Abstract Transposable elements are ubiquitous genomic parasites with an ancient history of coexistence with their hosts. A few cases have emerged recently where these genetic elements have been recruited for normal function in the host organism. We have identified an expressed hobo/Ac/Tam (hAT) family transposase-like gene in cereal grasses which appears to represent such a case. This gene, which we have called gary, is found in one or two copies in barley, two diverged copies in rice and two very similar copies in hexaploid wheat. No gary homologues are found in Arabidopsis. In all three cereal species, an apparently complete 2.5 kb transposase-like open reading frame is present and nucleotide substitution data show evidence for positive selection, yet the predicted gary protein is probably not an active transposase, as judged by the absence of key amino acids required for transposase function. Gary is expressed in wheat and barley spikes and gary cDNA sequences are also found in rice, oat, rye, maize, sorghum and sugarcane. The short inverted terminal repeats, flanked by an eight-nucleotide host sequence duplication, which are characteristic of a hAT transposon are absent. Genetic mapping in barley shows that gary is located on the distal end of the long arm of chromosome 2H. Wheat homologues of gary map to the same approximate location on the wheat group 2 chromosomes by physical bin-mapping and the more closely related of the two rice garys maps to the syntenic location near the bottom of rice chromosome 4. These data suggest that gary has resided in a single genomic location for at least 60 Myr and has lost the ability to transpose, yet expresses a transposase-related protein that is being conserved under host selection. We propose that the gary transposase-like gene has been recruited by the cereal grasses for an unknown function.

Keywords Transposon · Genome evolution · Ac · Transposable element · Gene evolution · Transposon domestication

Introduction

Transposable elements (TEs) are an ancient and ubiquitous component of genomes (Craig et al. 2002). TEs have been divided into Class I elements, which transpose to new genomic locations via RNA transposition intermediates, and Class II elements, which use DNA intermediates for transposition (Finnegan 1990). Each class is sub-divided into a number of ‘superfamilies’, which share common structures, sequence homologies and detailed transposition mechanisms. The first TEs to be described originated from McClintock’s studies of ‘controlling elements’ in maize (McClintock 1948, 1951). The two major TE systems studied by McClintock,
Ac/Ds and Spm/dSpm have become prototypes for the hobo/Ac/Tam (hAT) and CACTA superfamilies of Class II TEs, respectively. Members of the hAT superfamily, which this work concerns, have since been observed in fungi, animals and plants (Calvi et al. 1991; Rubin et al. 2001).

McClintock’s observation, that the somatic effects of controlling elements follow developmental programing, prompted her to speculate that these genetic elements operated to control development in maize (McClintock 1951). Her studies also clearly demonstrated that these TEs were capable of causing chromosome breaks and translocations. The issue of whether TEs are beneficial, neutral or harmful has been controversial ever since. An extensive body of evidence has accumulated demonstrating a variety of damaging mutations that TEs can cause to their hosts, either directly via their insertions in or near genes or indirectly via unequal genetic exchange between duplicated TEs, which leads to a variety of lesions in the surrounding DNA such as duplications, inversions, translocations and deletions (Craig et al. 2002).

Conversely, a small but increasing number of cases are accumulating where TEs or their component gene(s) can apparently be beneficial to their host. The first of these to be discovered was the recruitment of the Class I LINE retrotransposons HeT-A and TART to preserve the telomeres of Drosophila (Biestaussmann et al. 1992; Levis et al. 1993). More recently, other cases have emerged in animal and diatom systems (Agrawal et al. 1998; Lynch and Tistem 2003; Kapitonov and Jurka 2004, 2005; Gao and Voytas 2005), and in plants two cases have emerged recently. First, the Arabidopsis FARI and HY3 genes that affect the far-red light response are members of the Maudr superfamily (Hudson et al. 2003). Second, the Arabidopsis DAYSLEEPER gene, a putative transcription factor which binds to Kubox1 motifs and is essential for plant development, is a member of the hAT superfamily (Bundock and Houseykaas 2005). In all of these cases the corresponding protein coding regions appear to be intact and, for all tested cases, under positive selection. Additionally, the cis DNA elements required for transposition (long terminal repeats or terminal inverted repeats) are absent. This suggests strongly that all these genes encode functional proteins, which are no longer a part of TEs, and have been recruited by their respective hosts.

Plant genomes typically contain multiple different members of the hAT superfamily (Xu and Dooner 2005). In maize, there are at least three distinct types of hAT element in addition to Ac, rice has at least four, and Arabidopsis thaliana at least five. Some of these form loose phylogenetic clades, suggesting an ancient diversification of the superfamily before the monocot–dicot split (Xu and Dooner 2005). At least one pair of maize hAT transposons do not interact genetically with each other, implying that the hAT superfamily members have evolved to act independently of each other. Each hAT transposon is typically present at between 1 and 50 copies per genome in locations that are highly polymorphic within the species (Kunze et al. 1997).

This work describes the existence of an expressed, apparently functional gene, gary, related to the hAT transposase gene family, that is highly conserved within cereal grass genomes and not found in Arabidopsis. Consistent with the examples described above, this gene lacks the cis sequences required for transposition and appears not to have transposed for at least 60 Myr, suggesting that it has been recruited by the cereal grass genome for an as yet unknown function.

Results

Isolation of gary from barley

The barley gary sequence (b-gary) was first detected as an expressed sequence tag (EST; accession BI948834) in cDNA libraries of Fusarium graminearum–infected barley and wheat. To investigate the copy number and insertional polymorphism of b-gary in the barley genome, DNA gel blot analysis was carried out using a wide variety of wild and cultivated barley germplasm (Fig. 1). Surprisingly, only one or two gary-specific bands were seen per plant and relatively little polymorphism for a supposed TE is apparent across such a wide diversity of germplasm. A screen of 73,728 bacterial artificial chromosome (BAC) clones, prepared from cv Morex, representing approximately 1.5 haploid barley genomes yielded two clones homologous to b-gary. Both clones contain the 9.0 kb HindIII gary-specific restriction fragment visible in lane 9 in Fig. 1 and both share very similar HindIII restriction fragment patterns (data not shown), suggesting that they derive from the same locus. Collectively, these data suggest that b-gary is present in one or two copies in barley.

The gary-containing HindIII restriction fragment was subcloned and the b-gary-containing region sequenced. The subclone contains a 2.5-kb open reading frame (ORF) interrupted by a single intron, deduced from

![Fig. 1 Southern blot analysis of b-gary in diverse barley germplasm. HindIII digested barley DNAs used were 1 B83 (Thomas et al. 1998), 2 ICB 180049, 3 cv Lina, 4 cv Cyrrhus, 5 ICB 180049, 6 ICB 181277, 7 cv Haruna, 8 cv Haruna, 9 cv Morex, 10 ICB 180044, 11 cv Chime, 12 ICB 180994, 13 Mt. meron, 14 cv Derkado, 15 cv Steptoe, 16 cv Galleon 2. Marker sizes in kb are shown](image)
Identification of gary homologues from other cereal grasses and the relationship of the gary family to other hAT transposases

Searches of the complete rice genome sequences (indica and japonica) revealed two b-gary homologues, both containing apparently full-length ORFs (Figs. 2, 3). One of these two homologues, which we have called r-gary1 displays closer homology to the b-gary ORF than the other (70 and 58% predicted amino acid identities, respectively). The two r-garys share the C-terminal truncation and mutated DDE motifs with b-gary.

Further database searches against EST libraries showed close homologues of b-gary for wheat, maize, sorghum, rye and sugarcane (data not shown). To isolate wheat gary homologues, a pooled BAC library from Triticum aestivum, cv Renan (Chalhoub et al. 2003) was screened by PCR (Experimental procedures). Five BAC clones were obtained from ca. 10⁶ clones, representing approximately seven haploid T. aestivum genome equivalents. Diagnostic PCR with b-gary primers showed that these represent two different gary homologues (data not shown). The genomic regions containing both wheat garys (w-gary1 and w-gary2) were amplified from the corresponding BAC clones by PCR and their sequences determined. The two w-garys contain ORFs of very similar size to b-gary and the predicted peptides are 90–95% identical, both with each other and with b-gary, including the SVE motif and C-terminal truncation (Supplementary Fig. 1). Comparing b-gary with either w-gary gives ratios of synonymous to non-synonymous substitutions (Ks/Ka) between 3.0 and 3.1 and between the two w-garys the ratio is 2.84, indicating that the coding sequences of the Triticeae garys are under positive selection.

All five complete sequences available for gary from barley, wheat and rice were compared against other members of the hAT transposase superfamily from plants (Fig. 4). The gary sequences form a clade that is quite far diverged from the other hAT superfamily.

**Fig. 2** Structures of gary genes of barley and rice. Flanking host TSDs and inverted repeats of hAT TEs, exemplified by Ac of maize are shown by closed and open triangles, respectively. Positions of translational start (ATG) and stop (TAA and TAG) codons of long ORFs (dark grey boxes) are shown. Introns (horizontal shaded boxes), 5' and 3' untranslated regions (light grey boxes) and polyadenylation sites (A⁺) are also shown. The location of repetitious DNAs flanking b-gary is shown by black boxes and the positions of PCR primers used for Fig. 5 are shown as arrows. DNAs of unknown function are unshaded.
transposases. The closest relative to the gary clade is DAYSLEEPER but the homology (23% amino acid identity to \( b \)-gary) is similar to that between far-diverged members of the hAT superfamily (for example, \( Ac \) shares 25 and 23% amino acid identity with Tam3 and Mx, respectively). The gary clade thus represents a new divergent member of the hAT superfamily. Evidence for immobility of gary in the cereal genome

The low copy number and low DNA gel blot polymorphism in diverse barley germplasm implies that \( b \)-gary might be immobile in the barley genome. To explore this possibility, we examined the upstream boundary of \( b \)-gary for the 16 diverse germplasm samples used in

![Fig. 3 Alignment of predicted proteins encoded by gary genes of barley and rice with the Ac transposase. Amino acids shared by four, three and two of the polypeptides are boxed in black, dark grey and light grey, respectively. The positions of the three amino acids comprising the conserved DDE motif which is required for transposase function (Zhou et al. 2004) are arrowed and the six conserved regions A–F identified by Rubin et al. (2001) are shown by lines](image)
Fig. 1. The h-gary sequence from the barley BAC subclone is flanked by an internal int gene fragment of the Wis-2-1A retrotransposon (Fig. 2). PCR analysis of the 16 barley DNAs, using a primer pair from within the gary ORF and the Wis-2-1A fragment, respectively, produced a similarly sized PCR product (Fig. 5) and sequence analysis verified that all 16 sequences have the int/b-gary junction (Supplementary Fig. 2). These sequences show a high level of polymorphism, both SNP and indel, around the junction region, which is consistent with the hypothesis that this sequence arrangement is ancient.

A second experiment involved DNA gel blot analysis on transgenic wheat and barley plants to search for transposition during the tissue culture and transformation conditions, which promote transposition for other TEs [reviewed by (Wessler 1996)]. Forty-two wheat and ten barley transgenic lines were examined. All plants exhibited the same banding patterns as the non-transformed parental plants (examples shown in Supplementary Fig. 3), showing that these genomic stress conditions do not induce detectable transposition of gary in barley or wheat.

Transposition-proficient hAT TEs have conserved inverted sequence duplications of ca. 8–14 nucleotides at their termini (Kunze et al. 1997; Rubin et al. 2001), flanked by terminal sequence duplications (TSDs) of 8 bp of the host target site (these are indicated by open and closed triangles, respectively, in Fig. 2). The inverted repeats are absolutely required for transposition (Hehl and Baker 1989; Healey et al. 1993). Searches for such motifs within and between the barley and rice gary sequences flanking the transcription units were unsuccessful (Experimental procedures), further suggesting that the barley and rice garys are incapable of transposition.

The evidence that gary may be immobile in the barley genome raised the possibility that it might share a common genomic location between different cereal species. Therefore, the chromosomal locations of gary in barley, wheat and rice were determined. For b-gary, the restriction fragment length polymorphism data (Fig. 1) allowed the determination of its genetic map position in a mapping population between cvs Morex and Steptoe (the parental DNAs are lanes 9 and 15, respectively, in Fig. 1). The deduced genetic map location for b-gary is at the distal end of chromosome 2H in bins 14–15 (BIN location based on WSU Barley genomics website http://www.barleygenomics.wsu.edu/; 122 cM, in a 158 cM map, Fig. 6).

The approximate positions of wheat garys on the physical map of T. aestivum were obtained by physical BIN mapping using chromosome segment deletion lines (Experimental procedures). Three positions were obtained, corresponding to the three hexaploid wheat progenitor genomes and all are in the same approximate genomic location, namely the long arm of chromosome 2 (0–0.85, 0.89–1.00 and 0.76–1.00 for wheat 2AL, 2BL and 2DL, respectively, where 0 corresponds to the centromere and 1.00 corresponds to the telomere).

These physical map positions correspond well with the genetic map position for b-gary, with the caveat that the two maps are based upon different analytical methodologies (cytogenetics vs. recombination frequency) and the physical bin sizes are significant fractions of entire chromosome arms. Last, the physical map locations of the two r-garys in the rice genome were deduced from the complete sequence of rice Oryza sativa, cv. Nipponbare (japonica type). r-gary1, the closer of the two rice homologues of b-gary, is located at ca. nucleotide position 31.5 Mbp on rice chromosome 4, approximately 4.2 Mbp from the telomere. Rice chromosome 4 is syntenic to the distal regions of chromosome 2 of wheat and barley (Moore et al. 1995) and this position corresponds to the distal end of the long arm of chromosome 2 for these species (Fig. 6). r-gary2 is located at ca. 4.70 Mbp, towards the top of rice chromosome 8. Collectively, these data are consistent with the hypothesis that b-gary, w-gary1 and w-gary2 and r-gary1 are orthologous, immobile genetic elements occupying...
syntenic locations in the three corresponding cereal grass genomes and \( r\text{-}g\text{ary}2 \) is a paralogous element.

Attempts were made to search for microsynteny between the \( g\text{ary} \)-containing regions of barley, wheat and rice. The two BAC clones containing \( b\text{-}g\text{ary} \) were sample-sequenced at ca. twofold coverage to identify genes flanking \( b\text{-}g\text{ary} \). BLAST and TBLASTX searches with all these sequences identified three candidate genes (genes 24, 45 and 65), plus \( b\text{-}g\text{ary} \) itself. PCR screens of the \( w\text{-}g\text{ary} \)-containing wheat BACs, using primers specific for genes 24, 45 and 65 were unsuccessful, suggesting that the wheat orthologues of these genes are not in the close vicinity of \( w\text{-}g\text{ary}1 \) or \( w\text{-}g\text{ary}2 \). BIN mapping of genes 24, 45 and 65 in wheat suggests that genes 24 and 65 have undergone duplication and local translocation on chromosome 2D and gene 45 has experienced a duplication/local translocation on chromosome 2B (Supplementary Table 1). These data suggest that the genomic region containing \( b\text{-}g\text{ary} \) has been rather unstable in the Triticeae, with duplications, deletions and local translocation disrupting microsynteny. Next, searches were carried out to identify the genomic positions of the rice orthologues for genes 24, 45 and 65. Gene 24 produced multiple high homology rice hits (> 85% predicted amino acid identity), none of which map close to \( r\text{-}g\text{ary}1 \) on the rice genome. This gene encodes a member of the lipid transfer protein multi-gene family. BLAST searching with gene 45 identified a single rice homologue with 87% predicted amino acid identity, which maps 0.9 Mbp from \( r\text{-}g\text{ary}1 \) on chromosome 4 (2.6% of the chromosome length). Searching with gene 65 identified a single 87% identical sequence on rice chromosome 1. In summary, only one out of three genes which flank \( b\text{-}g\text{ary} \) in barley is found in the genomic vicinity of \( r\text{-}g\text{ary}1 \) in rice.

Finally, the 20 closest predicted genes flanking \( r\text{-}g\text{ary}1 \) in the rice genome (ten each side) were screened for their corresponding cytogenetic bin-map locations, to test the general level of synteny between rice and wheat for the region containing \( r\text{-}g\text{ary}1 \). Three of these predicted genes have been bin mapped in wheat as part of the International Triticeae EST Cooperative Mapping Project (http://www.wheat.pw.usda.gov/mgaze/; Supplementary Table 1). All three cytogenetic locations are

![Fig. 6 Schematic comparison of genomic locations of \textit{gary} in barley, wheat and rice. Genetic map position of \textit{b-gary} to chromosome 2 in barley Steptoe × Morex mapping population, cytogenetic BIN-map locations on chromosomes 2 for \textit{gary} in wheat chromosome deletion lines and physical location of \textit{r-gary}1 on the sequence of the syntenic rice chromosome 4. The approximate correspondence between rice chromosome 4 and wheat/barley chromosomes 2 is shown by dotted lines. Chromosome sizes are not to scale.](image-url)
on the long arm of chromosome 2, with five of the eight bin-map locations identical to those of *gary* in wheat (one gene failed to map to the A genome). In summary, the genomic region containing *gary* is somewhat labile between rice, barley and wheat, with only some genes retaining their genomic locations, and these data alone cannot tell us whether *gary* has occupied the same genomic position since the divergence of rice and the Triticeae species.

Expression of *b-gary* in barley

The presence of numerous *gary* cDNA sequences in EST collections suggests that *gary* is transcribed in the cereal grasses. To gain further information of the RNAs encoded by *b-gary* and to investigate the response of the element to stress, a Northern gel blot of RNAs isolated from barley spikes, either inoculated with *Fusarium graminearum* or water, was probed with *b-gary* (Fig. 7). The results show equivalent *b-gary* RNA accumulation in both *Fusarium*-infected and water control plants. We conclude that *b-gary* is expressed constitutively in barley spikes and that expression is not increased by *Fusarium* inoculation.

### Discussion

Evidence for immobility of *gary* in the cereal genome

Several lines of evidence point to the conclusion that *gary* is immobile in the cereal grass genome. First, the Southern data (Fig. 1) and upstream flanking PCR (Fig. 5) show no evidence for insertional polymorphism among 16 very diverse *Hordeum* genotypes, including five wild *H. spontaneums*. Bona fide plant *hAT* TEs show much higher polymorphism levels across wide germplasm (Federoff et al. 1983; Xu and Dooner 2005). It is possible that *b-gary* is polymorphic in other barley samples, but the high level of sequence polymorphism at the junction with flanking repetitious DNA (Supplementary Fig. 2) shows that this arrangement is ancient. The absence of observable insertional polymorphism following passage through cell culture (Supplementary Fig. 3), which leads to derepression of transposition in many TE systems (Wessler 1996) supports the view that *gary* transposition cannot be remobilised.

Further evidence supporting the hypothesis of *gary* immobility comes from the observation that *b-gary* in barley and its close homologues in wheat and rice share the same approximate chromosomal position. All of these localisation estimations are very approximate but the coincidence is compelling and suggests that *gary* may have been immobile in the cereal grass genome for at least 60 Myr. Intriguingly, the two rice *gary*s occupy regions which may have been duplicated from a single region before the origin of the grasses (Yu et al. 2005), suggesting an even greater antiquity for *gary*s immobility. No homologue for *gary* has been found in the *Arabidopsis*, *Medicago* or *Lotus* sequence databases (data not shown), so it is likely that *gary* became recruited after the emergence of the monocot lineage. It is unclear at present how widespread *gary* is within other monocots.

Unfortunately, the microsynteny analysis provides insufficient evidence to prove this hypothesis. This is not particularly surprising. Cereal genomes are highly variable with regard to the amount of repetitious DNAs interspersed between genes (SanMiguel et al. 1996; Shirasu et al. 2000; Li and Gill 2002; Chantret et al. 2005) and genes which are adjacent to each other in one species may be separated by large distances in another. Furthermore, microsynteny between rice and the Triticeae is uneven. Nineteen per cent of wheat ESTs, that map to rice chromosome 4, BIN map to wheat chromosomes other than group 2 chromosomes (Sorrels et al. 2003), 23% of barley ESTs from chromosome 2 map to ‘non-syntenic’ chromosomes (Rostoks et al. 2005) and inter-chromosomal translocation of a homoeologous region to a non-homoeologous chromosome has been observed (Li and Gill 2002).

![Fig. 7](image.png)

**Fig. 7** Northern gel blot analysis of *gary* RNAs in barley spikes. Barley spikes from a cv. Chevron-derived barley line and a cv. Fredrickson-derived breeding line were inoculated with either water (lanes 1–2, 5–6) or *Fusarium graminearum* (lanes 3–4, 7–8). RNAs were isolated 48 h (lanes 1, 3, 5, 7) or 96 h (lanes 2, 4, 6, 8) following inoculation. The blot was first photographed under UV illumination to visualise rRNA (gel loading control; *lower frame*) then probed with *b-gary* (*upper frame*).

The structure and function of the *gary* protein

All five *gary* sequences described here appear to encode nearly full-length *hAT* transposase-related proteins. These polypeptide sequences group together into a clade that is quite far diverged from the other *hAT* family transposases (Fig. 4). There are several large indels between *Ac* and *gary* in the region corresponding to amino acids 189–216 of *Ac* (Kunze et al. 1997) and might point to an altered DNA binding specificity for *gary*. Additionally, all *gary*s studied have a C-terminal truncation of 55 amino acids, relative to *Ac*. This tail is also present in other *hAT* transposases of plants, including *Mx* and *Tam3* (Xu and Dooner 2005) but is missing in *hobo* of *Drosophila* (data not shown). The C
terminus contains region 3 of Kunze et al. (1997), which is involved in oligomerisation to form the active transposase. Dimerisation and transposition studies by Essers et al. (2000) have shown that this domain is located between residues 674 and 754 of Ac. All of the sequenced garys are quite closely homologous to Ac up to position 750, where they terminate. It therefore seems likely that the gary proteins are capable of oligomerisation but they may lack another function or subfunction of this activity. All six conserved regions identified by Rubin et al. (2001) are present in all sequenced garys (Fig. 3). Additionally, other regions of the hAT transposases, with no functions identified to date, show strong conservation between all garys and Ac. In particular, large parts of the predicted protein sequence between regions 2 and 3 of Rubin et al. (2001) are well conserved (Fig. 5).

Gary is evolving as a low copy gene under positive selection in many and perhaps all cereal genomes. This suggests that it is fulfilling a function in these species—what might this be? It is unlikely that the basic hAT transposase function of DNA strand breakage and recombination is preserved in the gary proteins, because two or more of the three critical amino acids comprising the DDE motif (Zhou et al. 2004) are absent (Fig. 3). Several other possibilities exist. First, gary might act as a repressor of unwanted transposition in cereals. There are clear precedents for this, particularly in Drosophila (Río 1990; Hartl et al. 1997) but in these systems the only elements being repressed are the corresponding non-defective elements themselves. As gary is a low copy number TE which apparently lacks the cis-DNA components required for transposition this seems an unlikely reason to keep it; in fact, a simple knockout mutation to gary would make the entire system superfluous.

A related possibility is that gary is repressing transposition of other hAT TEs. There is no available evidence which addresses cross-element repression but there is information on cross-element mobilisation. Two members of the hAT superfamily, Hermes and hobo can mobilise each other’s transposition in insects (Sundararajan et al. 1999), suggesting that cross-interaction among different but related hAT transposases is possible. In contrast, the maize Mx and Ac elements do not induce each other’s transposition (Xu and Dooner 2005). This discrepancy may be explained by the fact that Hermes and hobo are 55% identical at the protein level, whereas Mx and Ac are 23% identical. The non-gary hAT elements of rice are between 16 and 21% identical to r-gary1 (data not shown), suggesting that cross-interactions are unlikely but there remains the testable possibility that gary can inhibit transposition of other hAT superfamily members via the formation of inactive multimers. It is particularly intriguing that the dimerisation region, corresponding roughly to regions D–F of Rubin et al. (2001) (Fig. 3), is among the best conserved regions for the seven rice hAT superfamily members in Fig. 4 (data not shown).

Other possibilities for gary’s function involve a departure from the simple transposition process to fulfill a different need in the host, as illustrated in Arabidopsis by the DAYSLEEPER gene (Bundock and Hooykaas 2005), which probably acts as a transcription factor. If gary has evolved to perform a function unrelated to transposition for the cereal grasses, the key to understanding will come from studies of gary mutants and the identification of the DNA target(s) for the recombination events catalysed by gary. No mutant is yet available from the rice insertional mutagenesis program (http://www.genoplante-info.infobiogen.fr/OryzaTagLine/insertion.htm) and it will be very interesting to observe the phenotype of such a mutant when it is found.

Expression of gary

Gary is clearly transcribed in many cereal grasses, as judged by EST entries. Only 59 gary-homologous EST sequences ($e < 0.1$) were obtained by BLAST searching ca. 2.2 M cereal grass EST sequences (http://www.ncbi.nlm.nih.gov/dbEST/). This represents ca. 0.0025% of the total EST population and shows that gary is expressed either with a narrow tissue and/or developmental specificity or at low levels. The former possibility is plausible, since for each species only a limited spectrum of tissues has yielded gary ESTs (for example, only spike and unclassified barley EST libraries have yielded gary ESTs to date). Our expression studies have been restricted to spikes and there is a clear need to further explore the specificity of gary’s expression in the cereal grasses.

Conclusion

These studies show that gary, a member of the hAT transposase gene superfamily, is immobile in the cereal grass genome. In barley, $b$-gary shows little or no insertional polymorphism and no evidence can be found in barley or rice for the flanking sequence motifs which are conserved among hAT family members and required for transposition (terminal inverted repeats) or indicative of recent transposition (8 bp flanking TSD). All predicted gary transposases lack amino acids that are absolutely required for transposase function. Furthermore, gary orthologues in barley, wheat and rice all map to approximately the same chromosomal position. Collectively, these data suggest strongly that gary has been immobile in the cereal grass genome since before the divergence of rice and the Triticeae ca. 60 Myr ago. Despite this, gary is present in one or two highly conserved and transcribed copies in a variety of cereal grass genomes and the encoded protein sequence shows evidence for positive selection, suggesting that the gary gene has been recruited by the cereal grass genome for an as yet unknown function.
Experimental procedures

Plant materials

Barley accessions for Southern analysis (see Fig. 1) were gifts from Luke Ramsay (Scottish Crop Research Institute). ICB accessions in Fig. 1 are from the ICARDA Hordeum spontaneum collection, Syria (http://www.icarda.org/). Barley lines for northern analysis (Fig. 7) were (1) a cv. Chevron-derived chromosome 2 (BIN 8) fusarium head blight (FHB) resistance Quantitative Trait Locus (QTL) (de la Pena et al. 1999), backcrossed five times into University of Minnesota M69 breeding line; (2) a breeding line that carries the cv. Fredrickson-derived chromosome 3 (BIN 6) FHB mycotoxin-susceptible allele for the QTL (Smith et al. 2004).

DNA and RNA methods

Genomic barley DNAs were isolated from fresh leaf tissues using a modified urea extraction procedure described in Caldwell et al. (2004). Preparation of BAC DNAs was by Phase Prep™ BAC DNA kit (Sigma). All PCRs used Qiangen Hotstart Taq DNA polymerase in unmodified Qiagen buffer (no extra Magnesium or Q buffer) and 0.2 pmol/ul of each primer. Hot-start PCR conditions were 95°C 15 min then 30x [94, 55, 72°C, all 1 min] then 72°C 7 min. All primers were designed for melting temperatures of 55–65°C in 50 mM cation concentration. For sequencing, PCR fragments were purified through NucleoFast® 96 PCR plates. Primers used for Fig. 5 were CTTGTTCATGATTTGCAGCC (Wis-2-IA) and CGACCTGCGGGATCACAGCA (b-gary), see Fig. 2 for locations.

RNAs were isolated from barley spikes inoculated either with F. graminearum in 0.04% (v/v) Tween 20 or mock inoculated with water plus Tween 20. Three biological replicates of the experiment were conducted at 24, 48, 72, 96 and 144 h after infection. Eight spikes per time point were flash frozen in liquid nitrogen and stored at −80°C. The pooled spikes were ground in liquid nitrogen and RNA was extracted from 1 g of tissue using the Trizol™ (Invitrogen, Carlsbad, CA, USA) protocol. Contaminating DNA was destroyed using RNase-free DNase and the RNA was purified on RNeasy® columns (Promega, Madison, WI, USA).

Ten micrograms of each RNA was electrophoresed on 1.2% formaldehyde-agarose gel and transferred to Hybond-N+ membrane (Amersham Pharmacia Biotech, Piscataway, NJ, USA). DNA probes were labelled with 32P-dATP using Prime-a-Gene Labeling System following the manufacturer’s instructions (Promega). Pre-hybridisations and hybridisations were carried out using the modified Church and Gilbert buffer (Church and Gilbert 1984) at 65°C. Washes were conducted in 1x SSC, 0.1% SDS at 65°C. Membranes were exposed to a phosphor screen for 24 h and scanned by a Phosphor-Imager (Molecular Dynamics, Inc., Sunnyvale, CA, USA).

HindIII-digested BAC DNAs were subcloned into HindIII digested plasmid. DNA gel blot analysis was conducted according to Sambrook et al. (1989).

DNA sequencing

Sample sequencing of BACs containing b-gary involved shearing the BAC DNA using a nebuliser. The nebulised DNA was A-tailed and cloned into PGM T-easy vector (Promega). The inserts from 600 randomly picked colonies were PCR amplified and sequenced by BigDye™ v 3.0 automated sequencing (PE Biosystems), yielding ca. 300,000 bp of total sequence information.

Mapping genomic locations for gary

Physical map locations for gary in wheat were obtained by DNA gel blot analysis of the wheat deletion lines as described in Conley et al. (2004). Wheat physical map locations for rice genes flanking r-gary1 were obtained by BLAST searching for BIN-mapped ESTs at the International Triticeae EST Cooperative website (http://www.wheat.pw.usda.gov/wEST/blast/). Rice genomic analysis for gary and flanking genes was carried out by BLAST analysis at the Gramene website (http://www.gramene.org/).

Bioinformatics

Searches for terminal inverted repeats motifs within and between the barley, wheat and rice gary genomic sequences were carried out using two independent programs, namely DOTTER (http://www.cgb.ki.se/egb/groups/sonnhammer/Dotter.html) and einverted (http://www.emboss.sourceforge.net/). Screening parameters were first optimised using genomic DNAs containing known hAT superfamily TEs. Regions screened included the entire region of b-gary between the flanking repetitive DNA and the transcribed region, 1,000 bp both upstream and downstream of the transcribed region for r-gary1, 2,382 bp upstream and 2,100 bp downstream of the r-gary2 ORF (this roughly equates to the region screened for r-gary1) and 897 bp downstream of the polyadenylation site of w-gary1. Einverted used gap penalty of 9, minimum score threshold of 15, match score of 3 and mismatch score of −4. Every inverted repeat pair was analysed visually for 8 bp flanking direct repeats, or degenerate derivatives thereof. BLAST searches were carried out at the NCBI (http://www.ncbi.nih.gov/BLAST/) or ITEC (http://www.wheat.pw.usda.gov/wEST/blast/) sites, using default parameters. Ratios of synonymous to non-synonymous nucleotide substitutions were calculated using DIVERGE.
Accession numbers for *b-gary*, *w-gary1*, *w-gary2* and genes 24, 45 and 65 are AM087608–AM087613, respectively. The ORF of *r-gary1* is nucleotides 35,392–32,805 of accession AK166298 and a full-length cDNA for *r-gary1* is also available (Ak066078). The ORF of *r-gary2* is nucleotides 65,383–67,799 of accession Ap004692.

Acknowledgements GJM was supported by a McKnight Landgrant Professorship for sabbatical leave to the Scottish Crop Research Institute. BSB was supported by a Boycast Fellowship from the Indian government. We thank Maggie Knox for help with performing *Kc/Kc* determinations, the Genoplante consortium (http://www.genoplante.com) for making available the haplobloc DNA library, Emily Conley and Jim Anderson for help with deletion mapping and Robbie Waugh and Noel Ellis for many useful discussions.

References


References


