Targeted Genomic Mapping of a Red Seed Color Gene (R-A1) in Wheat

Vasu Kuraparthy, Shilpa Sood, and Bikram S. Gill*

Wheat Genetic and Genomic Resources Center, Dep. of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502. Received 31 Aug. 2007. *Corresponding author (bsgill@ksu.edu).

Abstract
Seed color is an important trait affecting flour yield and quality in wheat. Seed color also is either tightly linked to or pleiotropically controls seed dormancy in wheat, because most of the red-seeded wheats are tolerant to preharvest sprouting in comparison to white-seeded wheats. Seed color in hexaploid wheat is controlled by the dominant red seed color genes \( R-A1 \), \( R-B1 \), and \( R-D1 \) located in orthologous positions on chromosome arms 3AL, 3BL, and 3DL, respectively. By using wheat ESTs and synteny with rice, we identified one STS marker and one EST marker flanking \( R-A1 \) in a 4.4-cM interval by using an RIL and F\(_2\) populations of Langdon (LDN)/LDN-DIC3A (disomic substitution of \( T. turgidum \) subsp. dicoccoides chromosome 3A for 3A of LDN). Physical mapping of the \( R-A1 \) gene using tightly linked markers on a set of deletion lines specific to the long arms of group-3 chromosomes indicated that the red seed color genes are located in the distal region (less than 10% of the chromosome arm 3L), which is a high-recombination, gene-rich region in wheat. Comparative genomic analysis indicated that, except for a very minor rearrangement of gene sequences in wheat relative to rice, macrocolinearity is well conserved between the consensus distal deletion bin of wheat 3L0.80–1.00 and rice chromosome arm 1L. The \( R-A1 \) gene of wheat was targeted into a single PAC of rice using colinear flanking markers.

Cultivated wheats can be classified as red or white types based on seed color. The seed color locus (\( R \)) is one of the major agronomic loci where all three orthologous genes are functional even after 10,000 years of polyploid wheat evolution (Huang et al., 2002; Flintham and Humphrey, 1993; Flintham and Gale, 1995; Nelson et al., 1995). Although different wheat cultivars possess \( R \) genes ranging from zero to all three copies, all wild species of wheat (at all ploidy levels) possess functional \( R \) alleles, indicating a strong positive selection pressure for the red seed color in wild species and suggesting that the \( R \) loci were involved in the adaptive evolution of the wild wheats. Red-kernel wheats are usually more resistant to preharvest sprouting (PHS) than white-kernel wheats (Dyck et al., 1986; Flintham and Gale, 1995). Thus, red seed color has been used in wheat breeding programs as a marker for resistance to PHS. The association between PHS and seed color might be due either to the pleiotropic effect of the genes controlling seed color or to the tight genetic linkage between these genes with PHS. Many efforts have been made to identify quantitative trait loci (QTL).

Abbreviations: CS, Chinese Spring; ESTs, expressed sequence tags; FHB, Fusarium head blight; NSF, National Science Foundation; PCR, polymerase chain reaction; PHS, preharvest sprouting; QTL, quantitative trait loci; RILs, recombinant inbred lines; SSR, simple sequence repeat; TIGR, The Institute for Genomic Research.
controlling seed dormancy and PHS tolerance in wheat (Anderson et al., 1993; Flintham et al., 2002; Groos et al., 2002; Kato et al., 2001; Mares and Mrva, 2001; Roy et al., 1999; Zanetti et al., 2000). QTL in similar chromosomal locations were identified in various studies, indicating similar alleles controlling seed dormancy and PHS tolerance. Although not much is known about the exact relationship between seed dormancy and PHS, the influence of seed color on seed dormancy was unambiguously demonstrated in wheat (Watanabe and Ikebata, 2002; Torada and Amano, 2002; Flintham and Humphrey, 1993; Flintham, 2000).

The red seed color of wheat is relevant to polyploidy and quantitative inheritance in plants. Orthologous R genes show additive effects where the intensity of the red seed color depends on the number of functional alleles present at the R gene loci (Nilsson-Ehle, 1909; Flintham and Humphrey, 1993; Flintham and Gale, 1995; Nelson et al., 1995). This additive action of R gene expression in polyploid wheat helped solve one of the main conflicts in genetics by Nilsson-Eale in 1909. In his 'multiple factor hypothesis', Nilsson-Eale (1909) gave experimental evidence for the genetic basis of continuous phenotypic variation. This work on wheat seed color had far-reaching implications for addressing questions related to quantitative genetic variation in natural and breeding populations (Provine, 1971).

Due to the polyploid nature of wheat, the availability of genetic stocks and cultivars with varying numbers of R genes and the simple inheritance of individual R genes offer a unique opportunity to investigate the molecular genetic basis of quantitative inheritance.

Seed color in hexaploid wheat is controlled by the dominant red seed color genes R-A1, R-B1, and R-D1 located in orthologous positions on chromosome arms 3AL, 3BL, and 3DL, respectively (Sears, 1944; Allan and Vogel, 1965; Metzger and Silbaugh, 1970). Previous mapping efforts showed that the R loci were mapped in an 8 to 12 cM interval flanked by orthologous alleles of RFLP markers Xbcd131 and Xabc174 on homeologous chromosome arm 3L of wheat (Flintham and Humphrey, 1993; Flintham and Gale, 1995; Nelson et al., 1995; Nalam et al., 2006). These mapping studies were in polyploid wheat where low polymorphism hinders high-resolution mapping of target traits. Therefore, other mapping approaches should be pursued to find closely linked and informative molecular markers for R loci in wheat.

Recently the National Science Foundation (NSF) funded wheat EST project and other public and private entities have generated more than 60,0000 expressed sequence tags (ESTs) from wheat and closely related species (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html; verified 8 Jan. 2008). In the wheat NSF-EST project, more than 16,000 EST loci were mapped to specific chromosome deletion bins by using a panel of the chromosome deletion lines (Qi et al., 2004). The EST sequence and mapping data provide a valuable resource for genome analysis, identification of candidate genes for traits of interest, prediction of biological function of genes, and comparative genomic analysis. Comparative mapping experiments among wheat and other members of the Poaceae including rice, barley, rye, oats, and maize have revealed remarkable similarities in gene content and marker colinearity at the chromosome (macro) level (for review, see Devos and Gale, 2000). The genomes of distantly related cereals such as oat, rice, and maize can be divided into linkage blocks that have homology to corresponding segments of the wheat genome (Ahn et al., 1993; Van Deynze et al., 1995, 1998). Furthermore, comparative genomic analysis between wheat and rice at the sequence (micro) level indicated good levels of conservation (Yan et al., 2003; Chantret et al., 2004; Distelfeld et al., 2004). Most of the studies reported the occurrence of multiple rearrangements in gene order and content (Bennetzen, 2000; Feuillet and Keller, 2002; Li and Gill, 2002; Sorrells et al., 2003; Francki et al., 2004; Lu and Faris, 2006).

Nevertheless, the rice genome sequence is a valuable tool for map-based cloning of genes in wheat (Yan et al., 2003), and a useful source of markers for saturation and high-resolution mapping of target genes in wheat (Liu and Anderson, 2003; Distelfeld et al., 2004; Mateos-Hernandez et al., 2005). The objective of the present study was to localize the red seed color gene R-A1 to a specific genomic region in wheat, establish its most syntenous location in rice, and identify closely linked markers using wheat ESTs and synteny with rice.

Materials and Methods

Plant Material

An F2 population of 120 plants derived from the cross Triticum monococcum L. subsp. monococcum (TA4342–96)/T. monococcum subsp. aegilopoides (Link) Thell. (TA4342–95) was used for ordering the EST and STS markers and exploiting the high level of polymorphism available in diploid wheat. Eighty-three recombinant inbred lines (RILs) developed by Joppa and Williams (1988) were used for the molecular mapping of R-A1 of wheat using ESTs and STS markers that were mapped in the genomic region of the seed color gene. An F3 population of 349 plants derived from Langdon (LDN)/LDN-DIC3A (disomic substitution of T. turgidum subsp. dicoccoides

March 2008 ■ No. 1

S-38 The Plant Genome [A Supplement to Crop Science] ■ March 2008 ■ No. 1
chromosome 3A for 3A of LDN) was used to identify closely linked markers to R-A1.

Because Chinese Spring (CS) wheat possesses only one functional red seed color gene (R-D1) on chromosome 3DL (Sears, 1944; Allan and Vogel, 1965; Metzger and Silbaugh, 1970), R-D1 was physically localized with respect to CS deletion lines by using terminal deletions of chromosome 3DL of CS developed by Endo and Gill (1996). Linked molecular markers were used to further physically map the R loci using CS deletion lines. Ditolomomic lines (Sears and Sears, 1978) of group-3 chromosomes, in which a specific chromosome arm pair is missing, were used to identify fragments hybridizing to specific chromosome arms. For deletion bin mapping of the R-A1 gene, eight lines, four for 3AL and two each for 3BL and 3DL of CS, with terminal chromosomal deletions in the long arms of the group-3 chromosomes (Endo and Gill, 1996) were used. Included in the four 3AL deletion lines were TA4526-L3 (3DS-3/3AL) and TA4536-L5 (3BS-3/3AL), which have much smaller terminal deletions in the long arm of chromosome 3A (Qi et al., 2003). All plants were grown in a regular greenhouse soil mix at 20 to 24°C under supplemental sodium-vapor lighting with a 16/8 h day/night cycle.

**Grain Color Evaluation**

In the case of LDN-DIC3A F_{2,3} population, F_{2} grains were analyzed and the F_{2} lines were subsequently classified as red or white. Five to ten seeds of each line or plant were soaked in 5% (w/v) sodium hydroxide (NaOH) solution for 30 to 45 min in a 10 mL glass test tube and placed against a white background. Grain color was classified as red or white with reference to the white durum wheat cultivar LDN and the red wheat line LDN-DIC3A (Fig. 1).

**Molecular Mapping of R-A1 Gene**

A total of 18 simple sequence repeat (SSR) markers previously mapped in the distal region of chromosome arm 3AL of wheat (Roder et al., 1998; Gupta et al., 2002; Somers et al., 2004; Guyomarc’h et al., 2002) were used for mapping R-A1 in the RIL population. The PCR reactions for SSR markers were performed as described by Kurapathy et al. (2007a).

We used three RFLP probes (BCD131, ABC174, and BCD1431) that were previously reported to be linked with R loci on chromosome arm 3L of wheat. Physically mapped wheat ESTs of the CS deletion bin 3DL3-0.81-1.00, and markers developed based on synteny with rice, were used for molecular mapping in diploid and tetraploid wheat populations. Both EST and STS markers were mapped as RFLPs. DNA isolation, and Southern hybridizations were as reported in Kurapathy et al. (2007a). DNA of the parents of the diploid and polyploid mapping populations was digested with six restriction enzymes (DraI, EcoRI, EcoRV, HindIII, SacI, and XbaI) for polymorphism study.

Comparative genomic analysis using physically and genetically mapped ESTs with the rice annotated BAC/PAC sequence was used to establish the syntenic and macrocolinearity in the genomic region encompassing the R-A1 gene of wheat. Full length cDNAs or tentative consensus sequences of the mapped ESTs were extracted using The Institute for Genomic Research (TIGR) wheat gene index TaGI release 10.0 (http://compbio.dicicenter.edu/cgi-bin/tgi/gimain.pl?gudb=wheat; verified 23 Jan. 2008). These sequences were then used to search the rice genome database (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osal; verified 23 Jan. 2008) using BLASTn (Altschul et al., 1997) to identify the syntenic rice BAC/PACs (Table 1). Sequences in the target region of the rice genome were also used as queries in BLASTn (Altschul et al., 1997) searches of the wheat EST database (http://tigrblast.tigr.org/cgi/ to identify additional unmapped wheat ESTs that were potentially linked to R-A1, and to develop STS markers (Table 2, Table 3) for further mapping and genomic targeting of R-A1. Predicted rice sequences of the syntenic rice BACs were further subjected to tBLASTx searches against the wheat EST clusters using the Gramene (Ware et al., 2002) database (http://www.gramene.org/Multi/blastview; verified 8 Jan. 2008) for confirming the homology. In the BLASTn searches, a significant match was declared when there was at least 65% nucleotide identity for at least half of the query sequence, but not less than 150 bases, and with an e value of less than e-20. For tBLASTx searches, significance was declared when there was at least 40% amino acid identity over at least half of the TC or EST sequence, but no less than 150 amino acids, and an e value of less than e-11. Whenever there were several significant matches for a single predicted rice gene sequence, only the best match was reported.

Primer design, polymerase chain reaction (PCR) amplification, cloning, and probe development of the STS markers were done as described in Kurapathy et al., 2007a. DNA of the parents of the diploid and polyploid mapping populations was digested with six restriction enzymes (DraI, EcoRI, EcoRV, HindIII, SacI, and XbaI) for polymorphism study.

Comparative genomic analysis using physically and genetically mapped ESTs with the rice annotated BAC/PAC sequence was used to establish the syntenic and macrocolinearity in the genomic region encompassing the R-A1 gene of wheat. Full length cDNAs or tentative consensus sequences of the mapped ESTs were extracted using The Institute for Genomic Research (TIGR) wheat gene index TaGI release 10.0 (http://compbio.dicicenter.edu/cgi-bin/tgi/gimain.pl?gudb=wheat; verified 23 Jan. 2008). These sequences were then used to search the rice genome database (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osal; verified 23 Jan. 2008) using BLASTn (Altschul et al., 1997) to identify the syntenic rice BAC/PACs (Table 1). Sequences in the target region of the rice genome were also used as queries in BLASTn (Altschul et al., 1997) searches of the wheat EST database (http://tigrblast.tigr.org/cgi/) to identify additional unmapped wheat ESTs that were potentially linked to R-A1, and to develop STS markers (Table 2, Table 3) for further mapping and genomic targeting of R-A1. Predicted rice sequences of the syntenic rice BACs were further subjected to tBLASTx searches against the wheat EST clusters using the Gramene (Ware et al., 2002) database (http://www.gramene.org/Multi/blastview; verified 8 Jan. 2008) for confirming the homology. In the BLASTn searches, a significant match was declared when there was at least 65% nucleotide identity for at least half of the query sequence, but not less than 150 bases, and with an e value of less than e-20. For tBLASTx searches, significance was declared when there was at least 40% amino acid identity over at least half of the TC or EST sequence, but no less than 150 amino acids, and an e value of less than e-11. Whenever there were several significant matches for a single predicted rice gene sequence, only the best match was reported.

Primer design, polymerase chain reaction (PCR) amplification, cloning, and probe development of the STS markers were done as described in Kurapathy et al., 2007a. DNA of the parents of the diploid and polyploid mapping populations was digested with six restriction enzymes (DraI, EcoRI, EcoRV, HindIII, SacI, and XbaI) for polymorphism study.
et al. (2007b). The computer program Mapmaker (Lander et al., 1987) version 2.0 for Macintosh was used to calculate linkage distances using the Kosambi mapping function (Kosambi, 1944) with an LOD threshold of 3.00.

**Results**

**Wheat–Rice Comparative Mapping**

We selected 28 EST markers from the deletion bin 3AL5-0.78-1.00 harboring previously mapped flanking markers to \( R \) loci, for molecular mapping in the \( F_2 \) population of diploid wheat. The parents \( T. monococcum \) subsp. \( monococcum \) and \( T. monococcum \) subsp. \( aegilopoides \) are both red grained, and, thus, no phenotypic segregation for \( R-A1 \) was observed in the \( F_2 \) population. Of the 28 ESTs surveyed, 23 (82%) were polymorphic with one or more enzymes. Of these, 13 ESTs were mapped in the present study which gave a genetic map length of 105.7 cM (Fig. 2). Included in the map were four RFLP markers and one SSR marker that were mapped previously (Kuraparthy et al., 2007a). Only two EST markers (\( XBF474820 \) and \( XBE493800 \)) were mapped in the interval flanked by \( Xbcd131 \) and \( Xpsr1205 \) (Fig. 2), a region that encompasses the \( R \) loci.

Comparative genomic analysis of the 13 genetically mapped ESTs with the rice genome sequence showed that nine ESTs had significant homology (Table 1). All these markers were mapped proximal to the marker \( Xbcd131 \). Except for two EST markers,
Table 2. Wheat STS markers (developed based wheat–rice synteny) that were used for targeting the seed color gene.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Source</th>
<th>Forward primer (5′→3′)</th>
<th>Reverse primer (5′→3′)</th>
<th>T_a (°C)</th>
<th>Fragment size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>XSTS-SC3L2</td>
<td>TC273647†</td>
<td>AGACATTGAGCGGAG6AAA</td>
<td>TATGCTCGCTGTCTTCAGG</td>
<td>60</td>
<td>216</td>
</tr>
<tr>
<td>XSTS-SC3L3</td>
<td>TC238077</td>
<td>AATTTGCGAGGACGATTCAC</td>
<td>ACCACGCTGTCTCTGTTT</td>
<td>60</td>
<td>249</td>
</tr>
<tr>
<td>XSTS-SC3L7</td>
<td>TC237568</td>
<td>TGAAGATGCGAAAGGACAG</td>
<td>GGCTGAGTCTGTGGATTG</td>
<td>60</td>
<td>235</td>
</tr>
<tr>
<td>XSTS-SC3L8</td>
<td>TC270002</td>
<td>ACCACGGTCTGCTTCC</td>
<td>GTGAAAGCTGCCGCTCAAAT</td>
<td>60</td>
<td>171</td>
</tr>
<tr>
<td>XSTS-SC3L9</td>
<td>TC253312</td>
<td>CCGGAGAATGCGAAGGAG</td>
<td>CAGGCCCAGGGAGTTATGG</td>
<td>55</td>
<td>202</td>
</tr>
<tr>
<td>XSTS-SC3L11</td>
<td>TC244291</td>
<td>GTATTGCGAGGATCAAGCAG</td>
<td>GATTTAGACAGGGCCACAA</td>
<td>60</td>
<td>199</td>
</tr>
<tr>
<td>XSTS-SC3L13</td>
<td>TC237544</td>
<td>GAAACGGCGATTGACCAT</td>
<td>TGGGGAAGAGAAGAGTGG</td>
<td>60</td>
<td>204</td>
</tr>
<tr>
<td>XSTS-SC3L29</td>
<td>TC265912</td>
<td>AAAGAAGGGAACCGACAGA</td>
<td>GCIGCCCTCAACTTGGAC</td>
<td>60</td>
<td>165</td>
</tr>
</tbody>
</table>

†TC, tentative consensus.

Table 3. Annotated rice sequences in the syntenic rice PAC P0504E02 which spans the R-A1 gene of wheat and their wheat orthologues based on the best BLASTn and tBLASTx hits in the wheat gene indices.

<table>
<thead>
<tr>
<th>Locus identifier</th>
<th>Putative function</th>
<th>Wheat EST/TC†‡</th>
<th>e value</th>
<th>Wheat BLASTn</th>
<th>Wheat tBLASTx</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC_Os01g71050</td>
<td>VAP27-2, putative, expressed</td>
<td>TC265912†</td>
<td>1.2e-157</td>
<td>1.2e-101</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71060†</td>
<td>xylanase inhibitor, putative, expressed</td>
<td>N/A</td>
<td>5.3e-36</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71070</td>
<td>xylanase inhibitor TAXI-IV, putative, expressed</td>
<td>TC236890</td>
<td>1.1e-64</td>
<td>7.2e-63</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71080</td>
<td>xylanase inhibitor TAXI-IV, putative, expressed</td>
<td>TC235814</td>
<td>1.1e-63</td>
<td>3.5e-61</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71090</td>
<td>xylanase inhibitor, putative, expressed</td>
<td>TC236890</td>
<td>7.6e-92</td>
<td>6.1e-98</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71094</td>
<td>expressed protein</td>
<td>TC236890</td>
<td>2.9e-39</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71100</td>
<td>expressed protein</td>
<td>TC244291</td>
<td>3.3e-136</td>
<td>3.9e-63</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71106</td>
<td>NBS-LRR disease resistance protein, putative, expressed</td>
<td>CF132872</td>
<td>9.8e-53</td>
<td>2.0e-48</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71114</td>
<td>disease resistance protein RGA4, putative, expressed</td>
<td>CF132872</td>
<td>9.8e-53</td>
<td>2.0e-48</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71130</td>
<td>xylanase inhibitor TAXI-III, putative, expressed</td>
<td>BQ295496</td>
<td>3.5e-26</td>
<td>3.0e-31</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71140</td>
<td>xylanase inhibitor TAXI-IV, putative, expressed</td>
<td>BQ295496</td>
<td>7.0e-33</td>
<td>2.9e-52</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71150</td>
<td>hypothetical protein</td>
<td>N/A</td>
<td>NS</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71160</td>
<td>xylanase inhibitor TAXI-III, putative, expressed</td>
<td>BQ295496</td>
<td>2.9e-31</td>
<td>1.1e-57</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71170</td>
<td>expressed protein</td>
<td>N/A</td>
<td>NS</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71180</td>
<td>pentatricopeptide repeat protein PPR1106-17, putative</td>
<td>TC248674</td>
<td>3.4e-28</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71190</td>
<td>photosystem II reaction center W protein, putative, expressed</td>
<td>TC237544</td>
<td>3.8e-80</td>
<td>3.9e-68</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71200</td>
<td>RNA binding protein, putative, expressed</td>
<td>TC254021††</td>
<td>3.7e-232</td>
<td>1.1e-189</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71210</td>
<td>expressed protein</td>
<td>TC240116</td>
<td>1.2e-25</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71220</td>
<td>RING zinc finger protein, putative, expressed</td>
<td>CD892606</td>
<td>3.2e-86</td>
<td>1.1e-72</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71230</td>
<td>nascent polypeptide-associated complex alpha subunit-like protein 3, putative, expressed</td>
<td>TC238598</td>
<td>2.6e-80</td>
<td>1.4e-52</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71240</td>
<td>calcium-transporting ATPase 11, plasma membrane-type, putative, expressed</td>
<td>TC248924</td>
<td>4.2e-101</td>
<td>3.1e-134</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71250</td>
<td>expressed protein</td>
<td>TC248924</td>
<td>4.2e-101</td>
<td>3.1e-134</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71256</td>
<td>expressed protein</td>
<td>N/A</td>
<td>NS</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71262</td>
<td>expressed protein</td>
<td>N/A</td>
<td>NS</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71270</td>
<td>eukaryotic peptide chain release factor subunit 1-1, putative, expressed</td>
<td>TC235279</td>
<td>4.5e-177</td>
<td>2.1e-198</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71040</td>
<td>expressed protein</td>
<td>N/A</td>
<td>NS</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>LOC_Os01g71280</td>
<td>glycerol-3-phosphate dehydrogenase, putative, expressed</td>
<td>TC237660</td>
<td>1.9e-258</td>
<td>1.0e-177</td>
<td></td>
</tr>
</tbody>
</table>

†TC, Tentative Consensus.
‡Designations of ESTs (Genbank) and TCs (TIGR) as of December 2006.
††Closest EST-based STS marker placed in the linkage maps.
N/A, TC sequence not available.
NS, No significant similarity.
††Syntenic distally mapped RFLP marker Xabc174.
The Plant Genome [A Supplement to Crop Science]  ■ March 2008  ■ No. 1

Figure 2. Genetic mapping of the R-A1 gene region in diploid wheat (T. monococcum subsp. monococcum) using wheat ESTs and STS markers. The top of each map is toward the centromere and the bottom is toward the telomere. In the T. monococcum map, markers that are highly syntenic to the colinear rice genomic sequence are indicated in green. Gene sequences homologous among the two species are connected by arrows. Rice BACs/PACs used for synteny based wheat STS marker development are indicated in green. All 3AL-STS markers were developed based on syntenic rice genomic sequence.
the order of the genetically resolved ESTs is consistent with the physical order of the orthologous rice sequences on chromosome arm 1L (Fig. 2). Furthermore, the orientation of the telomeric end of the wheat chromosome 3L corresponded well with that of chromosome arm 1L of rice, where the distal region of 3L corresponds to distal region of chromosome arm 1L of rice (Fig. 2). Macrocolinearity in the region proximal to the R-A1 gene is, thus, well conserved between wheat and rice. However, no significant rice homologous sequences were found for the two EST markers (XBE474820 and XBE493800) mapping in the region flanked by Xbcd131 and Xpsr1205 (Table 1; Fig. 2). Two EST markers (XBE518446 and XBE497740) showed high homology with noncolinear rice sequences (Table 1). The RFLP markers Xbcd131 and Xpsr1205 flanking the seed color gene region also did not show any homologous sequences in the syntenic region of rice. This tentatively suggested the wheat–rice synteny at or in the immediate genomic region of the R-A1 locus was not highly conserved.

Based on the wheat–rice synteny map (Fig. 2), five rice BACs physically mapped in the region orthologous to R-A1 were selected for STS marker development (Table 2). Of the eight STS markers developed, six were polymorphic between the diploid parents, one (XSTS-SC3L11) was monomorphic, and one (XSTS-SC3L13) produced multiple bands. Out of six polymorphic STS markers, only five were mapped in the diploid wheat mapping population. All but one (XSTS-SC3L8), of the five loci showed a genetic order similar to that of the homologous sequences in the rice BAC based physical map (Fig. 2). This fact further suggested that the colinearity in this region is well conserved between wheat and rice at the macrolevel. Comparative genomic analysis of the annotated rice sequences from rice PAC P0504E02 as queries in the BLASTn searches against the wheat EST database showed that 67% of the rice sequences had corresponding homologues in wheat, tentatively suggesting the conserved colinearity at the R-A1 genomic region (Table 3).

**Molecular Genetic Mapping of R-A1 Gene in Tetraploid Wheat**

For segregation analysis and molecular mapping of the R-A1 gene, an RIL population and an F2 population derived from cross LDN/LDN-DIC3A were used. Initial molecular mapping was done using genetically and/or physically mapped SSRs (Roder et al., 1998; Gupta et al., 2002; Somers et al., 2004; Guyomarc’h et al., 2002) and the ESTs and STS markers of the *T. monococcum* map that mapped in the genomic region of the seed color gene. Out of 23 SSRs surveyed, six (26%) were polymorphic between the tetraploid parents LDN and LDN-DIC3A and were mapped in the RIL population. The SSR marker order agreed with their previous mapping positions (Fig. 3). One additional STS marker (XSTS-SC3L29) not mapped in the *T. monococcum* F2 population was mapped in the RIL population (Fig. 3). All RFLP and STS markers mapped in the same genetic positions as expected based on the *T. monococcum* map (Fig. 3, Fig. 2). The STS marker XSTS-SC3L29 cosegregated with the seed color gene R-A1 and the EST marker XBE493800, which cosegregated with STS markers XSTS-SC3L2 and XSTS-SC3L3, and mapped 0.6 cM distally to the R-A1 gene in the RIL map (Fig. 3). However, all the six linked markers, including the R-A1 gene, showed significant segregation distortion at p < 0.05, where an excess of LDN-DIC3A alleles were observed (Fig. 3).

To resolve the markers cosegregating with the R-A1 region, an F1 population of 349 plants derived from cross LDN/LDN-DIC3A was used for molecular mapping. The F1 population segregated 258 red-seed and 91 white-seed phenotypes, which was a good fit for the monogenic segregation ratio of 3:1. Molecular mapping showed that XSTS-SC3L29 mapped 1.8 cM proximally, and XBE493800 mapped 2.6 cM distally to the R-A1 gene (Fig. 3). Both linked markers also showed a clear monogenic codominant marker segregation ratio of 1:2:1 (p = 0.01) in the F2 population indicating no segregation distortion at the R-A1 locus in the F2 population.

Previous reports indicated that RFLP marker Xabc174 mapped close to the R loci on the distal side in wheat (Flintham and Humphrey, 1993; Flintham and Gale, 1995; Nelson et al., 1995). BLASTn searching of the Xabc174 marker sequence against the rice BAC/PAC sequence data suggested that Xabc174, which codes for a ribonucleoprotein, showed very high homology (e-value: 3.7e-232) to an annotated marker sequence against the rice ESTDB database. The Xabc174 marker was mapped in the F2 population indicating no segregation distortion at the R-A1 locus in the F2 population.

Physical Mapping of the R-A1 Gene Using Deletion Lines

The R loci were previously mapped on the distal region of the long arm of the homeologous group-3 chromosomes of wheat (Flintham and Humphrey, 1993; Flintham and Gale, 1995; Nelson et al., 1995). Because CS wheat possesses only one functional gene (*R-DI*) on chromosome 3DL, evaluation of the CS 3DL deletion lines for seed–color phenotype...
indicated that the line with deletion 3DL3-0.81-1.00 was white seeded, suggesting that R loci can be mapped in the distal 20% of the chromosome arm 3L of wheat. Physical mapping of the R-A1 gene using closely linked flanking marker probes STS-SC3L29 and BE493800 on a set of deletion lines also revealed that the R-A1 gene maps in chromosome deletion bins 3AL-5 (FL 0.78–1.0), 3BL-7 (FL 0.63–1.00), and 3DL-3 (FL 0.81–1.00) of chromosomes 3A, 3B, and 3D of wheat, respectively (Fig. 3, Fig. 4). In a consensus physical map of the group-3 chromosomes of wheat (Delaney et al., 1995), the R-A1 gene is positioned in the distal 20% of the long arms. Furthermore, we used two new deletion lines of 3AL, TA4526-L3 (3DS-3/3AL) and TA4536-L5 (5BS-5/3AL), with smaller deletions (less than 10% of the distal region of 3AL) for higher resolution physical mapping. Both lines have the same new terminal deletion because identical sets of 3AL-specific EST fragments were missing in both lines (Qi et al., 2003). Markers XSTS-SC3L29 and XBE493800 showed diagnostic polymorphism in these two new deletion lines (Fig. 4), indicating that the R-A1 gene was actually located in the distal 10% of the long arm of chromosome 3A (Fig. 3).

**Discussion**

In the present study, the red seed color gene R-A1 of wheat was mapped within an interval of 4.4 cM flanked by one EST and one STS marker, which were developed based on synteny with rice. A highly polymorphic mapping population of the diploid wheat T. monococcum was used to resolve the wheat ESTs and identify markers linked at the R gene region. The red seed color gene was genomically targeted with respect to a syntenic rice genomic sequence and physically mapped in the distal high-recombination gene-rich region of chromosome arm 3L of wheat.

Since the initial molecular mapping of R genes on the long arms of group-3 homeologous chromosomes (Flintham and Humphrey, 1993; Flintham and Gale, 1995; Nelson et al., 1995; Nalam et al., 2006), little progress has been made for further mapping and genomic analysis. Major impediments to the limited molecular mapping of R genes have been segregation distortion (Nalam et al., 2006) and lack of polymorphism in the mapping populations used (M.E. Sorrells, personal communication, 2005; Flintham and Gale, 1995; Nalam et al., 2006). Using an F2 mapping population derived from highly polymorphic, diploid A-genome parental species, we overcame the problem of limited polymorphism available in polyploid wheat. Also, because of the lower ploidy level and less complex DNA hybridization patterns, a greater number of codominant markers with a higher degree of confidence were mapped in T. monococcum than would have been possible in polyploid wheat. Although the R-A1 gene did not segregate in the diploid wheat mapping population used, a high level of polymorphism facilitated the development of a dense genetic map using EST and STS markers, which, in turn, facilitated the identification of markers mapping in the genomic region of the R-A1 gene (Fig. 2). Newly added markers in the region of the seed color gene were then used to map R-A1 in segregating populations of tetraploid durum wheat. A high level of polymorphism and the diploid nature of T. monococcum and Ae. tauschii were also exploited for map-based cloning of agronomic genes Vrn1, Vrn2, Vrn3, and Lr21 (Yan et al., 2003, 2004, 2006, Huang et al., 2003) and for genome analysis in wheat (Chantret et al., 2004, 2005; Li et al., 2004). In the present study, the problem of segregation distortion in the R-A1 region of wheat in RILs was overcome by using an F2 population of a cross LDN/LDN-DIC3A. Both the R
gene and the linked markers showed the expected segregation ratios in this population. This suggests that the biallelic nature of the F2 population development had a masking effect on segregation distortion in the R-A1 region. Segregation distortion due to preferential transmission of male gametes in RIL populations developed through the use of monosomics as female parents in F1 hybrids has frequently been observed in wheat (Joppa, 1993, Kumar et al., 2007).

Since the discovery of synteny in the grass family, the rich genetic information of the rice genomic sequence has been exploited to map genes in other cereal crops. RFLP markers from the corresponding region of the rice genome were used directly to map genes of other cereal crops (Van Deynze et al., 1998), or DNA of probes developed from large-insert clones of rice were used for mapping (Kilian et al., 1997). Recently, sequence-based comparative mapping has become possible because of the rapid progress of rice genome sequencing and cereal EST projects. Although complexity in macro- as well as microcolinearity between wheat and rice at different orthologous positions has been reported (Bennetzen, 2000; Feuillet and Keller, 2002; Li and Gill, 2002; Liu et al., 2006, Sorrells et al., 2003; Francki et al., 2004; Lu and Faris, 2006), wheat–rice synteny has been very helpful in genomic targeting and marker enrichment of the adult-plant leaf rust resistance gene Lr46 (Mateos-Hernandez et al., 2005), the QTL Qfhns.ndsu-3BS for Fusarium head blight (FHB) resistance (Liu and Anderson, 2003; Liu et al., 2006) and grain protein content (GPC) (Distelfeld et al., 2004) genes in wheat. Conserved wheat–rice synteny was also used to clone the vernalization VRN1 gene using map-based cloning (Yan et al., 2003). Mapping of the R-A1 gene to a specific genomic location in wheat chromosome arm 3AL and its genomic targeting to a specific rice PAC P0504E02 suggested that the physically mapped EST sequences of wheat and wheat–rice synteny was very helpful in molecular mapping of R-A1. Our data confirm the synteny reported for 3AL of wheat and 1L of rice (La Rota and Sorrells, 2004) and the effectiveness of the rice genome as a base for targeting a gene of interest.

The homologous group-3 chromosomes have been shown to be the best conserved of all the homologous groups in wheat compared to rice chromosomes (Sorrells et al., 2003; La Rota and Sorrells, 2004). Our results agree, because we observed good colinearity between wheat and rice at the genomic region spanning the R-A1 gene of wheat (Fig. 2, Table 1). Comparative genomic analysis of annotated rice sequences from rice PAC P0504E02 with the TIGR wheat gene index TaGI showed that 67% of the rice sequences had corresponding homologs in wheat, tentatively suggesting the conserved colinearity at the R-A1 genomic region (Table 3). The absence of sequence homology with the colinear rice sequence for few of the mapped wheat ESTs and one of the flanking markers of the R-A1 gene in the genetic map (Fig. 2) suggests that either these wheat EST sequences were translocated in wheat from nonsyntenic locations, or they might have substantially changed since the divergence of wheat and rice. The homologous rice sequences might also have been deleted after wheat–rice divergence. Although recombination has been implicated in the sequence divergence between rice and wheat (See et al., 2006), illegitimate recombination has also been shown to be involved in the deletion of gene sequences in the distal regions of wheat (Chantret et al., 2005).

The pigment contributing to wheat seed color (phlobaphene) is produced through the flavonoid synthesis pathway (Miyamoto and Everson, 1958; Grotewold et al., 1994), in contrast to the red pigment in rice grains (proanthocyanidin) synthesized through the anthocyanin pathway (Winkel-Shirley, 2001; Oki et al., 2002). Cloning and characterization of different seed or grain color genes suggested that most of them are transcriptional factors, for example the Myb transcription factor of rice (Sweeney et al., 2006), TT2 zinc finger proteins, and anthocyanin-less2 of Petunia and Arabidopsis (Kubo et al., 1999). In the syntenic region of rice encompassing the
R-A1 gene, no candidate gene coding for the above transcription factors was found, suggesting that either the R-A1 gene of wheat codes for a novel protein, or that there have been microrearrangements at the sequence level in the orthologous region of R-A1. This further suggests that although wheat–rice synteny was pivotal in marker enrichment through targeted mapping of R-A1 gene, resources for chromosome walking or positional cloning have to be developed in wheat for molecular cloning of the R-A1 gene.

Physical mapping of the R loci using the closely linked flanking marker probes STS-SC3L29 and BE493800 on a set of CS aneuploid stocks, including two new deletion lines with terminal chromosome deletions, revealed that the R loci can be mapped in the distal 10% of the long arms of group-3 chromosomes of wheat (Fig. 3, Fig. 4). Deletion-bin based physical mapping of the R genes using the CS deletion lines is important to determine the genomic location of the R genes on chromosome 3A with respect to recombination and gene space. Comparing physical maps with recombination-based maps led to the discovery that gene density and recombination at the distal regions of the wheat chromosomes is very high (Werner et al., 1992; Gill et al., 1996; Akhunov et al., 2003) where gene density in such regions is comparable to that of rice (Feuillet and Keller, 1999). The same trend of higher recombination in the distal 20% of the long arms of group-3 chromosomes also was demonstrated unequivocally by Delaney et al. (1995). The gene density of the orthologous wheat R-A1 gene region in rice is also very high. Twenty-seven predicted genes are present in the 137.45 kb rice PAC P0504E02 that spans the seed color gene of wheat (Table 3). The overall gene density of this BAC region is one gene for every 5 kb. Physical mapping of R-A1 in less than 10% of the distal region of chromosome 3L suggests that R genes are physically located in the high-recombination gene-rich regions of wheat. These regions of wheat are suitable for the map-based isolation of genes. Thus, physical mapping of R loci in such high recombination regions of the genome could facilitate map-based cloning of R genes.

Molecular characterization of seed color genes could lead to a better understanding of gene evolution, expression with respect to polyploidy, domestication, species adaptive evolution, and quantitative inheritance of agronomic genes. Furthermore, identification and molecular characterization of genes involved in seed color and dormancy is an essential prerequisite for elucidating the molecular mechanism of grain development and seed dormancy for cereal crop improvement. Genetic mapping and genomic targeting of the red seed color gene reported here provides a starting point for the molecular dissection of these agronomically important domestication traits in wheat.

Acknowledgments
This research was supported by the Kansas Wheat Commission and a special USDA grant to the Wheat Genetic and Genomics Resources Center. We thank Dr. Justin Faris for supplying the Langdon durum wheat Recombinant Inbred Line population. We also thank Drs. Allan Fritz, Justin Faris and Scot Hulbert for reviewing the manuscript. We extend a special note of thanks to Duane Wilson and Jon Raupp for their excellent technical assistance. This paper is contribution number 07-284-J from the Kansas Agricultural Experimentation Station, Kansas State University, Manhattan, Kansas.

References

Kurapathy et al.: Seed Color of Wheat


