rice, and sorghum genes located in a syntenic section were extracted for subsequent anchoring in the GenomeZipper. Unigenes significantly matching a syntenic gene were assigned to a particular perennial ryegrass chromosome, if the matched gene was defined as syntenic to exactly one distinct perennial ryegrass chromosome. ESTs with ambiguous matches to two or more syntenic genes located in syntenic regions associated to different perennial ryegrass chromosomes were neglected.

Construction of the perennial ryegrass GenomeZipper

The linear gene order model for perennial ryegrass was established using the GenomeZipper approach as described by Mayer et al. (2011). Chromosome scaffolds were created by combining the known order of the available genetic markers and a gene template, which is built on synteny between the model grass species Brachypodium, rice, and sorghum as described above. For ordering and structuring the genome scaffolds, the syntenic genes were first joined to the marker scaffold using bbh comparisons. To complete the gene template, syntenic genes without marker-associations were arranged based on the principle of closest evolutionary distance. Furthermore, barley fl-cDNAs associated with the entities were included into the genome scaffold as identified by bbh sequence comparisons to perennial ryegrass markers and syntenic genes. Finally, the chromosome-assigned unigenes were integrated by joining them to the matched backbone elements. The development of the perennial ryegrass GenomeZipper was summarised and illustrated in Supplemental Figure S4.

K_a/K_s analysis

Putative orthologous gene pairs between perennial ryegrass and barley, Brachypodium, rice, and sorghum were identified based on a stringent bbh BLAST strategy. For barley fl-cDNAs protein sequence were predicted by using OrfPredictor (Min et al., 2005). Then, BLASTX analysis was used to determine protein alignments between perennial ryegrass EST sequences and orthologous genes form barley,

Brachypodium, rice, and sorghum. The best scoring alignments spanning at least 50 amino acids without any internal stop codon were filtered for further analysis. For each alignment of orthologous gene pairs the rate of non-synonymous (K_a) to synonymous (K_s) substitutions was estimated using the yn00 module of the PAML 4 suite (Yang, 2007). For further statistical investigations, only Ks values <2.0 were considered. Divergence times were calculated by $Ks/2T=\lambda$ assuming a substitution rate of $\lambda=5.1\text{-}7.1x10^{-9}$ substitutions per site per year (Wolfe et al., 1989).

The perennial ryegrass GenomeZipper online

The perennial ryegrass GenomeZipper including GenBank accession numbers and detailed marker information is available online under: http://mips.helmholtz-muenchen.de/plant/lolium/index.jsp

ACKNOWLEDGMENTS

The authors would like to acknowledge Mr. Manuel Spannagl from the Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich, Germany, for the integration of the perennial ryegrass Genome Zipper into the MIPS plant genome database.

LITERATURE CITED

Alm V, Busso CS, Ergon Å, Rudi H, Larsen A, Humphreys MW, Rognli OA (2011) QTL analyses and comparative genetic mapping of frost tolerance, winter survival and drought tolerance in meadow fescue (*Festuca pratensis* Huds.). Theor Appl Genet **123:** 369-382

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

Asp T, Byrne S, Gundlach H, Bruggmann R, Mayer KFX, Andersen JR, Xu M, Greve M, Lenk I, Lübberstedt T (2011) Comparative sequence analysis of *VRN1* alleles of *Lolium perenne* with the co-linear regions in barley, wheat, and rice. Mol Gen Genomics **286:** 433-447

Asp T, Frei UK, Didion T, Nielsen KK, Lübberstedt T (2007) Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*. BMC Plant Biol **7:** 36

Barcaccia G, Mazzucato A, Albertini E, Zethof J, Gerats A, Pezzotti M, Falcinelli M (1998) Inheritance of parthenogenesis in *Poa pratensis* L.: auxin test and AFLP linkage analyses support monogenic control. Theor Appl Genet **97:** 74-82

Bennett MD, Smith JB (1991) Nuclear DNA amounts in angiosperms. Phil Trans R Soc Lond B **334:** 309-345

Berkman PJ, Skarshewski A, Lorenc MT, Lai K, Duran C, Ling EYS, Stiller J, Smits L, Imelfort M, Manoli S, et al. (2011) Sequencing and assembly of

low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. Plant Biotechnol J **9:** 768-775

Bouchenak-Khelladi Y, Salamin N, Savolainen V, Forest F, van der Bank M, Chase MW, Hodkinson TR (2008) Large multi-gene phylogenetic trees of the grasses (Poaceae): Progress towards complete tribal and generic level sampling. Mol Phylogenet Evol **47:** 488-505

Byrne S, Hedegaard J, Panitz F, Bendixen C, Studer B, Farrell JD, Swain S, Armstead I, Caccamo M, Asp T (2011) *De novo* genome sequencing of perennial ryegrass (*Lolium perenne* L.). *In* EUCARPIA - 29th Fodder Crops and Amenity Grass Section Meeting, Dublin, Ireland, p 23

Catalán P, Müller J, Hasterok R, Jenkins G, Mur LAJ, Langdon T, Betekhtin A, Siwinska D, Pimentel M, López-Alvarez D (2012) Evolution and taxonomic split of the model grass *Brachypodium distachyon.* Ann Bot **109:** 385-405

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol **8:** 155-162

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell **12:** 637-646

Deynze AE, Nelson JC, O'Donoughue LS, Ahn SN, Siripoonwiwat W, Harrington SE, Yglesias ES, Braga DP, McCouch SR, Sorrells ME (1995) Comparative mapping in grasses. Oat relationships. Mol Gen Genet **249:** 349-356

Doležel J, Greilhuber J, Lucretti S, Meister A, Lysák MA, Nardi L, Obermayer R (1998) Plant genome size estimation by flow cytometry: Inter-laboratory comparison. Ann Bot **82:** 17-26

Feuillet C, Keller B (2002) Comparative genomics in the grass family: Molecular characterization of grass genome structure and evolution. Ann Bot **89:** 3-10

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science **282:** 656-659

Gaut BS (2002) Evolutionary dynamics of grass genomes. New Phytol **154:** 15-28

Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). Ann Mo Bot Gard **88:** 373-457

Hackauf B, Rudd S, van der Voort J, Miedaner T, Wehling P (2009) Comparative mapping of DNA sequences in rye (*Secale cereale* L.) in relation to the rice genome. Theor Appl Genet **118:** 371-384

Harper J, Armstead I, Thomas A, James C, Gasior D, Bisaga M, Roberts L, King I, King J (2011) Alien introgression in the grasses *Lolium perenne* (perennial ryegrass) and *Festuca pratensis* (meadow fescue): the development of seven monosomic substitution lines and their molecular and cytological characterization. Ann Bot **107:** 1313-1321

Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G (2006) Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using BAC landing with fluorescent in situ hybridization. Genetics **173:** 349-362

Hernandez P, Martis M, Dorado G, Pfeifer M, Gálvez S, Schaaf S, Jouve N, Šimková H, Valárik M, Doležel J, et al. (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. Plant J **69:** 377-386

Huo N, Garvin DF, You FM, McMahon S, Luo M, Cheng, Gu YQ, Lazo GR, Vogel JP (2011) Comparison of a high-density genetic linkage map to genome features in the model grass *Brachypodium distachyon*. Theor Appl Genet **123:** 455-464

Ingvarsson PK, Street NR (2010) Association genetics of complex traits in plants. New Phytol **189:** 909-922

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature **436:** 793-800

Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphreys MO, King IP, Kishida T, Yamada T, Balfourier F, 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

Kellogg EA, Bennetzen JL (2004) The evolution of nuclear genome structure in seed plants. Am J Bot **91:** 1709-1725

King J, Armstead IP, Donnison IS, Thomas HM, Jones RN, Kearsey MJ, Roberts LA, Thomas A, Morgan WG, King IP (2002) Physical and genetic mapping in the grasses *Lolium perenne* and *Festuca pratensis*. Genetics **161:** 315-324

King J, Armstead IP, Donnison SI, Roberts LA, Harper JA, Skøt K, Elborough K, King IP (2007) Comparative analyses between *Lolium/Festuca* introgression lines and rice reveal the major fraction of functionally annotated gene models is located in recombination-poor/very recombination-poor regions of the genome. Genetics **177:** 597-606

Kopecký D, Havránková M, Loureiro J, Castro S, Lukaszewski AJ, Bartoš J, Kopecká J, Doležel J (2010) Physical distribution of homoeologous recombination in individual chromosomes of *Festuca pratensis* in *Lolium multiflorum.* Cytogenet Genome Res **129:** 162-172

Krishnan A, Guiderdoni E, An G, Hsing Y-iC, Han C-d, Lee MC, Yu S-M, Upadhyaya N, Ramachandran S, Zhang Q, et al. (2009) Mutant resources in rice for functional genomics of the grasses. Plant Physiology **149:** 165-170

Liu Z, Pagani M, Zinniker D, DeConto R, Huber M, Brinkhuis H, Shah SR, Leckie RM, Pearson A (2009) Global cooling during the Eocene-Oligocene climate transition. Science **323:** 1187-1190

Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, et al. (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. Proc Natl Acad Sci USA **106:** 15780-15785

Massa AN, Wanjugi H, Deal KR, O'Brien K, You FM, Maiti R, Chan AP, Gu YQ, Luo MC, Anderson OD, et al. (2011) Gene space dynamics during the evolution of *Aegilops tauschii*, *Brachypodium distachyon*, *Oryza sativa*, and *Sorghum bicolor* genomes. Mol Biol Evol **28:** 2537-2547

Mayer KFX, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, et al. (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell **23:** 1249-1263

Mayer KFX, Taudien S, Martis M, Šimková H, Suchánková P, Gundlach H, Wicker T, Petzold A, Felder M, Steuernagel B, et al. (2009) Gene content and virtual gene order of barley chromosome 1H. Plant Physiology **151:** 496-505

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeted screening for induced mutations. Nat Biotech **18:** 455-457

Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res **33:** W677-W680

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al. (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature **457:** 551-556

Porceddu A, Albertini E, Barcaccia G, Falistocco E, Falcinelli M (2002) Linkage mapping in apomictic and sexual Kentucky bluegrass (*Poa pratensis* L.) genotypes using a two way pseudo-testcross strategy based on AFLP and SAMPL markers. Theor Appl Genet **104:** 273-280

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol **13:** 174-180

Salse J, Abrouk M, Bolot S, Guilhot N, Courcelle E, Faraut T, Waugh R, Close TJ, Messing J, Feuillet C (2009) Reconstruction of monocotelydoneous proto-chromosomes reveals faster evolution in plants than in animals. Proc Natl Acad Sci USA **106:** 14908-14913

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell **20:** 11-24

Salse J, Feuillet C (2007) Comparative genomics of cereals. *In* R Varshney, R Tuberosa, eds, Genomics-Assisted Crop Improvement, Vol 1. Springer-Verlag, Dordrecht, The Netherlands, pp 177-205

Sandve SR, Rudi H, Asp T, Rognli OA (2008) Tracking the evolution of a cold stress associated gene family in cold tolerant grasses. BMC Evol Biol **8:** 245

Sandve SR, Fjellheim S (2010) Did gene family expansions during the Eocene–Oligocene boundary climate cooling play a role in Pooideae adaptation to cool climates? Mol Ecol **19:** 2075-2088

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science **326:** 1112-1115

Shinozuka H, Cogan N, Smith K, Spangenberg G, Forster J (2009) Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* L.). Plant Mol Biol **72:** 343-355

Sim S, Chang T, Curley J, Warnke SE, Barker RE, Jung G (2005) Chromosomal rearrangements differentiating the ryegrass genome from the Triticeae, oat, and rice genomes using common heterologous RFLP probes. Theor Appl Genet **110:** 1011-1019

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res **13:** 1818-1827

Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney R, Perovic D, Grosse I, et al. (2007) A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. Theor Appl Genet **114:** 823-839

Studer B, Byrne S, Nielsen RO, Panitz F, Bendixen C, Islam MS, Pfeifer M, Lübberstedt T, Asp T (2012) A transcriptome map of perennial ryegrass (*Lolium perenne* L.). BMC Genomics **13:** 140

Swain S, Thomas A, Kudrna D, Huang L, Byrne S, Asp T, King J, King IP, Armstead I (2011) The development of a BAC-based physical map of perennial ryegrass (*Lolium perenne*) derived from high information content fingerprints and BAC-end sequences. *In* EUCARPIA - 29th Fodder Crops and Amenity Grass Section Meeting, Dublin, Ireland, p 16

Tamura K, Yonemaru J, Hisano H, Kanamori H, King J, King IP, Tase K, Sanada Y, Komatsu T, Yamada T (2009) Development of intron-flanking EST markers for the Lolium/Festuca complex using rice genomic information. Theor Appl Genet **118:** 1549-1560

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature **463:** 763-768

Vitulo N, Albiero A, Forcato C, Campagna D, Dal Pero F, Bagnaresi P, Colaiacovo M, Faccioli P, Lamontanara A, Šimková H, et al. (2011) First survey of the wheat chromosome 5A composition through a next generation sequencing approach. PLoS ONE **6:** e26421

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Šimková H, Kubaláková M, Choulet F, Taudien S, et al. (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. The Plant Cell **23:** 1706-1718

Wolfe KH, Scharp PM, Li W-H (1989) Rates of synonymous substitutions in plant nuclear genes. J Mol Evol **29**: 208-211

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol **24:** 1586-1591

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. Science **292:** 686-693

Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, et al. (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat Biotech **30:** 549-554

FIGURE LEGENDS

Figure 1. The heatmaps illustrate the degree of macro-synteny between perennial ryegrass and barley. For each of the seven perennial ryegrass linkage groups (LGs), the EST sequences of mapped DNA markers were compared against full length (fl)-cDNA sequences of barley that were arranged in the virtual barley genome (Mayer et al., 2011) using sequence homology analysis. Connections indicate the position of the perennial ryegrass DNA marker and its associated barley fl-cDNA. The syntenic regions of each perennial ryegrass LG are indicated by the increased height of the heatmaps.

Figure 2. Syntenic relationship between perennial ryegrass and (A) Brachypodium (Bd), (B) rice (Os) and (C) sorghum (Sb). Heatmaps represent entire syntenic chromosomes of Brachypodium, rice and sorghum. Coloured bars visualize that part of the chromosome which was defined as syntenic to perennial ryegrass via the barley bridge. Chromosomes are assigned according to the colour key. The colour of the heatmaps illustrates the density of perennial ryegrass marker sequences matching the Brachypodium, rice and sorghum genome.

Figure 3. Venn diagrams of the perennial ryegrass GenomeZipper. Each figure represents the number of Brachypodium (Bd), rice (Os) and sorghum (Sb) genes anchored in the GenomeZipper for each individual perennial ryegrass chromosome L1 to L7, as well as for the combined zipped-up genome. Intersections between circles indicate the number of genes which were anchored at a single unambiguous locus in all species.

Figure 4. Analysis of sequence divergence between perennial ryegrass, barley, Brachypodium, rice and sorghum based on synonymous substation rates (Ks). (A) Frequency distribution of synonymous substitution rates based on protein alignments of perennial ryegrass genes to 3,301 orthologous barley, 3,789 Brachypodium, 3,434 rice and 3,528 sorghum genes. (B) Based on mean Ks rates against barley (0.31), Brachypodium (0.33), rice (0.53) and sorghum (0.59), the divergence times of perennial ryegrass from barley and Brachypodium, rice and sorghum was estimated to 22-30, 23-32, 37-52 and 42-58 million years ago (Mya), respectively.

Figure 5. Micro-synteny between perennial ryegrass and barley. Conserved blocks between the seven chromosomes of perennial ryegrass (horizontal axis, L1 through L7) and barley (vertical axis, 1H through 7H) are shown by comparison of common anchored Brachypodium genes. Each dot represents a Brachypodium gene (coloured according to its chromosomal origin) that was anchored in both the perennial ryegrass and the barley GenomeZipper. The x- and y-axis are scaled according to the anchoring loci in the perennial ryegrass and barley GenomeZipper, respectively. Grey rectangles indicate loci of the barley GenomeZipper that are located in the genetic centromere region of the corresponding chromosome.

Table 1. Statistics of input data used for the construction of the perennial ryegrass GenomeZipper

Linkage group (LG)	**Number of mapped genes**	**Sequence length [bp]**				**Map Size [cM]**
		Min	**Max**	**Average**	**Total**	
LG 1	88	309	2,007	855	75,259	97.7
LG 2	125	263	1,590	826	103,351	151.5
LG 3	111	293	1,779	871	96,757	63.3
LG 4	154	324	2,278	902	138,934	119.2
LG 5	79	331	2,723	929	73,443	89.1
LG 6	86	248	1,810	924	79,530	115.2
LG 7	119	193	3,623	910	108,310	113.7
Total	762			888	675,584	750
Unmapped perennial ryegrass unigenes	8,876	100	2,720	575	5,108,379	

Table 2. Syntenic grass genes and chromosome-assigned unigenes incorporated in the GenomeZipper for each of the seven perennial ryegrass chromosomes L1 to L7

	L1	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**Sum**
Marker sequences matching full length-cDNAs of barley	47	41	54	45	33	35	46	301
Syntenic Brachypodium genes	513	563	535	705	692	437	463	3,926
Syntenic rice genes	443	524	522	598	387	400	381	3,255
Syntenic sorghum genes	436	485	504	524	470	383	436	3,238
Unambiguously chromosome-assigned perennial ryegrass unigenes	476	558	538	447	408	425	463	3,315

Table 3. General overview of the perennial ryegrass GenomeZipper: Results are given for each of the seven perennial ryegrass chromosomes L1 to L7 and summed-up for the whole genome.

	L1	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**Sum**
Loci with perennial ryegrass gene evidence	565	690	636	586	500	487	571	4,035
Marker sequences matching Brachypodium, rice or sorghum genes	19	13	28	26	15	18	13	132
Anchored barley full length-cDNAs	395	408	400	341	277	276	341	2,438
Anchored perennial ryegrass unigenes (bidirectional hit)	408	501	461	384	344	372	395	2,865
Anchored perennial ryegrass unigenes (first-best hit)	476	558	538	445	408	425	463	3,313
Anchored Brachypodium genes	397	451	450	375	365	350	370	2,758
Anchored rice genes	322	408	403	313	205	311	308	2,270
Anchored sorghum genes	334	390	409	320	257	297	344	2,351

Table 4. Comparison of syntenic genes from Brachypodium, rice, and sorghum that were anchored in the perennial ryegrass, barley, and in both GenomeZippers

	Perennial ryegrass only	**Barley only**	**Perennial ryegrass and barley**	**Sum**
Brachypodium	383	681	2,375	3,439
Rice	313	581	1,957	2,851
Sorghum	343	473	2,008	2,824

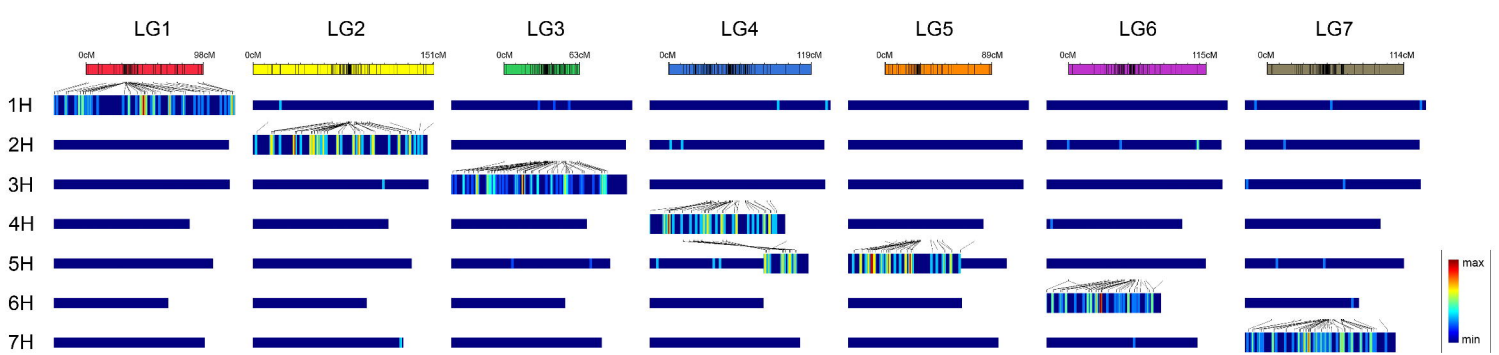

Figure 1. The heatmaps illustrate the degree of macro-synteny between perennial ryegrass and barley. For each of the seven perennial ryegrass linkage groups (LGs), the EST sequences of mapped DNA markers were compared against full length (fl)-cDNA sequences of barley that were arranged in the virtual barley genome (Mayer et al., 2011) using sequence homology analysis. Connections indicate the position of the perennial ryegrass DNA marker and its associated barley fl-cDNA. The syntenic regions of each perennial ryegrass LG are indicated by the increased height of the heatmaps.

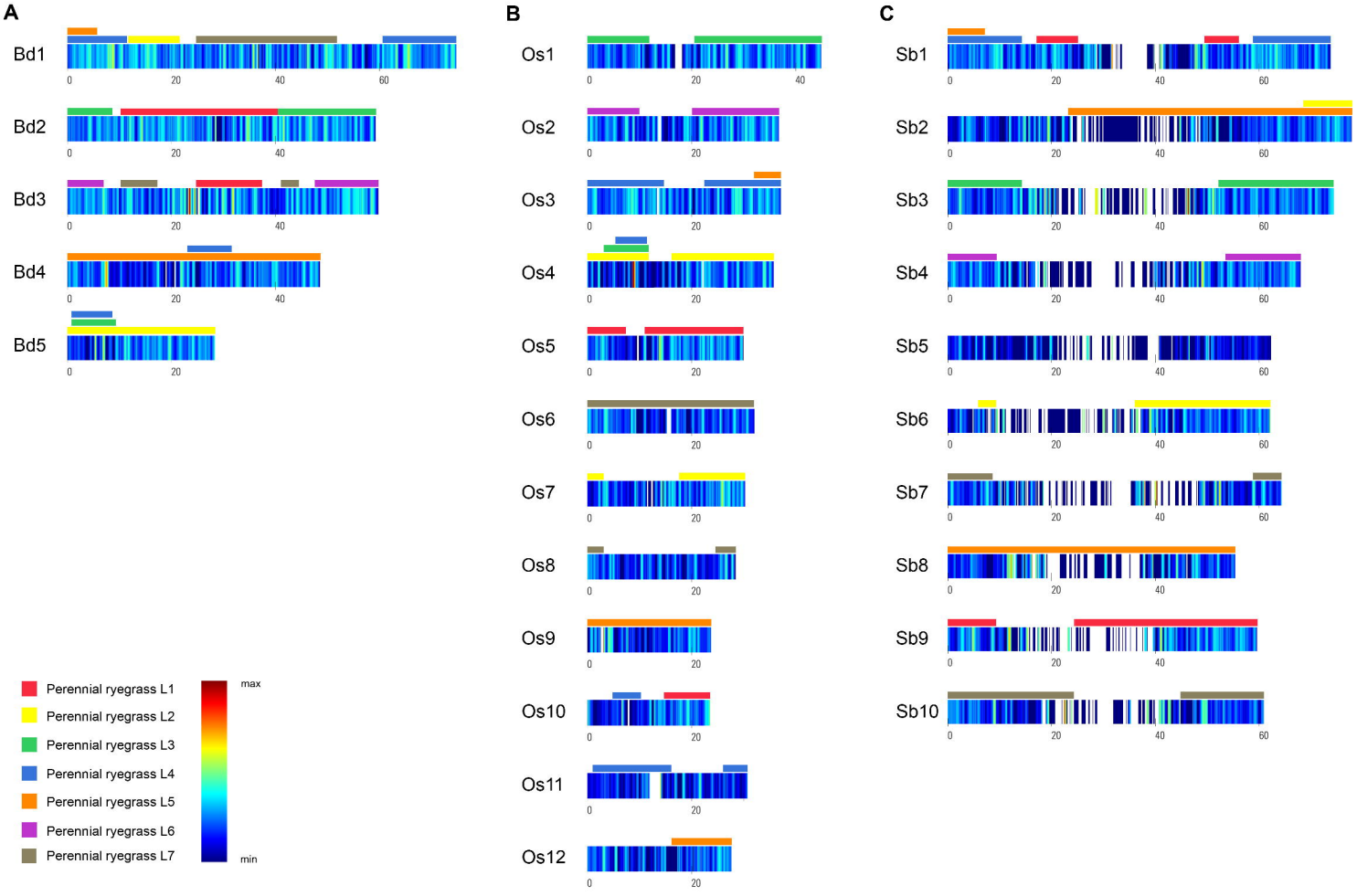

Figure 2. Syntenic relationship between perennial ryegrass and (A) Brachypodium (Bd), (B) rice (Os) and (C) sorghum (Sb). Heatmaps represent entire syntenic chromosomes of Brachypodium, rice and sorghum. Coloured bars visualize that part of the chromosome which was defined as syntenic to perennial ryegrass via the barley bridge. Chromosomes are assigned according to the colour key. The colour of the heatmaps illustrate the density of perennial ryegrass marker sequences matching the Brachypodium, rice and sorghum genome.

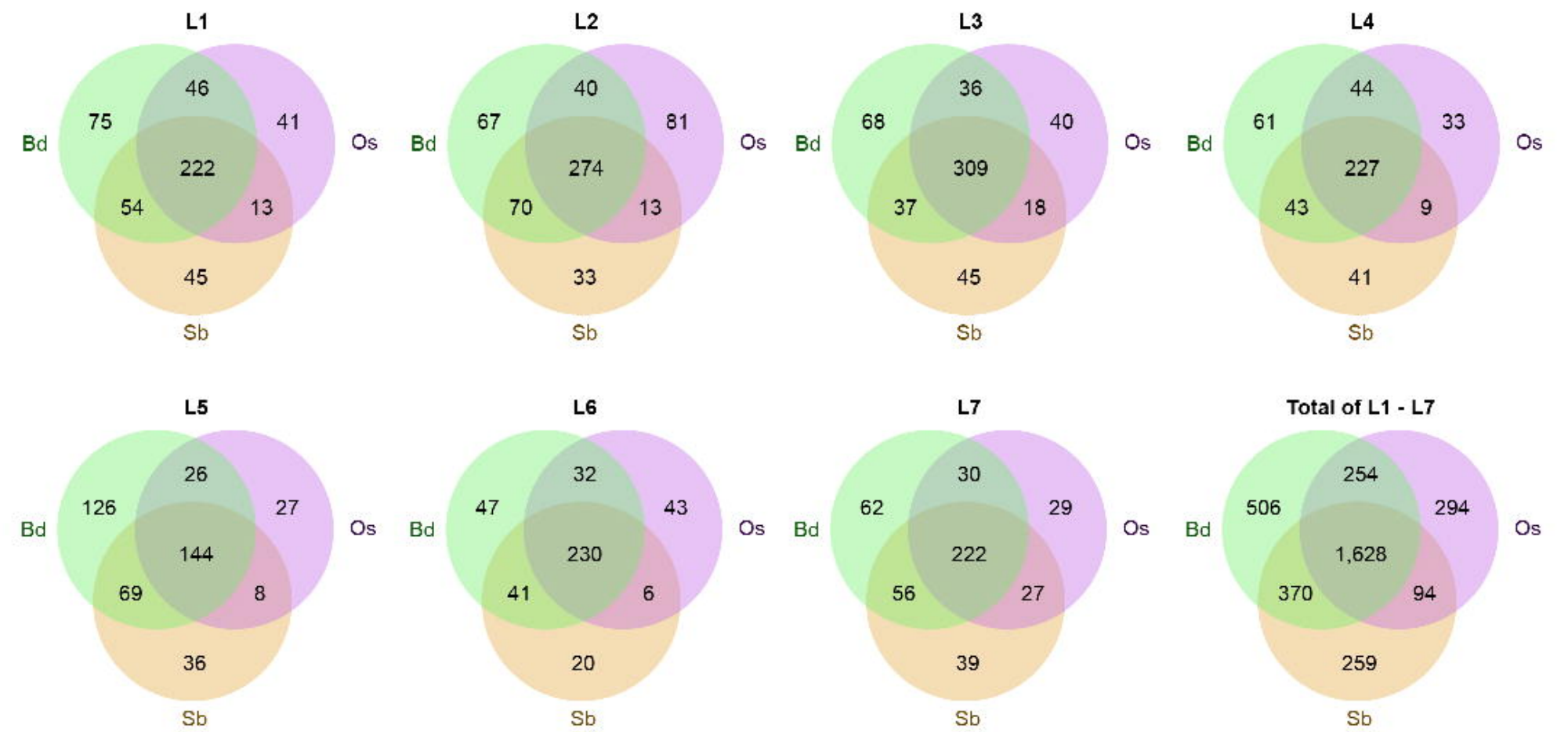

Figure 3. Venn diagrams of the perennial ryegrass GenomeZipper. Each figure represents the number of Brachypodium (Bd), rice (Os) and sorghum (Sb) genes anchored in the GenomeZipper for each individual perennial ryegrass chromosome L1 to L7, as well as for the combined zipped-up genome. Intersections between circles indicate the number of genes which were anchored at a single unambiguous locus in all species.

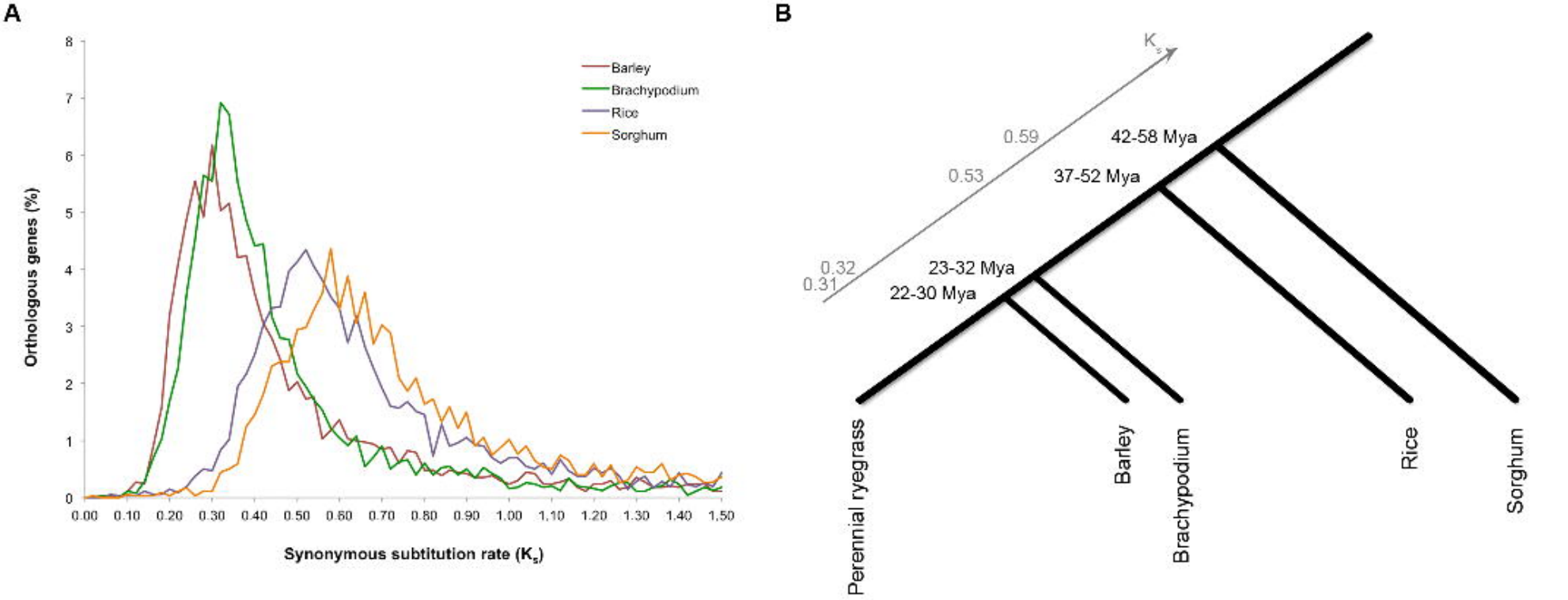

Figure 4. Analysis of sequence divergence between perennial ryegrass, barley, Brachypodium, rice and sorghum based on synonymous substation rates (K_s). (A) Frequency distribution of synonymous substitution rates based on protein alignments of perennial ryegrass genes to 3,301 orthologous barley, 3,789 Brachypodium, 3,434 rice and 3,528 sorghum genes. (B) Based on mean K_s rates against barley (0.31), Brachypodium (0.33), rice (0.53) and sorghum (0.59), the divergence times of perennial ryegrass from barley and Brachypodium, rice and sorghum was estimated to 22-30, 23-32, 37-52 and 42-58 million years ago (Mya), respectively.

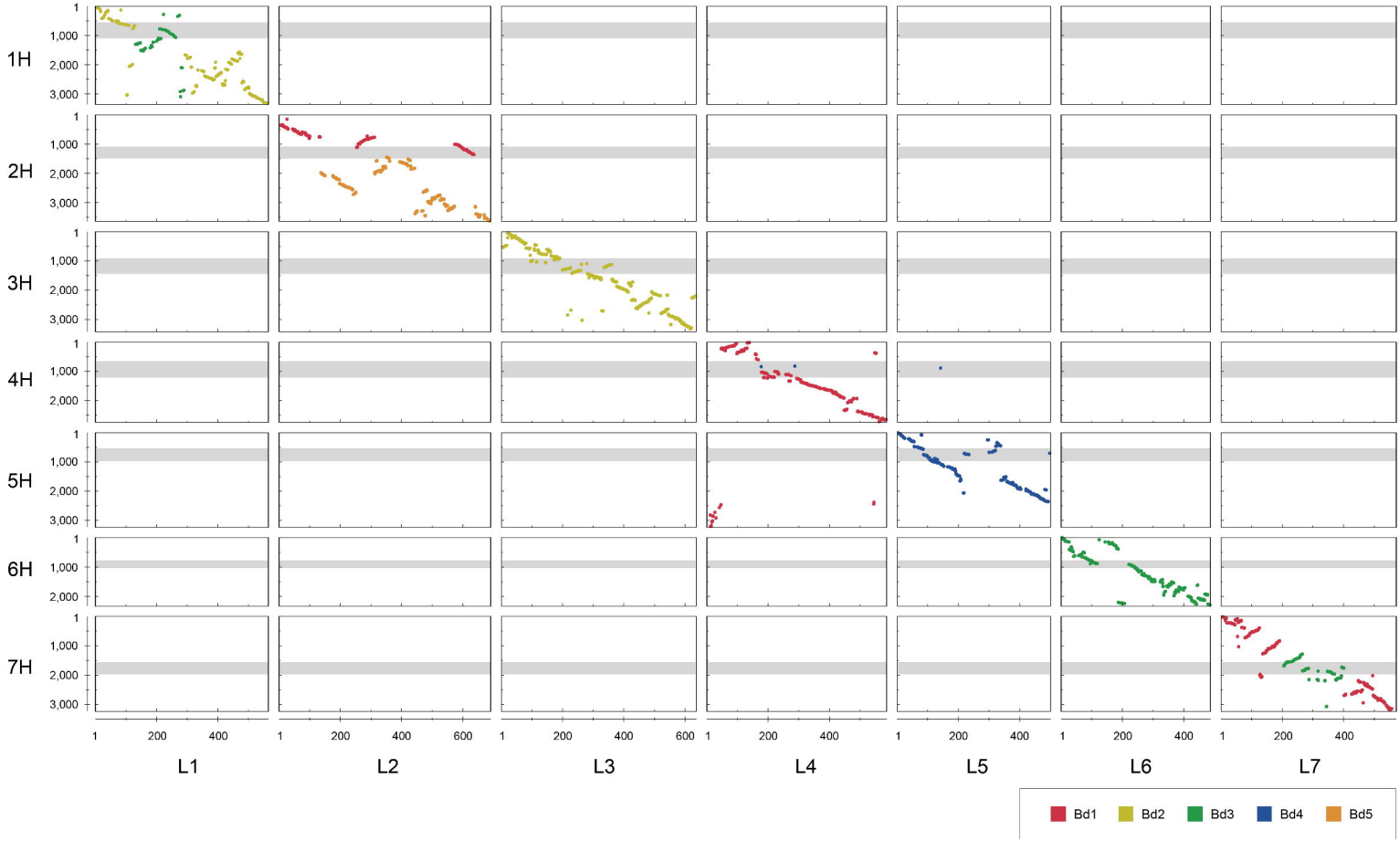

Figure 5. Micro-synteny between perennial ryegrass and barley. Conserved blocks between the seven chromosomes of perennial ryegrass (horizontal axis, L1 through L7) and barley (vertical axis, 1H through 7H) are shown by comparison of common anchored Brachypodium genes. Each dot represents a Brachypodium gene (coloured according to its chromosomal origin) that was anchored in both the perennial ryegrass and the barley GenomeZipper. The x- and y-axis are scaled according to the anchoring loci in the perennial ryegrass and barley GenomeZipper, respectively. Grey rectangles indicate loci of the barley GenomeZipper that are located in the genetic centromere region of the corresponding chromosome.