The Evolution of CONSTANS-Like Gene Families in Barley, Rice, and Arabidopsis

Simon Griffiths, Roy P. Dunford, George Coupland, and David A. Laurie*

John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom

The CO (CONSTANS) gene of Arabidopsis has an important role in the regulation of flowering by photoperiod. CO is part of a gene family with 17 members that are subdivided into three classes, termed Group I to III here. All members of the family have a CCT (CO, CO-like, TOC1) domain near the carboxy terminus. Group I genes, which include CO, have two zinc finger B-boxes near the amino terminus. Group II genes have one B-box, and Group III genes have one B-box and a second diverged zinc finger. Analysis of rice (Oryza sativa) genomic sequence identified 16 genes (OsA–OsP) that were also divided into these three groups, showing that their evolution predates monocot/dicot divergence. Eight Group I genes (HvCO1–HvCO8) were isolated from barley (Hordeum vulgare), of which two (HvCO1 and HvCO2) were highly CO-like. HvCO3 and its rice counterpart (OsB) had one B-box that was distantly related to Group II genes and was probably derived by internal deletion of a two B-box Group I gene. Sequence homology and comparative mapping showed that HvCO1 was the counterpart of OsA (Heading date 1), a major determinant of photoperiod sensitivity in rice. Major genes determining photoperiodic response have been mapped in barley and wheat (Triticum aestivum), but none corresponded to CO-like genes. Thus, selection for variation in photoperiodic response has affected different genes in rice and temperate cereals. The peptides of HvCO1, HvCO2 (barley), and Hd1 (rice) show significant structural differences from CO, particularly amino acid changes that are predicted to abolish B-box2 function, suggesting an evolutionary trend towards a one-B-box structure in the most CO-like cereal genes.

The control of flowering by photoperiod is an important adaptive characteristic in plants. Studies of the model dicot Arabidopsis have shown that the CO (CONSTANS) gene, isolated by Putterill et al. (1995), has an important role in the photoperiod pathway, which is one of four regulatory pathways controlling the timing of flowering (for review, see Mouradov et al., 2002; Simpson and Dean, 2002). CO acts between the circadian clock and genes controlling meristem identity (Samach et al., 2000; Suárez-López et al., 2001). In Arabidopsis, CO belongs to a family of 17 putative transcription factors defined by two conserved domains (Putterill et al., 1995; Robson et al., 2001). The first is a zinc finger region near the amino terminus that resembles B-boxes, which regulate protein-protein interactions in several animal transcription factors (Borden, 1998; Torok and Elkin, 2000). The second is a region of 43 amino acids near the carboxy terminus termed the CCT (CO, CO-like, TOC1) domain (Strayer et al., 2000; Robson et al., 2001). Studies using green fluorescent protein fusions show that the CCT domain is involved in nuclear localization of the CO protein but must have an additional role because the late-flowering co-7 mutant, which has an altered CCT domain, correctly localizes the protein (Robson et al., 2001).

Previous analysis of CO-like genes in Arabidopsis showed that the family is subdivided into three broad groupings (Robson et al., 2001). The first comprised CO and COL1 to COL5 (two B-box genes), the second comprised COL6 to COL8 and COL16 (one B-box genes), and the third comprised COL9 to COL15 (one CO-like B-box and one more diverged zinc finger domain). In this paper, these are referred to as Group I to III genes, respectively. The Group I genes had additional conserved regions, including a distinctive motif of six amino acids (consensus G-I-V-V-P-S/T-F) at the carboxy terminus of the predicted peptide.

CO-like genes have been identified in several dicots and in two cases (Brassica napus BnCOa1, Robert et al., 1998; and Phalaris nil PnCO, Liu et al., 2001) have been shown to complement a co mutant in Arabidopsis, demonstrating functional equivalence. The Hd1 (Heading date 1) gene of rice (Oryza sativa) is also homologous to CO (Yano et al., 2000). Conservation between short-day (SD) plants (rice and P. nil) and long-day (LD) plants (Arabidopsis and B. napus) suggests that CO is involved in a conserved pathway regulating flowering in response to inductive day length. Grafting experiments in dicots have shown that leaves from SD or LD plants grown under inductive conditions can promote flowering when grafted to LD or SD plants, respectively, growing under noninductive conditions (Zeevaart, 1976). Not
all such grafts are effective, but nevertheless these experiments also suggest common components in the response to inductive day length.

Conservation between rice and Arabidopsis suggests that CO-like genes are likely to be involved in flowering time control in other cereals such as barley (Hordeum vulgare), which, like Arabidopsis, is a quantitative LD plant. In this paper, we investigated the structure of the CO-like gene family in rice using previously published data (Song et al., 1998; Yano et al., 2000) and genomic sequence. This information was used to isolate Group I CO-like genes from barley. This allowed us to compare the evolution of CO-like genes in rice and Arabidopsis and the evolution of Group I genes in these species and barley. Genetic mapping of genes in barley determined their relationship to genes in rice and to previously mapped photoperiod response loci.

RESULTS

Identification of CONSTANS-Like Genes in Rice Genomic Sequence

Searches of rice genomic sequence using CO and COL1 to COL16 peptides identified a minimum of 16 genes, designated OsA to OsP, including four CO-like genes previously described (Hd1 [OsA], Yano et al., 2000; and S12569 [OsB], S3574 [OsN], and C60190 [OsP]; Song et al., 1998; Table I). Preliminary analysis of the rice genes showed that they could be divided into the Group I to III subfamilies seen in Arabidopsis and that OsA (Hd1) and OsB (S12569) were Group I genes that were the most CO like overall. These two rice genes also had the conserved carboxy-terminal motif seen in other Group I genes.

Isolation of CONSTANS-Like Genes from Barley

Barley CO-like sequences were first isolated using a 120-bp subclone of OsA (Hd1) (Yano et al., 2000), including the CCT domain and a 142-bp subclone of OsB (S12569) (Song et al., 1998), also containing the CCT domain. The CCT domain probes were used to screen a barley cv Igri lambda phage genomic library (Stratagene, Cambridge, UK) and a barley cv Morex bacterial artificial chromosome (BAC) library (Yu et al., 2000). Genomic library screens were used because the CO transcript is rare in Arabidopsis (Suárez-López et al., 2001), and orthologs might behave similarly in barley, making them difficult to isolate from cDNA libraries. In cases where Morex BAC clones were used, the corresponding gene from Igri was amplified by PCR from genomic DNA and sequenced. Sequences were obtained from Igri because it has an LD response (Laurie et al., 1995; Decousset et al., 2000) and, therefore, should contain functional alleles of any CO-like genes involved in the control of flowering by photoperiod.

The CCT domain probes from OsA (Hd1) and OsB (S12569) each identified two barley genes (four genes in total) in the preliminary screen. CCT domains from the barley genes were then used to test cross hybridization between the clones and, where appropriate, to rescreen the libraries. The CCT domains of the two barley genes detected by OsA (Hd1) gave no additional genes in further library screens, whereas the CCT domain of HvCO4 detected four additional genes. Thus, genomic library screens detected eight genes.

Nucleotide and predicted peptide searches of approximately 420,000 wheat (Triticum aestivum), 314,000 barley, and 8,900 rye (Secale cereale) ESTs available in November 2002 from GenBank (http://www.ncbi.nlm.nih.gov) gave no highly homologous matches outside the B-box or CCT domains to OsA (Hd1) or the two barley genes detected by the OsA (Hd1) CCT domain probe. However, several other CO-like genes were detected. These were highly homologous (>95% nucleotide identity) to sequences from the genomic clones except for one gene represented by six barley ESTs that, in combination, included part of the middle region, the CCT domain, and the COOH region. The combination of approaches therefore identified nine barley genes in all.

The Exon/Intron Structure of CONSTANS-Like Genes

The Arabidopsis CO, COL1 to COL8, and COL16 genes (Group I and II) have a single intron located between the B-box and CCT domains. COL9 to COL15 (Group III) have a different structure with three introns, two of which are between the B-box and CCT domains and the third within the CCT domain. For the barley genes, alignment to CO and related genes from Arabidopsis (Robson et al., 2001) and rice (Song et al., 1998; Yano et al., 2000) showed that the barley genes had one or two B-boxes, a CCT domain, and, for HvCO1 to HvCO6, one intron in a similar position to the single intron of CO (Fig. 1). HvC07 was identified only from cDNA sequences, HvC08 appeared to lack an intron, and HvC09 had no significant homology other than the CCT domain, and no intron was recognized. The barley genes were numbered based on the overall homology of their predicted peptides to CO (Arabidopsis) and Hd1 (rice). This relied primarily on the B-box and the CCT and COOH regions because the middle regions (between B-box2 and the CCT domain) were the least well conserved.

For HvCO1, the intron position was confirmed by sequencing reverse transcription-PCR products. The intron position of HvCO2 was then predicted by alignment. The intron of HvCO1 contained a 147-bp sequence with the characteristics of a Stowaway element (Bureau and Wessler, 1994). For HvCO3 to HvCO6, the intron position was identified by alignment with barley, wheat, and rye ESTs. Intron posi-
izations were confirmed by sequencing candidate full-length barley cDNA clones identified by EST searches. One wheat EST (BQ171773) matched HvCO9, but the clone was truncated and did not extend to candidate intron sites.

Spliced and unspliced ESTs were found for HvCO3 to HvCO6, but HvCO3 was the only gene for which alternatively spliced transcripts were found. The Igri gene was predicted to give a peptide similar to rice S12569 (HvCO3a in Fig. 1), and a matching transcript was found in four barley ESTs, including one fully sequence cDNA (Table I) and a rye EST (BE704660). Three barley ESTs, including a fully sequenced cDNA (Table I), used the same 3′ splice site and an alternative 5′ GT splice site, shortening the first exon (HvCO3b in Fig. 1). This variant was also seen in an Hordeum spontaneum EST (AV836099). The shorter transcript had a frame shift that would produce a truncated peptide lacking the CCT and COOH regions. Although the sequences flanking the intron

<p>| Table 1. Sequence accessions of CONSTANS and related Arabidopsis, rice, and barley genes |
|-----------------------------|-----------------------------|-----------------------------|</p>
<table>
<thead>
<tr>
<th>Gene</th>
<th>Genomic</th>
<th>cDNA/Expressed Sequence Tag (EST)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>X94937</td>
<td>NM_121589</td>
</tr>
<tr>
<td>COL1</td>
<td>Y10555</td>
<td>NM_121590</td>
</tr>
<tr>
<td>COL2</td>
<td>L81120</td>
<td>NM_111105</td>
</tr>
<tr>
<td>COL3</td>
<td>AC006585/gb AC007266</td>
<td>NM_128038</td>
</tr>
<tr>
<td>COL4</td>
<td>AF069716</td>
<td>NM_122402</td>
</tr>
<tr>
<td>COL5</td>
<td>AB018118</td>
<td>AY114006</td>
</tr>
<tr>
<td>COL6</td>
<td>AC011915</td>
<td>AY081541</td>
</tr>
<tr>
<td>COL7</td>
<td>AC016662</td>
<td>NM_106047</td>
</tr>
<tr>
<td>COL8</td>
<td>AC016041</td>
<td>NM_103803</td>
</tr>
<tr>
<td>COL9</td>
<td>AC009176</td>
<td>NM_111644</td>
</tr>
<tr>
<td>COL10</td>
<td>AB023039</td>
<td>AB023039</td>
</tr>
<tr>
<td>COL11</td>
<td>Z97338</td>
<td>NM_117613</td>
</tr>
<tr>
<td>COL12</td>
<td>AP000739</td>
<td>NM_113084</td>
</tr>
<tr>
<td>COL13</td>
<td>AC006072</td>
<td>NM_130356</td>
</tr>
<tr>
<td>COL14</td>
<td>AC002332</td>
<td>NM_128910</td>
</tr>
<tr>
<td>COL15</td>
<td>AC069471</td>
<td>NM_102570</td>
</tr>
<tr>
<td>COL16</td>
<td>AC079281</td>
<td>NM_102355</td>
</tr>
<tr>
<td>B. napus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BnCOa1</td>
<td>AF016009</td>
<td>–</td>
</tr>
<tr>
<td>P. nil</td>
<td></td>
<td>AF000700</td>
</tr>
<tr>
<td>Barley</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HvCO1</td>
<td>Igri, AF490467; Morex, AF490468</td>
<td>–</td>
</tr>
<tr>
<td>HvCO2</td>
<td>Igri, AF490469; Morex, AF490470</td>
<td>–</td>
</tr>
<tr>
<td>HvCO3</td>
<td>Igri, AF490473</td>
<td>HvCO3a: CI16155, AF490471; HvCO3b: Morex, AF490472</td>
</tr>
<tr>
<td>HvCO4</td>
<td>Igri, AF490474</td>
<td>Morex, AF490475</td>
</tr>
<tr>
<td>HvCO5</td>
<td>Igri, AY082958</td>
<td>Morex, AY082959</td>
</tr>
<tr>
<td>HvCO6</td>
<td>Igri, AY082960</td>
<td>Morex, AY082961, AY082962</td>
</tr>
<tr>
<td>HvCO7</td>
<td>–</td>
<td>CI16151, AY082963; Morex, EST BF629458</td>
</tr>
<tr>
<td>HvCO8</td>
<td>Igri, AY082964</td>
<td>Barke, ESTs AL503821, AL501239</td>
</tr>
<tr>
<td>HvCO9</td>
<td>Igri, AY082965</td>
<td>–</td>
</tr>
<tr>
<td>Rice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OsA (Hd1)</td>
<td>Ginzouzou, AB041840; Nipponbare, AB041837; Nipponbare, AP030044; 93-11, AAAA01018018</td>
<td>–</td>
</tr>
<tr>
<td>OsB (S12569)</td>
<td>93-11, AAAA01000110</td>
<td>Nipponbare, AB001887</td>
</tr>
<tr>
<td>OsC</td>
<td>93-11, AAAA01001728; Nipponbare, AL606627</td>
<td>–</td>
</tr>
<tr>
<td>OsD</td>
<td>93-11, AAAA01000818</td>
<td>–</td>
</tr>
<tr>
<td>OsE</td>
<td>93-11, AAAA01015873; Nipponbare, AP003579</td>
<td>Nipponbare, AU222658</td>
</tr>
<tr>
<td>OsF</td>
<td>93-11, AAAA01022688</td>
<td>–</td>
</tr>
<tr>
<td>OsG</td>
<td>93-11, AAAA01008321</td>
<td>–</td>
</tr>
<tr>
<td>OsH</td>
<td>93-11, AAAA01000838; Nipponbare, AC079874</td>
<td>–</td>
</tr>
<tr>
<td>OsI</td>
<td>93-11, AAAA01011539; Nipponbare, AP005307</td>
<td>–</td>
</tr>
<tr>
<td>OsJ</td>
<td>93-11, AAAA01003850</td>
<td>–</td>
</tr>
<tr>
<td>OsK</td>
<td>93-11, AAAA01010759</td>
<td>–</td>
</tr>
<tr>
<td>OsL</td>
<td>93-11, AAAA01023497; Nipponbare, AP004993</td>
<td>–</td>
</tr>
<tr>
<td>OsM</td>
<td>93-11, AAAA01003850</td>
<td>–</td>
</tr>
<tr>
<td>OsN (S3574)</td>
<td>93-11, AAAA01003074</td>
<td>Nipponbare, AB001888</td>
</tr>
<tr>
<td>OsO</td>
<td>93-11, AAAA01002984</td>
<td>–</td>
</tr>
<tr>
<td>OsP (C60910)</td>
<td>93-11, AAAA01015863</td>
<td>Nipponbare, AB001882</td>
</tr>
</tbody>
</table>
were not highly conserved, the positions of conserved N and H residues relative to other CO family genes suggests that S12569 and HvCO3 have evolved by insertion of 30 bp at the end of exon1 and that the HvCO3b variant may use the ancestral splice site (Fig. 1).

In rice, OsA to OsL resembled the barley genes in having a single intron, whereas OsM to OsP resembled the Arabidopsis COL9 to COL15 group in having three introns and a second zinc finger domain that was diverged from the CO-like B-box. Intron positions were confirmed for OsA (Hd1), OsB (S12569), OsN (S3574), and OsP (C60910), for which full-length cDNA sequences were available (Table I).

Sequence Alignment of CONSTANS-Like Genes

Alignments of the predicted peptides of CO, OsA (Hd1), and the most CO-like barley genes (HvCO1 and HvCO2) are shown in Figure 2 together with B-box (Fig. 2a) and CCT domain and COOH regions (Fig. 2c) for all the barley genes that were identified and their candidate rice orthologs. The alignment shows that the barley HvCO1 to HvCO8 peptides had consensus CO-like amino acid residues at the carboxy terminus. Variation within the B-box and CCT domains suggested that the CO-like genes could be further subdivided. For example, distinctive amino acid residues in the B-boxes (boxed in Fig. 2a) grouped Arabidopsis COL3 to COL5 peptides with HvCO4 to HvCO7 and OsC to OsF.

Conserved Motifs in the Middle Region

The middle region of the CO gene family was the most diverged, but comparison of cereal and dicot genes identified four small regions of conservation (boxed in Fig. 2b) that helped define the most CO-like genes. The E-X-S-W-L-L (box 1; Fig. 2b), L-V-D/G-Y (box 2; Fig. 2b), and G-X-D/E-X-I/V-V-P (box 3; Fig. 2b) motifs were in exon 1. Other genes in addition to those shown in box3 of Figure 2b also had V-P residues in similar positions, but due to a lack of conserved flanking residues, it was not clear whether these were the same motif. A group of the most CO-like genes also had a consensus S-X-E-X3-V-P motif close to the start of exon 2 (box 4 in Fig. 2b).

Phylogenetic Relationship of the Arabidopsis, Rice, and Barley Genes

To examine the relationship between genes in more detail, their nucleotide and predicted peptide sequences were used to determine genetic distances and to construct phylogenetic trees. Because the middle regions of the genes were the most diverged, they could not be aligned with confidence. Therefore, neighbor-joining (NJ) and maximum parsimony (MP) trees were constructed using B-box and CCT domain sequences following the alignments shown in Figure 2.

First, all genes were compared using the predicted peptide or nucleotide sequence of the CCT domain alone. This consistently grouped the genes into four principal clusters (Fig. 3a). Group III genes com-
Evolution of CONSTANS-Like Gene Families

Figure 2. Alignment of predicted peptides of CONSTANS and related genes from Arabidopsis, barley, and rice. 

(a) Amino termini and B-box domains. Conserved C and H residues and consensus spacing (Xn) defining B-box domains (Borden, 1998; Robson et al., 2001) are shown below. Residues altered in co-2 to co-7 mutant alleles or deleted in co-1 (Robson et al., 2001) are shown above. Boxed residues distinguish COL3- to COL5-like peptides from the most CO-like group. 

(b) Middle region. Alignments of whole sequences are shown for CO, Hд1, HвCO1, and HвCO2. Small conserved motifs found in other CO-like genes are shown in boxes (see text for details). The underlined region is predicted to form coiled structures in Hд1 and HвCO1 but not in CO. 

(c) CCT and COOH domains. Residues affected in co mutant alleles (Robson et al., 2001) are shown above. GenBank accession numbers for the sequences used in this figure are listed in Table I. Barley sequences are the Igri allele in all cases except for HвCO7, where sequence was from CI16151.
prised Arabidopsis and rice genes with two zinc finger domains, the second of which was diverged from the CO-type B-box. Group II genes comprised Arabidopsis and rice genes with a single B-box. Group I comprised the most CO-like genes and included the barley HvCO1 to HvCO8 genes. There were subdivisions within Group I, but these had low bootstrap values.

HvCO9 had the most diverged CCT domain of the barley genes, and the phylogenetic tree (Fig. 3a) placed HvCO9 and two related rice genes (OsH and OsI) between the Groups II and III genes. OsH was part of a sequenced and annotated BAC clone (Table I) and the predicted peptide (AAL79780) had no B-box. Searches of Arabidopsis genomic sequence using the predicted peptide showed no significant homology outside the CCT domain. Thus, HvCO9 and the related rice genes were identified as a distinct forth group that had no counterpart in Arabidopsis (Fig. 3a and c).
When genes were analyzed using concatenated B-box and CCT domain sequences, a similar overall pattern was observed, but the Group I genes could be divided into subgroups with higher bootstrap values. This was consistent in NJ and MP methods using nucleotide or predicted peptide sequences. There were minor variations at some nodes with low bootstrap values, but this did not affect the gene groups defined in Figure 3.

Group Ia comprised the most CO-like genes, including the B. napus and P. nil genes previously shown to complement co mutations in Arabidopsis, rice OsA (Hd1), and barley HvCO1 and HvCO2. Within the COOH terminus, Hd1, HvCO1, HvCO2, and two maize (Zea mays) Hdl-like peptides translated from ESTs (GenBank accession nos. BE051702 and BE640554) had T and W residues (Fig. 2c), suggesting that this may be a useful distinguishing feature for the most CO-like cereal genes.

Group I also contained the Arabidopsis COL3 to COL5 genes, rice OsC to OsF, and HvCO4 to HvCO7. This grouping reflected the distinctive amino acid residues in the B-boxes (the W, V, and VT residues marked in Fig. 2a). The phylogenetic tree (Fig. 3b) separated HvCO4/HvCO5 from HvCO6/HvCO7, and the latter also lacked the first domain of the middle region (box 1 in Fig. 2b). Therefore, these genes were designated as Groups Ic and Id, respectively.

Group I contained four genes (HvCO3, OsB [S12569], HvCO8, and OsG) that had a single B-box. This placement was consistent in all phylogenetic analyses of nucleotide or peptide sequences, and these genes were never placed with the one B-box genes of Group II. Nucleotide and peptide alignment of HvCO3/OsB gave the best fit to other Group I genes when the first part of the B-box was aligned to B-box1 and the remainder to B-box2 (Fig. 2a). The alignment of HvCO8/OsG was more ambiguous, but the alignment shown in Figure 2a was marginally the best. In addition to the carboxy-terminal motif, HvCO3/OsB and HvCO8/OsG had motifs in the middle region that placed them with the Group I genes (Figs. 2b and 3c). These results suggest that these genes originated by internal deletion of ancestral two-B-box genes. The COOH region of HvCO3/OsB was similar to HvCO1 and OsA (Hd1) (Fig. 2c), but the absence of conserved motifs 2 and 3 with the middle region suggested it was more closely related to Group Ic genes (Figs. 2b and 3c). Thus, the exact relationship of HvCO3/OsB to the Ia and Ic Groups was ambiguous. Apart from their B-box structure, HvCO3 and HvCO8 were not closely related, and HvCO8 also lacked middle region motif 1. Therefore, HvCO3/OsB was placed in the subgroup Ib, and HvCO8/OsG was placed in subgroup Ie.

B-Box Structure of the Group Ia Genes

Although HvCO1 and HvCO2 were the most CO-like overall (Fig. 2), and HvCO1 was the most similar to rice Hd1 in the middle region (generally the most diverged part of the genes; Fig. 2b) and the COOH region. However, B-box2 was not well conserved. The HvCO1 peptides of Igr1 and Morex were identical and lacked three highly conserved C residues (Fig. 2a) that would be predicted to abolish B-box2 function. There was also an A-to-V change affecting the residue that is altered in the Arabidopsis co-4 allele and is conserved in other genes, suggesting essential function (Fig. 2a). Interestingly, the functional alleles of Hd1 sequenced by Yano et al. (2000) also showed changes that would be expected to abolish B-box2 function. The Ginbouzou allele showed an H-to-Y change involving an H residue thought to be essential for B-box function. This change is found in Arabidopsis co-3, the most severely late-flowering mutant allele (Robson et al., 2001; Fig. 2a). The Nipponbare allele retained the H residue but had a deletion of 12 amino acids (nine from B-box2 and three from the adjacent middle region) that included part of the deletion in the Arabidopsis co-1 mutant allele. Nonfunctional alleles had either an H (Kasalath) or Y (HS66 and HS110) residue and additional insertion or deletion changes likely to result in nonfunctional proteins (Yano et al., 2000). The 93-11 sequence (not shown) had an H residue, predicting two functional B-boxes, but also had a 4-bp deletion in the CCT domain, giving a predicted peptide lacking the P and R residues affected in the Arabidopsis co-5 and co-7 mutant alleles and lacking the conserved residues at the carboxy terminus.

B-box2 from HvCO2 was more conserved but had two non-consensus amino acids in the position of the co-1 deletion (I-A-Q in place of L-A-R, Fig. 2a). Together, these observations suggest that B-box2 is less important to the function of the most CO-like genes in cereals. B-box1, in contrast, was well conserved, suggesting that a B-box is required for gene function in barley and rice.

A further difference between cereal CO-like peptides and CO itself was suggested by secondary structure analysis using facilities at the ExPASy Molecular Biology Server (http://www.expasy.ch/). This showed that HvCO1 was strongly predicted to form a coiled region in the central part of the peptide (underlined region in Fig. 2b; probabilities of 0.943, 0.994, and 0.995 for windows of 14, 21, or 28 amino acids, respectively). This was also strongly predicted for Hd1 (probabilities 0.925, 0.994, and 0.424, respectively), weakly predicted for HvCO2 (14-amino acid windows only, probability 0.38), and not predicted for CO. Torok and Elkin (2000) suggested that B-boxes may function to orientate coiled-coil domains that are the site of interaction with other proteins. However, in animal proteins, the coiled-coil region is separated from the B-box by five to eight amino acids. Therefore, it is unclear if the predicted secondary structure of the middle region would affect protein interaction of the plant proteins.
Figure 4. Genetic map locations of CONSTANS-like genes in barley and rice. 1, Genetic map lengths of barley chromosomes 1H, 2H, 5H, 6H, and 7H from Qi et al. (1996) indicating the approximate positions of partial maps that are drawn to larger scale. Solid ovals show positions of centromeres derived from other published maps. 2, Genetic map lengths of rice (Legend continues on facing page.)
Southern Hybridization Analysis of CONSTANS-Like Genes of Barley

To assess whether the nine HvCO genes comprised the whole CO-like family in barley, Southern blots of barley genomic DNA were hybridized with the CCT domain probes used for library screening, stripped, and reprobed with single-copy subclones for each gene in turn. The CCT domain of Hd1 detected two bands in barley, as did the CCT domains of HvCO1 and HvCO2, with reciprocal strength differences, consistent with the presence of two genes. Hybridization of these three probes to rice DNA cut with several restriction enzymes detected one band, suggesting that only one such gene (Hd1) is present in rice, consistent with the analysis of indica and japonica genomic sequence.

The CCT domains of S13576 and HvCO4 gave equivalent results, detecting up to six bands on Southern blots of barley genomic DNA, none of which corresponded to bands detected by HvCO1 or HvCO2 probes. When Southern blots were reprobed with single-copy subclones, all bands detected by the CCT domains could be accounted for. Thus, there was no evidence for additional CO-like genes in barley. However, this does not exclude the possibility that additional genes exist that have CO-like CCT domains with diverged nucleotide sequences (although no such genes were detected by TBLASTN searches) or CCT domains that comigrated with other bands on the Southern blots.

Sequence comparisons showed that although the predicted amino acid sequences of the CCT domains used as probes were similar, the nucleotide sequences were more diverged. This accounted for the different behavior of CCT domains on genomic Southern blots and the failure of the S13576 or HvCO4 CCT domain probes to detect HvCO1 and HvCO2. Thus, no single CCT domain probe could identify all members of the barley CO-like family at the stringencies used.

Genetic Mapping of Barley CONSTANS-Like Genes

Hybridizations of single-copy subclones to wheat/barley telosomic addition lines were used to assign barley genes to chromosome arms. Genes were then mapped in barley or, where polymorphism was lacking, Hordeum bulbosum or rye (Fig. 3), utilizing the known colinearity of their maps (Salvo-Garrido et al., 2001 and Devos et al., 1993, respectively). This showed that barley CO-like genes were present on several chromosomes, with three on chromosome 6H (HvCO2, 5, and 7) and three on chromosome 7H (HvCO1, 6, and 8). HvCO1 was in a region of 7H previously shown to be colinear with the region of rice 6 containing Hd1 (Van Deynze et al., 1995). HvCO3 was located close to the centromere on barley chromosome 5HL, whereas OsB (S12569) was mapped to the end of rice chromosome 9 (Song et al., 1998; Fig. 3). We confirmed that single-copy subclones from HvCO3 and OsB (S12569) cross-hybridized and detected the same loci in reciprocal hybridizations. The respective chromosome regions were shown previously to be colinear (Van Deynze et al., 1995; Foote et al., 1997).

Mapping in rice using CCT domain probes detected several loci and suggested that HvCO6 (7HL) had a counterpart on rice 6L (Fig. 4). This was confirmed by analysis of genomic sequence, which showed that OsE was in the predicted region (Table I). Rice 6 also contained OsL close to OsA (Hd1) but, as expected from differences in the nucleotide sequences, this was not detected on the Southern blots.

Single-copy subclones of barley genes, which for the most part were derived from the middle regions of the genes (Fig. 1), hybridized to rice with varying efficiency, and no systematic attempt was made to map CO-like genes in the latter. However, the HvCO4 subclone detected two loci. Strongly hybridizing bands detected a locus at the expected colinear position on rice chromosome 4L; this is probably OsC, which was found on a chromosome 4 sequence (Table I). Weakly hybridizing bands detected a second locus on rice chromosome 3S, which is possibly OsD.

Relationship of Barley CONSTANS-Like Genes to Known Photoperiod Loci

In wheat and barley, major genes regulating photoperiod response have been mapped on the short arms of the group 2 chromosomes, where their location relative to common RFLP markers suggests that...
they form a homoeoallelic series (Börner et al., 1998 and Sourdille et al., 2000 for wheat and Laurie et al., 1995 and Decousset et al., 2000 for barley). None of the barley CO-like genes mapped to the Ppd-H1 region on chromosome 2HS; therefore, there were no candidates for this major photoperiod response gene. HvCO9, although located in the same chromosome region as Ppd-H2 (Laurie et al., 1995), was not in the same map interval and because HvCO9 was the least CO-like barley gene it was not considered a likely candidate. Similarly, HvCO1 is clearly related to Hd1 but does not correspond to a known major flowering time gene in barley. HvCO1 (7H), HvCO4 (2H), HvCO3 (5H), HvCO2 and HvCO5 (6HL), and HvCO7 (7HS) are all in regions previously shown to contain QTLs for flowering time (Laurie et al., 1995; GrainGenes, http://wheat.pw.usda.gov/). However, in no case did the locus coincide with the maximum likelihood peak for QTL location. Further work would be needed to determine if any of the mapped sequences are candidates for these QTL. HvCO7 was located on the same chromosome arm as the recessive $\text{cam}^7$ (early maturity7) mutant, which is early flowering under LD and SD conditions (Stracke and Börner, 1998). However, HvCO7 occupied a more proximal position.

**DISCUSSION**

**The Evolution and Divergence of CONSTANS-Like Genes**

Previous analysis of CO-like genes sequences by Lagercrantz and Axelsson (2000) concluded that they evolve rapidly, particularly in the middle regions. Their analysis focused on B-box sequences and included the Arabidopsis STO (SALT TOLERANCE) gene and related sequences from rice. STO-like genes have B-boxes but no CCT domain.

In this paper, we restricted our analysis to genes with B-box and CCT domains. Analysis of this group shows that the major subgroups of Arabidopsis genes (Ia, Ic, II, and III in Fig. 3) are all present in cereals. Thus, they predate monocot/dicot divergence and have conserved distinctive B-box and CCT domain sequence characteristics since the subgroups also retain distinctive conserved motifs in the middle region and at the carboxy terminus of their peptides (Fig. 3b). Furthermore, the numbers of genes in each grouping was similar between species, particularly between barley and rice. This clearly suggests that pairs of barley and rice genes are orthologous, and this was supported by genetic mapping that showed that HvCO1/OsA (Hd1), HvCO3/OsB (S12569), and HvCO6/OsE were in regions previously shown to be colinear. Comparison of these genes showed the middle region to be the most diverged and, hence, the most rapidly evolving.

Cereals, however, possess two classes of CO-like genes that were not found in Arabidopsis. One class is Group I genes with a single B-box. Alignment with other genes and the presence of conserved domains in the middle and carboxy-terminal regions suggests that they evolved by internal deletion of a two B-box Group I gene. The HvCO3/OsB and HvCO8/OsG pairs are not closely related and probably evolved independently. The HvCO8/OsG pair were the only genes for which there was no evidence of an intron, but they have intact reading frames and were found in ESTs of barley and sorghum (Sorghum bicolor). These characters, together with differences in sequence, make it unlikely that they are pseudogenes derived from HvCO3/OsB. Also novel are the Group IV genes that lack B-box domains but have a CO-like CCT domain.

The most CO-like genes (Group Ia) differ in number between species. Southern-blot analysis and analysis of genomic sequence detected one rice gene (Hd1) whereas barley clearly has two (HvCO1 and HvCO2) and Arabidopsis has three (CO, COL1, and COL2). In Arabidopsis, CO and the closely related COL1 gene are arranged in tandem, separated by about 3.9 kb (accession no. AL391144). The rice PAC clone containing the Hd1 gene (accession no. AP003044) and corresponding 93-11 genomic contigs had no such duplication. To test the situation in barley, three restriction enzymes were used to digest barley BAC clones containing HvCO1 or HvCO2. Southern blots were tested with CCT domain probes that consistently revealed a single band. Therefore, there was no evidence in rice or barley for a tandem duplication.

A surprising feature of the most CO-like barley and rice genes was changes in B-box2 involving amino acid residues believed to be essential for B-box function (Fig. 2a). This suggests that these cereal genes do not require a second B-box for function, and in this respect, they differ from CO in Arabidopsis. Changes to B-box2 were not a general feature of cereals because the HvCO4 to HvCO7 genes of barley, and their rice counterparts, retained characteristic B-box residues. Further work is needed to understand why B-box2 has been altered in the most CO-like genes and whether the derived regions have novel roles in gene function. The presence of one B-box in HvCO3 and OsB, which were the most closely related to HvCO1 and Hd1 in the COOH region, might also reflect an evolutionary trend toward a one B-box structure.

The existence of common classes of CO-like genes in monocots and dicots clearly suggests that CO-like genes have an ancient origin. Database searches confirmed this, showing that CO-like CCT domains (defined as being more similar to CO than to other Arabidopsis genes such as TOCI) are present in Chlamydomonas reinhardtii and, therefore, evolved before the divergence of unicellular green algae from other green plants. CO-like genes (defined by one or more B-boxes, a CO-like CCT domain, and conserved...
residues at the COOH terminus) are present in mosses (Physcomitrella patens), gymnosperms, and angiosperms (Fig. 5), showing that they predate the evolution of these groups.

The Role of CONSTANS-Like Genes in Barley

Studies in wheat and barley have consistently shown that Ppd genes on the short arms of the group 2 chromosomes are the major factors controlling photoperiod response, and comparative mapping suggests that they are a series of orthologous genes (Börner et al., 1998; Sourdille et al., 2000; and Laurie et al., 1995, respectively). We found no evidence for a CO-like gene in the Ppd region. However, cereals clearly possess CO homologs, and at least one gene (OsA [Hd1]) plays a major role in photoperiod response in rice. OsA (Hd1) has a clear counterpart in barley (HvCO1), but to date we have found no correspondence between barley CO-like genes and flowering time loci. Thus, variation for photoperiod response in rice and temperate cereals has been achieved by different routes. To investigate this further, we are analyzing the role of barley CO-like genes using transgenic plants. In addition to overexpression studies within the relevant species, it would be interesting to determine whether Arabidopsis co mutants can be complemented by HvCO1 or HvCO2, either with their normal structure or with a second, more conserved, B-box2. It would also be interesting to determine if the barley HvCO1 gene can complement the hd1 mutation of rice and, if so, whether the resulting plant has SD or LD characteristics.

CLONING AND SEQUENCING OF BARLEY (Hordeum vulgare) CONSTANS-LIKE GENES

To clone barley CO-like genes, genomic libraries were first screened with CCT domain probes from two rice CO-like genes, Hdl (Yano et al., 2000) and S13576. The rice Hdl gene and sequence were kindly provided by Dr. Masahiro Yano (National Institute of Agrobiological Resources, Tsukuba, Japan). S13576 was one of four rice EST clones that were fully sequenced after having been identified by database searches as containing a CO-like B-box. Of the three other rice ESTs, two (R1479 and R2967) had no CCT domain, whereas the third (S3574) had a 43-amino acid CCT domain that was less homologous to CO than that of S13576 (60% and 88% amino acid identity, respectively). Therefore, the CCT domains of Hdl and S13576 were selected for screening barley libraries. S13576 was subsequently found to be identical in sequence with S12569, a cDNA previously described by Song et al. (1998) that is referred to as S12569 elsewhere in this paper. The Hdl-CCT fragment was amplified using primers HD1-CCTf (CAGGGAGGCC-CAGGGTGCTCAG) and HD1-CCTr (CTCTTGCCAAAGCCCTTGTG). The S13576-CCT fragment was amplified using primers S12569-CCTf (TACAGGGA-GAAGAGGGAAGCAGAG) and S12569-CCTr (AGAAAGACT-GGTCAAAGCTCAAG). 32P-labeled Hdl-CCT was hybridized to 3.13 × 10^6 barley BAC clones from the var. Morex yielding two CO-like genes. High-density filters of the Morex BAC library, described by Yu et al. (2000), and Morex, CI16151, and CI16155 cDNA clones were obtained from Clemson University Genomics Institute (SC; http://www.genome.clemson.edu). S12569-CCT and subsequent screens with barley CCT domain probes detected six CO-like genes (HvCO3-HvCO9) when hybridized to 1 × 10^6 barley genomic phage clones from a commercially available library (Stratagene Inc.) from the var. Igrid.

BAC DNA was extracted using a modified lysis procedure from Clemson University (http://www.genome.clemson.edu/protocols.j). Phage DNA was extracted from plate lysates according to Sambrook et al. (1989). Gene-containing fragments were subcloned by restriction digestion of 5 µg of DNA followed by fractionation in 0.8% (w/v) agarose for 20 h at 1 V cm^-1, and then blotted onto charged nylon membrane (Amersham, Buckinghamshire, UK) according to the manufacturer’s instructions. Hybridization with the same probe sequence used to screen the genomic library identified gene-containing restriction fragments. These fragments were excised from preparative agarose gels, purified by electrophoresis, and cloned into pBlue seront (Stratagene Inc.) using standard laboratory procedures. Subclone sequencing began using M13 forward and reverse primers and progressed through the insert with custom primers. Reactions were carried out using BigDye (Perkin-Elmer Applied Biosystems, Foster City, CA) chemistry and fractionated on a Perkin-Elmer Applied Biosystems 3700 automated sequencing machine. Alternatively, genes were sequenced directly from BAC or phage clones, initially using primers based on the highly conserved CCT

MATERIALS AND METHODS

Nomenclature

Italicized names refer to genes and gene segments (e.g. HvCO1 and Hdl), whereas nonitalic names (e.g. HvCO1, Hdl) refer to predicted peptides. Loci (Fig. 4) are italicized and prefixed with X (e.g. XhCO1 and XHdl).

Analysis of Rice (Oryza sativa) Genomic Sequence

The nucleotide and predicted peptide sequences of Arabidopsis CO and COL1 to COL16 (Table I) were used for BLASTN and TBLASTN searches of databases using facilities provided by the National Center for Biological Information Web site (http://www.ncbi.nlm.nih.gov). In addition to genomic sequence of indica rice 93-11 (Yu et al., 2002) and japonica rice Nipponbare (http://www.dna.affrc.go.jp), we also analyzed Nipponbare genomic sequence from the Torrey Mesa Research Institute (http://www.tmri.org; Goff et al., 2002).

Cloning and Sequencing of Barley (Hordeum vulgare) CONSTANS-LIKE GENES

To clone barley CO-like genes, genomic libraries were first screened with CCT domain probes from two rice CO-like genes, Hdl (Yano et al., 2000) and S13576. The rice Hdl gene and sequence were kindly provided by Dr. Masahiro Yano (National Institute of Agrobiological Resources, Tsukuba, Japan). S13576 was one of four rice EST clones that were fully sequenced after having been identified by database searches as containing a CO-like B-box. Of the three other rice ESTs, two (R1479 and R2967) had no CCT domain, whereas the third (S3574) had a 43-amino acid CCT domain that was less homologous to CO than that of S13576 (60% and 88% amino acid identity, respectively). Therefore, the CCT domains of Hdl and S13576 were selected for screening barley libraries. S13576 was subsequently found to be identical in sequence with S12569, a cDNA previously described by Song et al. (1998) that is referred to as S12569 elsewhere in this paper. The Hdl-CCT fragment was amplified using primers HD1-CCTf (CAGGGAGGCC-CAGGGTGCTCAG) and HD1-CCTr (CTCTTGCCAAAGCCCTTGTG). The S13576-CCT fragment was amplified using primers S12569-CCTf (TACAGGGA-GAAGAGGGAAGCAGAG) and S12569-CCTr (AGAAAGACT-GGTCAAAGCTCAAG). 32P-labeled Hdl-CCT was hybridized to 3.13 × 10^6 barley BAC clones from the var. Morex yielding two CO-like genes. High-density filters of the Morex BAC library, described by Yu et al. (2000), and Morex, CI16151, and CI16155 cDNA clones were obtained from Clemson University Genomics Institute (SC; http://www.genome.clemson.edu). S12569-CCT and subsequent screens with barley CCT domain probes detected six CO-like genes (HvCO3-HvCO9) when hybridized to 1 × 10^6 barley genomic phage clones from a commercially available library (Stratagene Inc.) from the var. Igrid.

BAC DNA was extracted using a modified lysis procedure from Clemson University (http://www.genome.clemson.edu/protocols.j). Phage DNA was extracted from plate lysates according to Sambrook et al. (1989). Gene-containing fragments were subcloned by restriction digestion of 5 µg of DNA followed by fractionation in 0.8% (w/v) agarose for 20 h at 1 V cm^-1, and then blotted onto charged nylon membrane (Amersham, Buckinghamshire, UK) according to the manufacturer’s instructions. Hybridization with the same probe sequence used to screen the genomic library identified gene-containing restriction fragments. These fragments were excised from preparative agarose gels, purified by electrophoresis, and cloned into pBlue sera (Stratagene Inc.) using standard laboratory procedures. Subclone sequencing began using M13 forward and reverse primers and progressed through the insert with custom primers. Reactions were carried out using BigDye (Perkin-Elmer Applied Biosystems, Foster City, CA) chemistry and fractionated on a Perkin-Elmer Applied Biosystems 3700 automated sequencing machine. Alternatively, genes were sequenced directly from BAC or phage clones, initially using primers based on the highly conserved CCT
domain of CO-like genes. Sequence reads were assembled into contigs using the Sequencer program (Gene Codes Corporation, Ann Arbor, MI). Multiple sequence alignments were constructed using ClustalX followed by manual optimization. GenBank sequence accession numbers for the barley genes are listed in Table I.

Southern Blotting, Hybridization, and Genetic Mapping
To assess the copy number of CO-like genes in barley, Southern blots were prepared from barley genomic DNA digested with EcoRI, EcoRV, DraI, or HindIII. CCT domain probes and gene-specific probes (Fig. 1) were amplified by PCR, radiolabeled, and used as described by Laurie et al. (1993). For heterologous probes, the filters were washed at a final stringency of 2× SSC or 0.2× SSC at 65°C. Otherwise, the final stringency was 0.2× SSC at 65°C. Wheat (Triticum aestivum)/barley telomeric addition lines (Islam, 1983) were used to assign single-copy subclones of barley genes to chromosome arms. Mapping populations used to locate CO-like genes in barley, Hordeum bulbosum, rye (Secale cereale), and rice are referenced in the legend to Figure 4. Genetic maps were produced using JoinMap v2.0 (Stam and Van Ooijen, 1995).

Phylogenetic Analysis
To compare Arabidopsis, barley, and rice genes, we used concatenated nucleotide or predicted peptide sequences of the B-box and CCT domains or individual domains for comparisons of genes where not all domains were present. Sequences were aligned using ClustalX or the pileup facility of the GCG package (Genetics Computer Group, Madison, WI) followed by manual adjustment. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 2.1 software (http://www.megasoftware.net; Kumar et al., 2001). Genetic distances between nucleotide sequences and predicted peptide sequences were calculated various parameters provided by the software, and phylogenetic trees were constructed using NJ and MP methods. Nucleotide comparisons excluded the third position in each codon. Data files used for the analysis are available from the corresponding author on request.

Received October 15, 2002; returned for revision December 3, 2002; accepted December 12, 2002.

LITERATURE CITED


Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis Software. Arizona State University, Tempe
Evolution of CONSTANS-Like Gene Families

Arabidopsis
clock gene TOC1, an autoregulatory response regulator homolog. Science 289: 768–771
Arabidopsis.
Nature 410: 1116–1119
Yano M, Katayose Y, Ashikara M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y et al. (2000) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the 
Arabidopsis
flowering time gene CONSTANS. Plant Cell 12: 2473–2483