The Distribution of Transgene Insertion Sites in Barley Determined by Physical and Genetic Mapping

Haroldo Salvo-Garrido,1,2 Silvia Travella,1,3 Lorelei J. Bilham,1 Wendy A. Harwood4 and John W. Snape

John Innes Centre, Department of Crop Genetics, Norwich Research Park, Norwich NR4 7UH, United Kingdom

Manuscript received November 6, 2003
Accepted for publication April 2, 2004

ABSTRACT

The exact site of transgene insertion into a plant host genome is one feature of the genetic transformation process that cannot, at present, be controlled and is often poorly understood. The site of transgene insertion may have implications for transgene stability and for potential unintended effects of the transgene on plant metabolism. To increase our understanding of transgene insertion sites in barley, a detailed analysis of transgene integration in independently derived transgenic barley lines was carried out. Fluorescence in situ hybridization (FISH) was used to physically map 23 transgene integration sites from 19 independent barley lines. Genetic mapping further confirmed the location of the transgenes in 11 of these lines. Transgene integration sites were present only on five of the seven barley chromosomes. The pattern of transgene integration appeared to be nonrandom and there was evidence of clustering of independent transgene insertion events within the barley genome. In addition, barley genomic regions flanking the transgene insertion site were isolated for seven independent lines. The data from the transgene flanking regions indicated that transgene insertions were preferentially located in gene-rich areas of the genome. These results are discussed in relation to the structure of the barley genome.

ONE of the main goals of plant genetic engineering is to achieve stable transgenic events that give predictable and reproducible levels of expression of a transgene and that are fully characterized in terms of the effect and implications of the transgene insertion for the plant. To attain this goal, it is necessary to obtain detailed information about the transgene insertion site and to characterize transgenic plants in detail at the physical, genetic, and molecular levels. An understanding of the transgene insertion site within the host genome is an essential first step toward understanding the importance of the transgene insertion position and in aiding the safety assessment process for a particular transgenic event. The positive position of transgenes has been detected by fluorescent in situ hybridization (FISH) in cereals such as barley, wheat, triticale, and oat (Pedersen et al. 1997; Leggett et al. 2000; Svitashev et al. 2000; Salvo-Garrido et al. 2001; Choi et al. 2002). However, equally important is the genetic map position of transgenes, and to our knowledge, this has not been studied in any crop species. Combining knowledge of the physical position of transgenes with their genetic map position provides confirmation of the insertion site and a better understanding of the pattern of transgene insertions. It is also a valuable starting point for understanding the genomic environment in which a transgene resides and for determining whether transgene insertion is indeed a random process.

Additional information on transgene insertions can be obtained from the analysis of transgene flanking regions. Large-scale studies of T-DNA insertions have been undertaken in Arabidopsis using this method (Forsbach et al. 2003; Qin et al. 2003a,b) and also in rice (Sha et al. 2004). Such studies are leading to an increased understanding of the preferences for transgene insertion and its implications. However, this has not yet been done for the large-grained cereals, such as barley, where ~80% of the genome is repetitive DNA.

In the present study, the physical position of transgenes in 19 independent transgenic barley lines has been determined using a highly sensitive and reproducible method of determining the chromosomal position of transgenes on metaphase chromosomes (Salvo-Garrido et al. 2001). These locations were then confirmed by genetic mapping of the transgenes in 11 of these independently transformed lines. This is the first report of both physical and genetic mapping of transgenes in cereals. In addition, further information on the transgene insertion environment was obtained by isolating and characterizing the genomic sequence flanking the transgenes in some of the lines. The patterns of trans-
TABLE 1
The transgenic lines used for physical and genetic mapping

<table>
<thead>
<tr>
<th>Cross parent</th>
<th>Transgenic line</th>
<th>Transgenic generation</th>
<th>Constructs</th>
<th>Gene transfer methods</th>
<th>Chromosomal location of insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nigrinudum AGN1A</td>
<td>T₁</td>
<td>pDM805</td>
<td>AMT</td>
<td>2H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum AGN4A</td>
<td>T₀</td>
<td>pDM805</td>
<td>AMT</td>
<td>5H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum H1B1A</td>
<td>T₁</td>
<td>pAHC25</td>
<td>PB</td>
<td>5H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum HC1B</td>
<td>T₁</td>
<td>pAL51 + pAL74</td>
<td>PB</td>
<td>4H</td>
<td></td>
</tr>
<tr>
<td>Hayakeizo HC8</td>
<td>T₁</td>
<td>pAL51 + pAL74</td>
<td>PB</td>
<td>4H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum HC12B</td>
<td>T₁</td>
<td>pAL51 + pAL74</td>
<td>PB</td>
<td>6H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum HG4B</td>
<td>T₀</td>
<td>pAL51 + R173</td>
<td>PB</td>
<td>2H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum HG3B</td>
<td>T₀</td>
<td>pAL51 + R173</td>
<td>PB</td>
<td>4H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum HI1A</td>
<td>T₁</td>
<td>pAL51 + pDAPH7</td>
<td>PB</td>
<td>6H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum HH2A</td>
<td>T₁</td>
<td>pAL51 + pPhyt</td>
<td>PB</td>
<td>5H</td>
<td></td>
</tr>
<tr>
<td>HH3E</td>
<td>T₁</td>
<td>pAL51 + pPhyt</td>
<td>PB</td>
<td>6H</td>
<td></td>
</tr>
<tr>
<td>HH4F</td>
<td>T₁</td>
<td>pAL51 + pPhyt</td>
<td>PB</td>
<td>5H</td>
<td></td>
</tr>
<tr>
<td>Nigrinudum HR18D</td>
<td>T₀</td>
<td>pAL51 + γ-gliadin</td>
<td>PB</td>
<td>4H</td>
<td></td>
</tr>
<tr>
<td>Bronze Tr1372D8</td>
<td>T₁</td>
<td>pAHC25 + pDC-2CP</td>
<td>PB</td>
<td>4H</td>
<td></td>
</tr>
<tr>
<td>Fanfare Tr130</td>
<td>T₁</td>
<td>pAHC25 + pDC-2CP</td>
<td>PB</td>
<td>2H</td>
<td></td>
</tr>
</tbody>
</table>

Details of crosses made, constructs, transgenic generation, gene-transfer method, and chromosomal location for each transgenic line are included. PB, particle bombardment; AMT, agrobacterium-mediated transformation.

Gene insertion observed are discussed in relation to patterns of insertion seen in other plants and to the structure of the barley genome.

MATERIALS AND METHODS

Plant materials: The plant materials studied by FISH and mapping procedures were T₁ transgenic plants and F₂ populations of crosses between T₀, T₁, T₂, or T₃ transgenic lines of barley (Hordeum vulgare cv. Golden Promise) with the non-transgenic barley varieties, Nigrinudum, Hayakeizo, Fanfare, and Bronze. Table 1 summarizes the crosses, constructs, and gene transfer methods used. All transgenic lines were produced at the John Innes Centre except lines Tr130 and Tr1372D8, obtained from Monsanto (Cambridge, United Kingdom). Most of the lines were produced by particle bombardment of immature embryos (Harwood et al. 2002), except for lines AGN1A and AGN4A, which were transformed by Agrobacterium following the method of Tingay et al. (1997). The plasmid pPhyt contained a phytase gene under the control of an α-amylase promoter, pDAPH7 contained the DapA gene involved in lysine biosynthesis, pAL74 contained a fungal glucoamylase gene under an α-amylase promoter, and pDC-2CP contained a viral coat protein. All other constructs—pAHC25 (Christensen and Quail 1996), pAL51 (Harwood et al. 2002), and the binary vector pDM805 (Tingay et al. 1997)—contained the bar gene as a selectable marker, which encodes resistance to the glufosinate group of herbicides, together with a reporter gene, either the gus gene (β-glucuronidase) or the firefly luciferase gene. All transgenic lines examined expressed the bar gene.

Physical mapping by FISH: FISH was performed according to Salvo-Garrido et al. (2001). Simultaneous hybridization was carried out with probes pTa71 (18S-5.8S-26S rDNA) and pTa794 (5S rDNA), which hybridize to ribosomal DNA to give distinct and specific chromosomal patterns that unambiguously identify all seven individual chromosomes in barley (Leitch and Heslop-Harrison 1993).

Marker probe pTa71 hybridized to specific sites of ribosomal DNA on the short arms of chromosomes 1H, 2H, 5H, 6H, and 7H. Marker probe pTa71 labels chromosomes 5H and 6H at the nucleolar organizer region (NOR) on the short arm to produce the strongest hybridization bands across both chromatids. Chromosomes 1H, 2H, and 7H are more weakly labeled by pTa71, but show a specific hybridization signal in each chromatid. It is possible to identify chromosomes 5H and 6H without a marker probe on the basis of morphology alone because of the presence of the NOR, which is visible as a secondary constriction in a chromosome spread. These chromosomes can be distinguished by observing that chromosome 5H has a much longer, long arm compared to its short arm whereas chromosome 6H has a short and long arm of equal length.

Marker probe pTa794 hybridized to regions of ribosomal DNA in the long arm of chromosomes 1H, 2H, 5H, 6H, and 7H. Marker probe pTa71 labels chromosomes 3H and 4H at the nuclear organizer region (NOR) on the short arm to produce the strongest hybridization bands across both chromatids. Chromosomes 1H, 2H, and 7H are more weakly labeled by pTa794, but show a specific hybridization signal in each chromatid. It is possible to identify chromosomes 5H and 6H without a marker probe on the basis of morphology alone because of the presence of the NOR, which is visible as a secondary constriction in a chromosome spread. These chromosomes can be distinguished by observing that chromosome 5H has a much longer, long arm compared to its short arm whereas chromosome 6H has a short and long arm of equal length.

Mapping procedures: Bulk segregant analysis (BSA) was initially applied to establish marker/transgene associations (Michelmore et al. 1991). F₂ plants at the three/four leaf stage, derived from the crosses described above, were pheno-
typed in terms of resistance and susceptibility to the herbicide. For BSA analysis, DNA of the resistant plants was bulked into one sample. DNA of susceptible plants was isolated individually and a bulked sample was also prepared. Bar PCR was performed on the sensitive plants to check for transgene silencing and any PCR-positive plants were discarded from the mapping population.

Molecular markers: Simple sequence repeat (SSR) and restriction fragment length polymorphism (RFLP) assays were employed. Three different sources of microsatellite primers were used (Becker and Heun 1995; Liu et al. 1996; Ramsay et al. 2000). The RFLP probes were those already widely mapped in barley and wheat.

Linkage analysis and estimation of recombination frequency: The physical mapping results provided a starting point for genetic mapping by locating transgenes to chromosomes and chromosome arms, as described above. Known polymorphic codominant markers for the target region were then genotyped on the susceptible population. Two-point recombination frequencies were obtained from a simple maximum-likelihood estimation equation and map order was calculated. Recombination frequency was converted to centimorgans using the Kosambi mapping function. Comparative mapping with published maps of barley for the corresponding regions was used to compare gene orders.

Isolation of transgene flanking regions: The lines selected for analysis of transgene flanking regions were all transformed with plasmid pAL51, containing the luciferase (luc) reporter gene. All lines had a low copy number and simple integration patterns as determined by Southern analysis. In addition, they did not contain an intact copy of the luciferase gene as the plasmid breakpoint was within the luciferase cassette. Genomic DNA was extracted using a CTAB protocol (Harwood et al. 2000). PCR was carried out using a range of primer combinations within the luc cassette. In this way the breakpoint within the luc cassette was identified to within 200–550 bp for each line. Adaptor-mediated PCR was then used to allow the specific amplification of the transgene flanking region. The method described by Spertino et al. (1999) was used with the following modifications. PCRs were performed in a 25-μl total reaction volume using 2 μl ligated DNA, 1 μl 2.5 mM nested primer, 1 μl 2.5 mM AP1/AP2 (Spertino et al. 1999), and 21 μl Abgene 1.1× Reddy Mix with 1.5 mM MgCl₂. The second PCR amplification was performed using 2 μl of product from the first PCR amplification. Following the second PCR amplification, the total PCR product was run out on a 1% agarose gel and bands were excised. DNA was extracted from the gel bands using a QIAAGEN (Valencia, CA) QIAquick gel extraction kit. The extracted DNA was cloned using the Promega (Madison, WI) pGem-T Easy Vector System I, prior to Perkin-Elmer (Norwalk, CT) Big Dye terminator version II sequencing. Sequence data were analyzed by BLAST against the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) and Duport databases.

Statistical analyses: An exact chi-square test (using StatExact 4) was carried out to determine whether the transgene insertions were uniform over the barley genome. The different sizes of the barley chromosomes were taken into account during the analysis.

RESULTS

Physical map position of transgenes: The improved protocol described by Salvo-Garrido et al. (2001) was used for the detection of transgenes by FISH. This method produced specific signals on metaphase chromosomes in all the lines under study. Simultaneous hybridization with karyotyping ribosomal probes (pTa71 and pTa794) allowed the clear identification of the seven barley chromosomes.

The physical location of transgenes was determined in 19 independent barley lines, and a total of 23 transgene integration sites were detected (Figures 1 and 2). Lines Tr130 and AGN4A showed two insertion sites on separate chromosomes; line HG3B showed three insertions on the same chromosome (Salvo-Garrido et al. 2001) while the remaining lines showed a single insertion site. Of the seven barley chromosomes, five (2H, 3H, 4H, 5H, and 6H) were involved in transgene integration, but no sites were detected on chromosomes 1H and 7H (Figure 2).

FISH detected 39% of the transgene signals as being in the telomeric or subtelomeric regions of both the long and the short arms. The centromeric regions contained 17% of the transgene signals. Other regions of the long arms of chromosomes contained 22% of the transgene signals while the short arms contained the remaining 22%. Most integration sites were found on chromosomes 5H (30%), 4H (30%), and 6H (22%). An exact chi-square test (using StatExact 4) was carried out to determine whether the transgene insertions were uniform over the barley genome (Table 2). The exact test takes into account the small sample size by conditioning the distribution of counts on the total; hence the possible arrangements of counts in the cells of the 1 × 7 chi-square table follow a multinomial distribution. The different sizes of the seven barley chromosomes were also taken into account in determining the expected transgene insertions for each chromosome (Laurie et al. 1992). Strong evidence was found against the null hypothesis that there was no difference in the frequency of transgene insertions between the different chromosomes (P > 0.05). Thus it was concluded, on the basis of the FISH results, that transgene insertion was not random in the barley lines analyzed in this study.

Transgenic line Tr130 was cotransformed with constructs pAH25 (bar gene) and pDC-2CP (viral coat protein). Two different integration sites were observed for pDC-2CP and only one integration site was mapped with pAH25. On chromosome 5H, both pDC-2CP and pAH25 shared the same integration site close to the centromere in chromosome 5H, while pDC-2CP gave an additional signal centromeric in chromosome 2H. These results indicate that the bar and the viral coat protein gene have integrated independently into the genome.

Genetic map position of the transgenes: The genetic location of transgenes was determined for 11 independent lines (Figure 1). All genetic maps showed collinearity to published SSR and RFLP barley maps. The physical integration sites showed complete agreement with the genetic map positions of the transgenes. The length distortion between the genetic and physical distance of
the transgenes is similar to that previously observed within cereal species (Gill et al. 1996).

The transgene in line HG4B was physically mapped to the telomeric region of the long arm of chromosome 2H (Figure 1a). In the genetic map, the bar gene showed linkage with the RFLP marker Xmag895, mapped in the long arm of chromosome 2H (Laurie et al. 1993), 27 cM from the centromere and 3 cM from the 5S rDNA site (designated 5SDna-H3, Leitch and Heslop-Harrison 1993). In line HG4B, the locus Xmag895 was mapped 10 cM from the bar gene, indicating that the transgene was genetically located close to the 5S rDNA locus. This confirmed the results of the physical mapping. In the cross H11A × Nigrinudum, SSR locus Bmag209 mapped in the proximal region of the short arm of 3H (Ramsay et al. 2000) and showed complete linkage to the transgene (Figure 1b).

All transgene insertions found by FISH on chromosome 4H were genetically mapped (Figure 1c). FISH indicated that the transgene in cross HG3B × Nigrinudum was telomeric in the long arm of chromosome 4H, and genetic mapping showed that it was 11 cM from the markers XzenC6 and HVM67. In the barley map developed by Liu et al. (1996), HVM67 was mapped 8 cM from the most terminal marker BAmy1 (Kleinhofs et al. 1993), confirming this result. In the cross Bronze × Tr1372D8, the transgene was expected to be in the centromeric region of chromosome 4H on the basis of FISH results. Total linkage was found with RFLP marker Xmag58, previously mapped close to the 4H centromeric region (Kleinhofs et al. 1993). The genetic position of the bar gene in line Tr1372D8 was therefore in the proximal region of the long arm of chromosome 4H.

SSR markers HVM3 and Bmag0375 were totally linked to the bar gene in cross HC8 × Hayakeizo, and both have been mapped to the 4H centromeric region (Liu et al. 1996; Ramsay et al. 2000). The transgene in the cross H11A × Nigrinudum showed total linkage to markers HVM3, Bmac181, and Bmag0375, centromeric on chromosome 4H and similar to the transgene in line HC8. With respect to the cross HR18D × Nigrinudum, the two SSR markers Bmag0375 and Bmac181, which cosegregated together, were mapped 4 cM from the bar transgene, again in the proximal region of the short arm of chromosome 4H, close to the centromere. These three independent transgene insertions therefore all mapped to within 4 cM on the short arm of chromosome 4H. The total map distance for chromosome 4H is ~120 cM, so to find three insertions within 4 cM suggests some preference for insertion into this region of the genome.

Transgene insertions in lines HB1A and Tr130 were genetically mapped to chromosome 5H (Figure 1d). The transgene in the cross HB1A × Nigrinudum was closely flanked by loci Xpar145 and XEmbp1 and mapped to the terminal region of chromosome 5H (Laurie et al. 1995), confirming the FISH results. SSR loci Ebmac970 and Ebmac684 cosegregated with the bar gene in the cross Fanfare × Tr130, deduced to be in the centromeric region of chromosome 5H.

**Figure 1.**—Summary of the 19 physically and 11 genetically mapped transgene insertion sites found on chromosomes (a) 2H, (b) 3H, (c) 4H, (d) 5H, and (e) 6H of barley. FISH was carried out using a biotin- or digoxigenin-labeled probe, made from fragments of the integrated plasmid, together with labeled marker probe pTa71 or pTa794, to metaphase chromosomes of barley. FISH data are represented on ideograms of each chromosome (a–e) with actual results shown only for chromosome 4H (c). Integration sites are represented by red spots and the physical locations of 5S rDNA, pTa794 (in yellow), and the 18S-5.8S-26S rDNA, pTa71 (in blue) are indicated. Transgenic line names are given above the ideogram of each chromosome and linked to the site of transgene insertion found by FISH. Published genetic maps are shown in purple with the source reference below. Genetically mapped transgene insertion sites are shown on yellow maps with the cross-reference below. The locations of SSR markers are shown in black type, RFLP markers are shown in blue type, and the site of probe pTa794 (X5S DNA-H3) is shown in green type. (a) includes SSR and RFLP genetic mapping results from the cross HG4b × N, Golden Promise is represented by G, Nigrinudum is represented by N, the bulk of resistant individuals is represented by R, and the bulk of the sensitive individuals is represented by S.
Transgene insertions from two independently transformed lines were mapped to the short arm of chromosome 6H (Figure 1e). The transgene in cross AGN4A × Nigrinudum was physically mapped to the middle of the short arm. RFLP marker $XemC61$ was totally linked to the transgene, confirming the FISH result, since this marker has been mapped 21 cM from the centromeric region on the short arm of chromosome 6H (Laurie et al. 1995). The transgene in the cross HC12B × Nigrinudum was mapped 12 cM from the SSR locus $Bmag0500$ on the short arm of chromosome 6H (Ramsay et al. 2000).

**Transgene flanking regions:** Genomic DNA flanking one side of the transgene insertion was isolated for 5 (HC1B, HC8, H18A, HH2A, and HH4F) of the 19 mapping lines together with two additional lines (HC3 and HC4C). Comparison of the sequence data with available databases showed that the isolated genomic DNA corresponded to recognized coding regions in 6 of these lines, suggesting that the transgene insertions have occurred in coding regions of the barley genome (Table 3).

A total of 500 bp of genomic junction sequence DNA was obtained from line HC1B. The strongest alignment was found when two regions of HC1B flanking DNA matched to a cytochrome B gene; a 107-bp region showed 95% identity and a 207-bp region showed 94% identity on the basis of the results of a BLASTn search of the NCBI nonredundant database.

With respect to line HC8, analysis of 700 bp of genomic DNA flanking the transgene insertion showed that a 183-bp region matched with 90% identity to a lipase-like protein from rice on the basis of the results of a BLASTn search of the Dupont wheat expressed sequence tag (EST) database. From line H18A, a total of 400 bp of flanking sequence was obtained. The most significant match found following BLASTn analysis of the Dupont wheat EST database was to a catalase gene. In this case a 133-bp region from H18A flanking DNA showed 98% identity.

For line HH4F, 500 bp of junction sequence DNA was isolated. The most significant alignment was to a barley bacterial artificial chromosome (BAC) on chromosome 5H, thus further confirming the chromosomal location of this transgene. Homology was also seen to a high-molecular-weight glutenin. In addition to this, a 267-bp region of HH4F flanking sequence showed 52% identity to a rice polypeptide following a BLASTx search of the Dupont wheat EST database. Line HH2A was the only line where homology to known coding regions was not found in the transgene flanking DNA. Good homology was found to another barley BAC and short regions of the 500-bp flanking region showed homology to a putative retroelement from Arabidopsis. The two additional lines, HC3 and HC4C, showed homology to a probable glucosidase protein and a tomato leaf-curl-disease-associated sequence, respectively.

**DISCUSSION**

Although there have been previous reports of the physical mapping of transgenes using FISH, this is, to our knowledge, the first report of the combined physical and genetic mapping of transgenes in a crop plant. For the 19 lines analyzed in this study, only five of the seven barley chromosomes (2H, 3H, 4H, 5H, and 6H) were found to contain transgene integration sites. In addition, there appeared to be specific regions of chromosomes 4H and 5H where clusters of transgene insertions were seen. The distribution of transgene integration sites was found to statistically deviate from a random expectation. The close proximity of transgene integration sites from three independent lines on the short arm of chromosome 4H (lines HC8, HC1B, and HR18D) and also clusters of three independent transgene insertions on both the long arm (lines HC2B, HH2A, and H19A) and the short arm (lines Tr130, HH4F, and AGN4A) of 5H suggests that these areas of the genome may be more amenable for transgene insertion than other regions of the genome. On chromosome 4H, three transgene insertions were found within a genetic distance of 4 cM of a total of 120 cM for chromosome 4H. It has been estimated that only 12% of the barley genome is occupied by the “gene space” (Barakat et al. 1997). This region is characterized by a specific GC percentage range and most genes are found within this genomic space. In wheat, genes have been found to

---

**TABLE 2**

Exact chi-square analysis of the distribution of transgene insertion sites

<table>
<thead>
<tr>
<th>Chromosome no. and size (10⁶ bp)</th>
<th>1H (644)</th>
<th>2H (811)</th>
<th>3H (797)</th>
<th>4H (779)</th>
<th>5H (756)</th>
<th>6H (740)</th>
<th>7H (825)</th>
<th>Totals (5352)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed transgene integration sites</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>Expected transgene integration sites</td>
<td>2.77</td>
<td>3.49</td>
<td>3.43</td>
<td>3.35</td>
<td>3.25</td>
<td>3.18</td>
<td>3.55</td>
<td>23.000</td>
</tr>
<tr>
<td>Chi square</td>
<td>2.77</td>
<td>0.07</td>
<td>1.72</td>
<td>3.98</td>
<td>4.33</td>
<td>1.04</td>
<td>3.55</td>
<td>17.455</td>
</tr>
</tbody>
</table>
Distribution of Transgene Insertion Sites

TABLE 3
Most significant alignments for the transgene flanking regions based on BLASTn and BLASTx searches of the NCBI nonredundant databases and the Dupont wheat and oat EST databases

<table>
<thead>
<tr>
<th>Line</th>
<th>BLASTn results</th>
<th>Accession</th>
<th>E value</th>
<th>BLASTx results</th>
<th>Accession</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1C1B</td>
<td>Cytochrome B</td>
<td>AF205088</td>
<td>e-111</td>
<td>Cytochrome B</td>
<td>AAK54411</td>
<td>1e-19</td>
</tr>
<tr>
<td>HC3</td>
<td>No good matches</td>
<td></td>
<td></td>
<td>Arabidopsis putative protein (D²)</td>
<td>NP_195322</td>
<td>7e-05</td>
</tr>
<tr>
<td>HC4C</td>
<td>Tomato leaf-curl-disease-associated sequence</td>
<td>AJ316036</td>
<td>3e-57</td>
<td>Probable glucosidase protein</td>
<td>NP_519202</td>
<td>3e-18</td>
</tr>
<tr>
<td>HC8C</td>
<td>Rice lipase-like protein (D²)</td>
<td>AP003710</td>
<td>2e-60</td>
<td>Arabidopsis lipase-like protein (D²)</td>
<td>NP_195590</td>
<td>2e-35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rice lipase-like protein</td>
<td>NP_917262</td>
<td>3e-34</td>
</tr>
<tr>
<td>H1A</td>
<td>Catalase (D¹)</td>
<td>X85130</td>
<td>4e-62</td>
<td>Catalase (D²)</td>
<td>140767</td>
<td>4e-17</td>
</tr>
<tr>
<td></td>
<td>mRNA clone</td>
<td>AJ309392</td>
<td>8e-06</td>
<td>Rice lipase-like protein</td>
<td>NP_917282</td>
<td>5e-14</td>
</tr>
<tr>
<td>HH2M</td>
<td>Barley BAC 184G9</td>
<td>AY268139</td>
<td>6e-45</td>
<td>Arabidopsis Athila retroelement (D²)</td>
<td>NP_178678</td>
<td>9e-06</td>
</tr>
<tr>
<td>HH4F</td>
<td>Barley chromosome 5H BAC 635P2</td>
<td>AY013246</td>
<td>2e-60 and 2e-11</td>
<td>Rice polyprotein (D²)</td>
<td>AF259976</td>
<td>4e-31</td>
</tr>
<tr>
<td></td>
<td>Triticum turgidum HMW-glutenin</td>
<td>AY368673</td>
<td>2e-11</td>
<td>Rice gag-pol precursor</td>
<td>AAL75983</td>
<td>8e-17</td>
</tr>
</tbody>
</table>

D¹, based on the results of a BLASTn search on the Dupont database. D², based on the results of a BLASTx search of the Dupont database.

occur in clusters along chromosomes (Gill et al. 1993), providing further evidence for gene-rich regions interspersed with large gene empty compartments in members of the Gramineae. In the present study, only lines expressing the bar gene were evaluated, not lines where the bar gene was present but not expressed. The non-random distribution of functional transgenes may therefore reflect the fact that they are found only in the 12% gene-rich region of the barley genome. The analysis of the transgene-flanking region in 7 lines supports this suggestion as most flanking regions were recognized as known coding regions when the sequences were compared to those held in databases.

Studies of T-DNA insertions in Arabidopsis (Forsbach et al. 2003; Qin et al. 2003a,b), identified by the isolation of T-DNA flanking sequences, suggest a random distribution of the T-DNA insertions among the five chromosomes. This is perhaps not surprising, considering the differences between the genomes of cereals like barley and Arabidopsis. As previously stated, barley is thought to have large noncoding regions interspersed with gene-rich regions. The much smaller Arabidopsis genome consists of 85% gene-rich regions (Barakat et al. 1998) compared to 12% in barley, so it would not be expected that any bias would be seen. Forsbach et al. (2003) examined the distribution of T-DNA copies within coding and noncoding regions of the Arabidopsis genome in further detail. They found that although the majority of the insertions were in noncoding regions, a large number of these were found in regions immediately upstream or downstream of coding sequences. In particular, they found that T-DNA insertions in 5' upstream regions were more common than might be expected from a completely random distribution. These results, therefore, support the conclusions of the present study showing that transgene insertion may not be random.

Qin et al. (2003b) analyzed 1194 T-DNA insertion sites in the Arabidopsis genome by matching T-DNA flanking sequence to the Arabidopsis genome sequence. The authors observed “hot regions” that had more insertions than other areas of the genome. This result corresponds to our finding of apparent “hot spots” for transgene insertion within the barley genome. In addition, of the 1194 T-DNA insertion sites mapped, 1010 were inserted in, or close to, genes. This is further strong evidence for a preferential insertion in gene-rich regions.

Barakat et al. (2000) compared the distribution of T-DNA insertions in Arabidopsis and rice by localizing T-DNA in fractions of plant DNA separated according to their GC levels. They found T-DNA in fractions representing 85% of the Arabidopsis genome, whereas it was found only in fractions representing 25% of the genome in rice. The fractions containing the T-DNA matched those corresponding to the gene-rich region in each species. This again supports the finding of the present study that transgenes tend to insert in gene-rich regions.
Sallaud et al. (2003) isolated the region flanking the left border of the T-DNA in 477 transgenic rice plants and comment that the detailed analysis of large populations of T-DNA insertion lines now in progress will provide further information on the preference of T-DNA integration in rice. A recent study by Sha et al. (2004) examined 361 T-DNA flanking regions in rice. These authors also detected preferential insertion of T-DNAs into protein-coding regions and an apparent nonrandom distribution of the T-DNA insertions.

Choi et al. (2002) examined transgene location in 19 barley lines using FISH only. In that study, transgene insertion appeared to be random but transgene insertions were not confirmed by genetic mapping or flanking sequence data. It is therefore possible that these authors did not observe the preferred regions for transgene insertion that we detected following more detailed analysis of insertion sites.

Reports of transgene detection in barley (Pedersen et al. 1997; Choi et al. 2002) and other cereal species (Svitashev et al. 2000) indicate a tendency toward the localization of transgenes in distal (subtelomeric and telomeric) chromosome regions. This was also seen in the present study where 39% of transgene insertions were found in telomeric or subtelomeric regions. This is similar to Choi et al. (2002) who found 21% of insertions in telomeric regions and 10% in subtelomeric regions in barley. In oat, Svitashev et al. (2000) found 18 of 25 integration events in telomeric and subtelomeric regions. These authors suggest that this may be due to the fact that these regions are enriched with coding DNA sequences. This tendency for telomeric and subtelomeric regions to contain more transgene insertions again supports the findings of the present study. If transgene insertions are more likely in coding regions of the genome, then it is likely that some preference for telomeric and subtelomeric regions will be observed in the cereals.

In summary, information on both the physical and the genetic location of transgenes, together with data on their flanking sequences, is an important starting point toward increasing our understanding of transgene integration. Such information may also be of value in examining the effect of insertion site on transgene stability and in looking for possible unintended effects of the transgene insertion. The apparent nonrandom distribution of transgene insertions observed for the transgenic barley lines examined in this study appears to reflect the preference for insertion into gene-rich regions. Data from transgene-flanking sequences provided additional evidence for preferred insertion into gene-rich sites. This preference has also been observed in other plant species although the physical pattern of transgene insertion observed is dependent on the structure of the genome being studied.

We gratefully acknowledge the statistical assistance of Sandro Leidi of the Reading statistical service. This work was supported by the Food Standards Agency of the United Kingdom. H. Salvo-Garrido acknowledges the support of the National Institute of Agriculture Research (INIA), Temuco, Chile, and S. Travella acknowledges the support from the European Commission (Agriculture and Fisheries program).

LITERATURE CITED


Christensen, A. H., and P. H. Quail, 1996 Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res. 5: 213–218.


