The Perennial Ryegrass GenomeZipper: Targeted Use of Genome Resources for Comparative Grass Genomics[1][C][W]

Matthias Pfeifer, Mihaela Martis, Torben Asp, Klaus F.X. Mayer, Thomas Lübberstedt, Stephen Byrne, Ursula Frei, and Bruno Studer*

Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany (M.P., M.M., K.F.X.M.); Department of Molecular Biology and Genetics, Faculty of Science and Technology, Research Centre Flakkebjerg, Aarhus University, 4200 Slagelse, Denmark (T.A., S.B.); Department of Agronomy, Iowa State University, Ames, Iowa 50011 (T.L., U.F.); and Institute of Agricultural Sciences, Forage Crop Genetics, Eidgenössisch Technische Hochschule Zurich, 8092 Zurich, Switzerland (B.S.)

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 have yet to be developed. Here, we present a model of the perennial ryegrass (Lolium perenne) genome on the basis of conserved synteny to barley (Hordeum vulgare) and the model grass genome Brachypodium (Brachypodium distachyon) as well as rice (Oryza sativa) and sorghum (Sorghum bicolor). 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 macrocollinearity and was then utilized to anchor a collection of perennial ryegrass genes in silico to their predicted genome positions. 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 spp.), rice (Oryza sativa), maize (Zea mays), barley (Hordeum vulgare), and sorghum (Sorghum bicolor) belong to the grass family. Significant investment in these crop species and the model grass species Brachypodium (Brachypodium distachyon) 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 (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 Gb [Bennett and Smith, 1991]) and barley (2n = 2x = 14, 5.1 Gb [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, 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 nonmodel 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 macrocollinearity 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, 2009).
More recent studies proposed that grass genomes have evolved from a five-chromosome ancestral genome into a 12-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 the subfamilies Bambusoideae and Ehrhartioideae (Grass Phylogeny Working Group, 2001). Within the subfamily Pooideae, the tribe Brachypodieae is more distantly related to the Poaeae than the Triticeae, Aveneae, and Bromeeae (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) is a diploid (2n = 2x = 14) member of the subfamily Pooideae, belonging to the tribe Poaeae, genus Lolium. First attempts to describe synteny between Poaeae, 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) and the Triticeae genomes by comparative mapping of 117 loci with known map positions (Alm et al., 2003). The chromosome translocation involving the long arms of chromosomes 4 and 5 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 (hereafter 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. First, the 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 led 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 colinearity (Salse et al., 2008). Second, 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 the 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 positions 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 Gb (1 C = 2,623 Mb [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 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 syntonic regions in these reference genomes and arranges the syntetic 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 chromosomes (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 locations of unmapped genes and allows gene-based marker development in specific target regions to accelerate fine-mapping and map-based cloning of genes or quantitative trait loci (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 (1) characterizing chromosomal rearrangements of syntenic genes between perennial ryegrass, barley, and sequenced grass species by means of sequence homology analysis, (2) predicting in silico the genomic locations of genes by using the identified syntenic relationships, and (3) applying the GenomeZipper to integrate barley, rice, sorghum, and Brachypodium genes in a synteny model that facilitates 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 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 3,623 bp, with an average of 889 bp and with a mean GC content of 49.6% (Table I). 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 (hereafter referred to as unigenes) were used for in silico mapping. These unigenes represent more than 5 Mb of nucleotide sequence information of the perennial ryegrass transcriptome (Table I).

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% or greater sequence identity over at least 100 bp). Using a sliding-window approach (250 kb window size, 50 kb window shift), the number of syntenic barley genes matched by marker sequences from each LG was calculated and illustrated as heat maps (Fig. 1). In total, 342 of 762 (45%) markers matched a barley full-length (fl)-complementary DNA (cDNA) sequence (Table II). A close syntenic relationship between the perennial ryegrass LGs and the barley chromosomes (hereafter referred to as 1H–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; Fig. 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 Triticaceae 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 2,540). On 5H, we identified three nonsyntenic loci (AK370577, AK365621, and AK374623) as well as one large syntenic segment located on the distal end of the long arm. This segment ranged from AK250782.1 (locus 2,380) to AK250137.1 (locus 2,973) and covered almost 20% of the chromosome. In comparison with 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 2,295 (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 by Mayer et al. (2011) into the perennial

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

<table>
<thead>
<tr>
<th>LG</th>
<th>No. of Mapped Genes</th>
<th>Sequence Length</th>
<th>Map Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bp</td>
<td>bp</td>
</tr>
<tr>
<td>LG 1</td>
<td>88</td>
<td>309</td>
<td>2,007</td>
</tr>
<tr>
<td>LG 2</td>
<td>125</td>
<td>263</td>
<td>1,590</td>
</tr>
<tr>
<td>LG 3</td>
<td>111</td>
<td>293</td>
<td>1,779</td>
</tr>
<tr>
<td>LG 4</td>
<td>154</td>
<td>324</td>
<td>2,278</td>
</tr>
<tr>
<td>LG 5</td>
<td>79</td>
<td>331</td>
<td>2,723</td>
</tr>
<tr>
<td>LG 6</td>
<td>86</td>
<td>248</td>
<td>1,810</td>
</tr>
<tr>
<td>LG 7</td>
<td>119</td>
<td>193</td>
<td>3,623</td>
</tr>
<tr>
<td>Total</td>
<td>762</td>
<td>762</td>
<td>888</td>
</tr>
</tbody>
</table>

Unmapped perennial ryegrass unigenes 8,876 100 2,720 575 5,108,379

The Perennial Ryegrass GenomeZipper

www.plantphysiol.org
ryegrass genome structure (Supplemental Fig. 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 (Fig. 2). Over all chromosomes, 3,926 syntenic genes of Brachypodium, 3,255 of rice, and 3,238 of sorghum were identified (Table II).

**In Silico Chromosome Assignment of Perennial Ryegrass Genes**

The perennial ryegrass unigenes were first assigned to their putative chromosomal origin by comparing the ESTs with protein sequences of annotated genes of Brachypodium, rice, and sorghum (BLASTX; 70% or greater sequence identity covering at least 30 amino acids) that 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%) tagged unigenes (between 408 and 558 on each chromosome) were assigned to chromosomes in silico (Table II).

**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 II). 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 III; Supplemental Data S1). 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 with 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 GenomeZipper. 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

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

<table>
<thead>
<tr>
<th>Gene</th>
<th>L1</th>
<th>L2</th>
<th>L3</th>
<th>L4</th>
<th>L5</th>
<th>L6</th>
<th>L7</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker sequences matching fl-cDNAs of barley</td>
<td>50</td>
<td>48</td>
<td>60</td>
<td>53</td>
<td>33</td>
<td>43</td>
<td>55</td>
<td>342</td>
</tr>
<tr>
<td>Syntenic Brachypodium genes</td>
<td>513</td>
<td>563</td>
<td>553</td>
<td>705</td>
<td>692</td>
<td>437</td>
<td>463</td>
<td>3,926</td>
</tr>
<tr>
<td>Syntenic rice genes</td>
<td>443</td>
<td>524</td>
<td>522</td>
<td>598</td>
<td>387</td>
<td>400</td>
<td>381</td>
<td>3,255</td>
</tr>
<tr>
<td>Syntenic sorghum genes</td>
<td>436</td>
<td>485</td>
<td>504</td>
<td>524</td>
<td>470</td>
<td>383</td>
<td>436</td>
<td>3,238</td>
</tr>
<tr>
<td>Unambiguously chromosome-assigned perennial ryegrass unigenes</td>
<td>476</td>
<td>558</td>
<td>538</td>
<td>447</td>
<td>408</td>
<td>425</td>
<td>463</td>
<td>3,315</td>
</tr>
</tbody>
</table>
model species. The number of loci exclusively tagged by Brachypodium genes (506) was considerably higher than the counterparts for rice (294) and sorghum (259; Fig. 3).

Genome-Wide Nonsynonymous/Synonymous Substitution Analysis and Estimation of Divergence Time

Sequence divergence analysis between perennial ryegrass, barley, Brachypodium, rice, and sorghum was performed based on the rate of nonsynonymous ($K_a$) 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 $K_a$, $K_s$, and $K_a/K_s$ were analyzed.

Table III. General overview of the perennial ryegrass GenomeZipper

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

Figure 2. Syntenic relationships between perennial ryegrass and Brachypodium (Bd; A), rice (Os; B), and sorghum (Sb; C). Heat maps represent entire syntenic chromosomes of Brachypodium, rice, and sorghum. Colored bars visualize that part of the chromosome that was defined as syntenic to perennial ryegrass via the barley bridge. Chromosomes are assigned according to the color key. The color of the heat maps illustrates the density of perennial ryegrass marker sequences matching the Brachypodium, rice, and sorghum genomes.
values are shown in Figure 4A and Supplemental Figure S2, A and B. 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 $K_s$ rates of orthologous gene pairs were 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; Fig. 4A), the divergence times of perennial ryegrass from barley, Brachypodium, rice, and sorghum was estimated to 22 to 30, 23 to 32, 37 to 52, and 42 to 58 Mya, respectively (Fig. 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 macrocollinearity (Fig. 2). For Brachypodium, the syntenic blocks, defined as colored bars, revealed a highly similar pattern when comparing Brachypodium with barley, Aegilops tauschii (the D genome donor of hexaploid wheat), and wheat (Fig. 2A; 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, and Os11 and Os3, Os9, and Os12, respectively (Fig. 2B). For sorghum chromosome 5, however, no orthologous gene relationship to perennial ryegrass was found (Fig. 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 IV). To visualize collinear blocks, their physical positions (i.e. the anchored loci in the GenomeZipper) were plotted against each other (Fig. 5; Supplemental Fig. S3). The chromosomal origin of each gene is color 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, 2011; Berkman 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 toward developing a reference genome sequence. Such chromosomal templates are also instrumental for genome resequencing and large-scale marker development strategies that will ultimately enable the 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 levels (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 Pooidae 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), rye (Secale cereale), as well as forage and turf grass species (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 with Triticeae and Poeae genomes. Comparative analysis of both the marker scaffold and the GenomeZipper from perennial ryegrass to the virtual barley 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 (N. Stein, personal communication). Indeed, barley is, when compared with Brachypodium, phylogenetically closer to rye-grasses and shares the same number of chromosomes (Bouchenak-Khelladi et al., 2008). The structural conservation of the perennial ryegrass and the barley genomes is surprisingly high, considering that their split dates back to 22 to 30 Mya, only shortly after the split

| Table IV. Comparison of syntenic genes from Brachypodium, rice, and sorghum that were anchored in perennial ryegrass, in barley, and in both GenomeZippers |
|-----------------------------------|-----------------|-----------------|-----------------|-----------------|
| Species                          | 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 4. Analysis of sequence divergence between perennial ryegrass, barley, Brachypodium, rice, and sorghum based on Ks. A, Frequency distribution of Ks 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 to 30, 23 to 32, 37 to 52, and 42 to 58 Mya, respectively.
with Brachypodium 23 to 32 Mya. Subsequent to the
split from Brachypodium, a core Pooideae ancestor
with seven new chromosomes evolved from a 12-
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 to 34 Mya (Zachos et al., 2001; Liu
et al., 2009), as been suggested by Sandve and Fjellheim
(2010).

Similar findings of synteny between perennial rye-
grass 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 nonsyntenic regions that indicated
large-scale chromosomal rearrangements (Jones et al.,
2002; Sim et al., 2005). However, it remained difficult to
resolve if the nonsyntenic 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 the 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 re
flects a closer
phylogenetic relationship of perennial ryegrass to
Brachypodium when compared with rice and sorghum
and is 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 peren-
nial ryegrass, because none of the chromosome fusions

Figure 5. Microsynteny between perennial ryegrass and barley. Conserved blocks between the seven chromosomes of per-
nennial ryegrass (horizontal axis; L1–L7) and barley (vertical axis; 1H–7H) are shown by comparison of common anchored
Brachypodium genes. Each dot represents a Brachypodium gene (colored according to its chromosomal origin) that was an-
chored in both the perennial ryegrass and the barley GenomeZipper. The x and y axes are scaled according to the anchoring
loci in the perennial ryegrass and barley GenomeZipper, respectively. Gray rectangles indicate loci of the barley GenomeZipper
that are located in the genetic centromere region of the corresponding chromosome.
are shared between the two species (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 (Van 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). This study confirmed the absence of the Triticeae-specific translocation in perennial ryegrass and further resolved its chromosomal break point. A detailed structural characterization of this region is of particular interest for comparative genomics of physiological processes such as vernalization, 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 the colocation of these QTL with orthologous QTL and stress tolerance genes such as dehydrins, C-repeat binding factors, or vernalization response genes of other grass species.

The described synteny will also prove highly useful for cytogenetic approaches such as fluorescence in situ hybridization to physically locate genes and chromosome landmarks (Hasterok et al., 2006). For example, syntenic genes and bacterial artificial chromosomes identified using the GenomeZipper can be used for precise and robust molecular karyotyping of ryegrasses or less characterized 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; 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 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 vernalization (Asp et al., 2011) or self-incompatibility (Shinozuka et al., 2010) 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 bacterial artificial chromosome-based physical map (Swain et al., 2011) and de novo sequencing (Byrne et al., 2011) of the perennial ryegrass genome are under way, 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 next-generation sequencing read data yield very fragmented assemblies. However, genomic 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 the 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 provides 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 the identification of candidate genes and the development of functional markers in these regions, alternative strategies based on linkage disequilibrium (Rafalski, 2010; Ingvarsson and Street, 2011), substitution and deletion lines (Harper et al., 2011), TILLING (McCallum et al., 2000), or transfer 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 Poaceae but, so far, are not well characterized. For *Poa* spp., *Dactylis* spp., and *Phleum* spp., for example, the perennial ryegrass GenomeZipper constitutes a unique tool for the 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* spp. and *Phleum* spp., 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, this 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.

CONCLUSION

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 (*Lolium perenne*) genes (Studer et al., 2012) served as a scaffold for the GenomeZipper. In brief, the VrnA two-way pseudotestcross mapping population consisting of 184 F2 perennial ryegrass genotypes was used to map EST-derived single-nucleotide polymorphisms. For in silico mapping, 8,876 nonredundant, 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 (*Brachypodium distachyon*), rice (*Oryza sativa*), and sorghum (*Sorghum bicolor*) via the virtual barley (*Hordeum vulgare*) genome, synteny comparison between the perennial ryegrass and barley genomes 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 ≤ 1e-5, sequence identity of 85% or greater, match length of 100 bp or greater). Matches were visualized as heat maps using a manual python script by counting the number of hits in a 250-kb window that was moved in 50-kb 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 1H, 2H, 3H, 6H, and 7H were conferred upon the perennial ryegrass chromosomes as reported previously 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 2,380, 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 2,379 from chromosome 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 with protein sequences of Brachypodium, rice, and sorghum genes using BLASTX sequence analysis. Only matches with E ≤ 1e-5, 70% or greater sequence identity, and match length of 30 or more 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 with the entities were included into the genome scaffold as identified by bbh sequence comparisons with 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 is summarized and illustrated in Supplemental Figure S4.

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 that 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 with 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 is summarized and illustrated in Supplemental Figure S4.

K$_+/K_-$ 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 flcDNAs, protein sequences were predicted by using OriPredictor (Min et al., 2005). Then, BLASTX analysis was used to determine protein alignments between perennial ryegrass EST sequences and orthologous genes from 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 $K_s$ to $K_t$ was estimated using the yn00 module of the PALM 4 suite (Yang, 2007). For further statistical investigations, only $K_s$ values of less than 2.0 were considered. Divergence times were calculated by $K_t/2T = \lambda$, assuming a substitution rate of $\lambda = 5.1 \times 10^{-9}$ substitutions per site per year (Wolfe et al., 1989).

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

Supplemental Data
The following materials are available in the online version of this article.
Supplemental Figure S1. Illustration of the barley bridge.
Supplemental Figure S2. Genome-wide Ka and Ks analysis.
Supplemental Figure S3. Microsynteny between perennial ryegrass and barley.
Supplemental Figure S4. GenomeZipper workflow.
Supplemental Data S1. The perennial ryegrass GenomeZipper.

ACKNOWLEDGMENTS
We thank Mr. Manuel Spannagl from the Institute of Bioinformatics and Systems Biology for the integration of the perennial ryegrass GenomeZipper into the Munich Information Center for Protein Sequences plant genome database.
Received September 14, 2012; accepted November 20, 2012; published November 26, 2012.

LITERATURE CITED

The Perennial Ryegrass GenomeZipper

581

Downloaded from on October 30, 2017 - Published by www.plantphysiol.org
Copyright © 2013 American Society of Plant Biologists. All rights reserved.
Pfeifer et al.


