Copyright © 2000 by the Genetics Society of America

Retrotransposon Evolution in Diverse Plant Genomes

Tim Langdon,* Charlotte Seago,* Michael Mende,† Michael Leggett,† Huw Thomas,† John W. Forster,*,‡ Howard Thomas,† R. Neil Jones* and Glyn Jenkins*

**Institute of Biological Science, University of Wales, Aberystwyth SY23 3DD, United Kingdom,* †*Institute of Grassland and Environmental Research, Aberystwyth SY23 3EB, United Kingdom and* ‡*Plant Biotechnology Centre, Agriculture Victoria, La Trobe University, Victoria 3083, Australia*

Manuscript received August 10, 1999
Accepted for publication May 1, 2000

ABSTRACT

Retrotransposon or retrotransposon-like sequences have been reported to be conserved components of cereal centromeres. Here we show that the published sequences are derived from a single conventional Ty3-gypsy family or a nonautonomous derivative. Both autonomous and nonautonomous elements are likely to have colonized Poaceae centromeres at the time of a common ancestor but have been maintained since by active retrotransposition. The retrotransposon family is also present at a lower copy number in the Arabidopsis genome, where it shows less pronounced localization. The history of the family in the two types of genome provides an interesting contrast between "boom and bust" and persistent evolutionary patterns.

RETROTRANSPOSONS represent a complex fraction of the repetitive DNA of most eukaryotes. The long terminal repeat (LTR) retrotransposons, in particular, have a high degree of autonomy and encode at least five distinct protein components required for their movement in the genome (Grandbastien 1992). For these elements, movement involves colonization of new genomic sites by intact copies of a parent that is not destroyed in the process, so that progressive rounds of retrotransposition have the potential to lead to massive amplification of elements. Some families may also encode an additional protein allowing movement between cells or even individuals, further blurring the distinction between repetitive DNAs and infective agents (Kim *et al.* 1994; Song *et al.* 1994; Wright and Voytas 1998). "Infection" appears to be a particularly apt description of the LTR retrotransposon families of higher plants, where >50% of the genome may be composed of family members (Bennetzen *et al.* 1998). In cereals, where distribution patterns have been best characterized, most LTR retrotransposon copies are relatively recent insertions into intergenic islands (Bennetzen *et al.* 1998). Waves of colonization may occur (San Miguel *et al.* 1998); in the Triticeae, BARE (*ba*rley *re*trotransposon) elements or their relatives currently appear to be highly active, composing up to 10% of the genome of barley, in which transcription and line-specific insertion site polymorphisms are readily detected (Suoniemi *et al.* 1996a,b; Waugh *et al.* 1997). There is currently some debate about the fate of genomes in which such aggressive colonization occurs. Phylogenetic analysis suggests that cereal and grass genomes have shown progressive inflation that may reflect retrotransposon amplification, at least in part (Bennetzen and Kellogg 1997); however, retrotransposon families represented only by ancient and decaying members have been reported for Drosophila and Arabidopsis, suggesting that new recruitment may eventually be blocked, followed by gradual erosion of existing sequences by deletion (Petrov 1997).

The fate of retrotransposons is inextricably linked to that of their hosts, and adaptations that minimize or even alleviate host genome disruption may be expected to evolve. The preferential distribution of yeast Ty elements to "silent" chromosomal regions is well known (Boeke and Devine 1998) and may be a strategy adopted by a number of retrotransposon families. Subtelomeric or telomeric regions appear to be particularly favorable refuges; in some cases, retroelement adaptation to these niches may have been efficient enough to augment or even replace conventional telomerase activity (Pardue *et al.* 1996). A LINE family has targeted telomeric repeats in Chlorella (Higashiyama *et al.* 1997); the concentration of Ty1-*copia* sequences in Allium subtelomeric domains may also reflect active targeting (Pearce *et al.* 1996). More localized targeting has also been inferred for BARE elements, which show nonrandom distribution of insertion sites into other members of the same family (Suoniemi *et al.* 1997). Recently, a number of groups have reported that cereal centromeres contain a high density of specific "retrotransposon-like" sequences (Ananiev *et al.* 1998; Dong

Corresponding author: Glyn Jenkins, Institute of Biological Sciences, Cledwyn Building, University of Wales, Aberystwyth, Penglais, Aberystwyth, Ceredigion SY23 3DD, United Kingdom.
E-mail: gmj@aber.ac.uk

et al. 1998; Miller *et al.* 1998a; Noutoshi *et al.* 1998; Presting *et al.* 1998). Individual clones, isolated on the basis of similarities to two "universal" cereal centromeric sequences identified in random genomic screens of Brachypodium (CCS1; Aragon Alcaide *et al.* 1996) and sorghum (pSau3A9; Jiang *et al.* 1996), contain degenerate retrotransposon reading frames and conserved LTR fragments. Both the centromeric location and conservation of these sequences indicate that the evolution of these sequences is of interest. It is possible that retrotransposon fragment(s) have been co-opted in centromere function in a similar way to telomeric elements, so that their presence is favored (Miller *et al.* 1998a; Presting *et al.* 1998). However, DNA at centromeric sites is typically organized in long arrays of tandem repeats that normally limit the insertion of mobile elements or other complex DNA (Lee *et al.* 1997; Sun *et al.* 1997). The question remains as to whether the retrotransposon-like sequences owe their concentration to active retrotransposition or whether they have been passively amplified.

We have reanalyzed published sequences and find that they fall into two classes. The first represents fragments of a highly conserved Ty3-*gypsy* family, which has a conventional organization and is closely related to a family in the Arabidopsis genome, while the second represents a nonautonomous family that encodes no enzymatic functions. Both classes of element share highly similar LTRs and show characteristics of recent retrotransposition, despite species-specific polymorphisms that indicate independent evolution since the divergence of species dating back to the origin of the Poaceae. The presence of retrotransposon-like sequences in Poaceae centromeres may be best understood as a transient exploitation of a novel host genomic niche. Following a massive initial amplification, the conventional retrotransposon family appears to be undergoing a slow extinction, presumably reflecting both the emergence of new centromeric organization and competition from nonautonomous elements.

MATERIALS AND METHODS

Materials: DNA was isolated from the following plant lines: inbred rye lines maintained by R.N.J. at UWA; Chinese Spring (CS; wheat); SunII (oat); maize, sorghum, grass, and Avena species from germ plasm maintained at IGER; Aegilops species kindly supplied by Dr. Steve Reader, John Innes Centre, Norwich, United Kingdom; wheat (CS)/rye (cv. Imperial) addition lines kindly supplied by Dr. Terry Miller, John Innes Centre, Norwich, United Kingdom.

PCR: All clones were derived from blunt-ended PCR products ligated to pUC19. Amplification of CCS1 junctions (t26 clone) was carried out with rye genomic DNA and oligonucleotide Hi10R (CGRTYGCTAAGGCGCA); cycle conditions were 94° for 40 sec, 43° for 90 sec, 72° for 2 min, for 3 cycles, followed by 30 cycles of 94° for 30 sec, 50° for 30 sec, 72° for 2 min. Reverse transcriptase (RT) amplification was carried out with S14F (GAGATCMWGCGTCARATWCAAGAAATNCT) and BOTYR (GGCATGACAAGCCACTCATA); conditions were 94° for 30 sec, 55° for 40 sec, 72° for 90 sec, for 30 cycles. Lower stringency (48°) annealing was also used with some species in an attempt to obtain more divergent sequences. No such differences were seen, however. Integrase (IN) amplification was carried out with SORGF1 (TKYTGCAGGAAKCGCATG GAGG) and INTR2 (TTTGTCCATCAGTYTGNGGRTG), conditions as for RT. The insert in clone UC6.7 was derived from PCR of rye DNA amplified with AW37 (TATGTKCTKATHTG GTGGGAYCARAT) and BOTYR (96° for 35 sec, followed by seven cycles of 92° for 40 sec, 48° for 60 sec, 68° for 4 min, and 22 cycles of 92° for 40 sec, 48° for 10 sec, 68° for 3 min); that in UC7.12 was derived from maize amplified with REPQS (CCTCAGTCMGATGGMCARACNGA) and UNIHI (AGGKG CCCGATCTTTCGRYGAG) (94° for 1 min, 47° for 2 min, 72° for 5 min, for 1 cycle, followed by five cycles of 94° for 35 sec, 52° for 60 sec, 72° for 3 min, and 25 cycles of 94° for 35 sec, 57° for 40 sec, 72° for 3 min); that in UC8.5 was derived from *Aegilops squarrosa* amplified with REPQS and UNIHI (conditions as for UC7.12).

Genomic consensus sequences were derived from gel-purified templates.

Fluorescence *in situ* hybridization (FISH): Preparation and pretreatment of the cytological preparations and FISH were performed according to published procedures (Heslop-Harrison *et al.* 1991; Leitch and Heslop-Harrison 1992; Pan *et al.* 1992; Zhong *et al.* 1996). Briefly, probe DNA was labeled either with digoxygenin-11-dUTP or biotin-11-dUTP and hybridized to pretreated chromosome preparations overnight at 37° in the presence *inter alia* of 50% deionized formamide in an Omnislide *in situ* hybridization system (Hybaid). Slides were washed stringently in 20% (v/v) formamide in 0.1× SSC at 42° before probe detection with FITC-conjugated antidigoxygenin antibodies or avidin-rhodamine as appropriate. Amplification of the signals was effected either by FITC-conjugated secondary antibodies or by anti-avidin-biotin followed by a second round of avidin-rhodamine binding. The chromosomes were counterstained in 4′,6-diamidino-2-phenylindole (DAPI) and mounted in Vectashield. Fluorescent images were captured by cooled CCD camera, assigned false color, and manipulated uniformly in Adobe Photoshop.

Sequence analysis: Database searches were performed with BLAST and further sequence analysis carried out with the Genetics Computer Group (Madison, WI) programs. Alignments of conserved regions were made using PILEUP and adjusted by eye. Sequence alignments were displayed using GeneDoc (K. B. Nicholas and H. B. Nicholas, distributed by the authors). K_s:K_a ratios were calculated using the GCG program Diverge. Phylogenetic analyses were carried out using the PHYLIP package, as implemented by the HGMP resource of the MRC. Phylogenetic distances were calculated with PROTDIST or DNADIST, and trees were constructed with NEIGHBOR (neighbor-joining method) and drawn with DRAWTREE.

Deletions indicated in Figure 4 are based on alignments of *crwydryn* elements in Arabidopsis bacterial artificial chromosomes (BACs) T9F8, F5K24, F9M13, T17A11, T1O16, and T27D20. Trees in Figure 5 are derived from peptide alignments analyzed with PROTDIST (Kimura method). Trees in Figure 6 are derived from DNA alignments analyzed with DNADIST (maximum-likelihood method, transition/transversion ratio of 1.4 estimated from relevant cereal data set or entire reading frame alignments of Arabidopsis BACs F5K24, F9M13, and T9F8, three category divisions based on codon position variation rates of 1.5:1:3). Full alignments and details are available from the authors.

Estimates of mutation rates and element ages: Synonymous and nonsynonymous substitution rates were calculated using the GCG program Diverge. For calculation of underlying mu-

tation rates, species consensus was used where more than two comparable sequences were available. IN rates calculated using the least divergent of pairs of sequences are also shown (indicated by brackets in Table 1), although these were not used when deriving an average rate for RT or IN regions. Synonymous substitution rates were converted to mutation rate estimates (Table 1) using the following species divergence times: rye/wheat 7.5 million years (my), rye-wheat/oat 20 my, rye-wheat-oat/maize-millet-sorghum-rice 60 my, rye-wheat-oat-rice-maize-millet/bamboo 80 my, rice/sorghum-millet 30 my, maize/millet-sorghum 40 my (Bennetzen and Freeling 1993, 1997). Element ages, representing the time required for nonsynonymous divergence from the relevant species consensus at the relevant (RT or IN) mutation rate, were as follows: rye RT average 1.01 my (22 sequences, range 0–3.3 my), IN average 0.6 my (7, 0–1.2 my); wheat RT 0.7 my (4, 0.5–0.9 my), IN 0.85 my (5, 0.8–1.2 my); oat RT 1.4 my (5, 0.4–2.9 my), IN 1.1 my (6, 0–2.3 my); sorghum RT 0.42 my (7, 0–1.1 my), IN 0.73 (3, 0.7–0.8 my); rice RT 0.22 my (3, 0–0.7 my), IN 0.7 my (2).

RESULTS

A single ancestral retrotransposon family has given rise to a variety of universal cereal centromeric sequences: No intact centromere-specific retrotransposon having a conventional complement of LTRs and retrotransposon reading frames has yet been found in the cereals. It has been suggested that the contemporary centromeric sequences may represent ancient rearrangements that have become fixed in the genomes by accident or by virtue of acquisition of novel function(s) (Miller *et al.* 1998a; Presting *et al.* 1998). We have therefore aligned the published sequences and isolated additional clones to clarify the organization of the centromere-specific sequences. Details of the clones used in this analysis are given in Table 2.

It is clear from the degree of overlapping homology that a single ancestral family has given rise to all the cereal centromere-specific retrotransposon fragments so far identified. We have named this family *crwydryn* (Welsh for "wanderer"; Figure 1). There is no evidence for unconventional additional components and the only rearrangement to be seen in more than one clone is an internal deletion leading to the loss of all enzymatic functions, resulting in elements having only LTRs, 5′ untranslated region (UTR), and a *gag* structural gene fragment, truncated before the canonical RNA-binding motif (Figure 2). The first member of this class to be described was CentA, in maize (Ananiev *et al.* 1998); the same organization is also seen in two elements in the rice clone RCB11 (Nonomura and Kurata 1999), one intact (here designated RCB11-1) and one truncated in cloning (here designated RCB11-2). Both CentA and RCB11-2 are inserted at different positions into Ty3-*gypsy* retrotransposon reading frames; these reading frames are closely related and show little degeneracy (see below), implying recent movement by both of these elements.

Other than the CentA-like deletions, the only unex-

TABLE 1

Mutation rates of RT and IN in the Poaceae using species consensus or least divergent pairs of sequences (in parentheses)

<table>
<tr><td>Rye/wheat</td><td>RT 0.66 × 10^{-8}
IN 0.68 × 10^{-8}</td><td>Oat/rice</td><td>RT 1.48 × 10^{-8}
IN 1.11 × 10^{-8}</td><td>(Wheat/maize</td><td>IN 1.46 × 10^{-8})</td><td>(Rice/millet</td><td>IN 1.47 × 10^{-8})</td></tr>
<tr><td>Rye/oat</td><td>RT 3.00 × 10^{-8}
IN 1.85 × 10^{-8}</td><td>Oat/sorghum</td><td>RT 1.34 × 10^{-8}
IN 1.42 × 10^{-8}</td><td>(Wheat/millet</td><td>IN 1.13 × 10^{-8})</td><td>(Maize/bamboo</td><td>IN 1.39 × 10^{-8})</td></tr>
<tr><td>Rye/rice</td><td>RT 1.70 × 10^{-8}
IN 1.05 × 10^{-8}</td><td>Rice/sorghum</td><td>RT 4.3 × 10^{-8}
IN 1.24 × 10^{-8}</td><td>(Oat/bamboo</td><td>IN 0.94 × 10^{-8})</td><td>(Maize/millet</td><td>IN 0.90 × 10^{-8})</td></tr>
<tr><td>Rye/sorghum</td><td>RT 1.65 × 10^{-8}
IN 1.45 × 10^{-8}</td><td>(Rye/bamboo</td><td>IN 1.15 × 10^{-8})</td><td>(Oat/maize</td><td>IN 1.15 × 10^{-8})</td><td>(Maize/sorghum</td><td>IN 1.58 × 10^{-8})</td></tr>
<tr><td>Wheat/oat</td><td>RT 3.13 × 10^{-8}
IN 2.18 × 10^{-8}</td><td>(Rye/maize</td><td>IN 1.40 × 10^{-8})</td><td>(Oat/millet</td><td>IN 1.33 × 10^{-8})</td><td>(Millet/bamboo</td><td>IN 1.26 × 10^{-8})</td></tr>
<tr><td>Wheat/rice</td><td>RT 1.82 × 10^{-8}
IN 1.27 × 10^{-8}</td><td>(Rye/millet</td><td>IN 1.17 × 10^{-8})</td><td>(Rice/bamboo</td><td>IN 1.00 × 10^{-8})</td><td>(Millet/sorghum</td><td>IN 1.17 × 10^{-8})</td></tr>
<tr><td>Wheat/sorghum</td><td>RT 1.75 × 10^{-8}
IN 1.49 × 10^{-8}</td><td>(Wheat/bamboo</td><td>IN 1.25 × 10^{-8})</td><td>(Rice/maize</td><td>IN 1.53 × 10^{-8})</td><td></td><td></td></tr>
</table>

TABLE 2

Details of the clones used in this analysis

Clone	Species	PCR primers	Reference
IN	(Integrase)	SORGF1 + INTR2	This work
RT	(Reverse transcriptase)	S14F + BOTYR	This work
UC8.5	*Aegilops squarrosa*	REPQS + UNIHI	This work
dea1	*Ananas comosus*	—	ACY12432
F9D12	*Arabidopsis thaliana*	—	AF077407
T6B13	*A. thaliana*	—	AF005398
T9F8	*A. thaliana*	—	AC005561
F5K24	*A. thaliana*	—	AF128395
T27D20	*A. thaliana*	—	AF076274
F9M13	*A. thaliana*	—	AC006267
T1016	*A. thaliana*	—	AC006304
T17A11	*A. thaliana*	—	AC006194
CCS1	*Brachypodium sylvaticum*	—	Aragon-Alcaide *et al.* (1996)
Hi10	*B. sylvaticum*	—	Abbo *et al.* (1995)
cereba	*Hordeum vulgare*	—	Presting *et al.* (1998)
del-1-46	*Lilium henryi*	—	X13886
Tomato	*Lycopersicon esculentum*	—	AF119040
maggy	*Magnaporthe grisea*	—	L35053
RCB11	*Oryza sativa*	—	Nonomura and Kurata (1999)
RCS1	*O. sativa*	—	Miller *et al.* (1998b)
RCH3	*O. sativa*	—	Dong *et al.* (1998)
retrosat	*O. sativa*	—	AF111709
t26	*Secale cereale*	Hi10R	This work
UC6.7	*S. cereale*	AW37 + BOTYR	This work
crw-r39	*S. cereale*	—	Langdon *et al.* (2000)
sau3A9	*Sorghum bicolor*	—	Jiang *et al.* (1996)
retrosor1	*S. bicolor*	—	AF061282
retrosorb	*S. bicolor*	—	AF061282
CentA	*Zea mays*	—	Ananiev *et al.* (1998)
UC7.12	*Z. mays*	REPQS + UNIHI	This work
gypsy	*Z. mays*	—	AF030633

The clones have been listed alphabetically by species.

pected rearrangement seen is in the barley clone described by Presting *et al.* (1998), originally proposed to contain a single element, cereba. Comparison with other clones demonstrates that the clone in fact contains two independent elements, here designated cereba-1 and cereba-2 (Figure 1A). Cereba-1 consists of a very short fragment of the upstream LTR, a long 5′ UTR, and most of the polyprotein; cereba-2 appears to have been a CentA-like element that has suffered insertions into the *gag* region of a sequence of unknown origin (which is not centromere specific, Presting *et al.* 1998) and into both LTRs.

Alignment of clones allows reconstruction of the ancestral *crwydryn* element (Figure 1B). The ancestor is a conventional Ty3-*gypsy* class retrotransposon, with a relatively large 5′ UTR (>1 kb), and a polyprotein reading frame that overlaps the downstream LTR. Both of these features are common to other cereal retrotransposon families (Marillonnet and Wessler 1998) and do not represent unusual adaptations. Other than the rearrangements described above or small (<1 kb) internal deletions, published fragments do not differ in organization from the ancestral sequence. Conservation is greatest between open reading frame fragments (typically >90% peptide identity), but is also seen in 5′ UTR regions. The reading frames interrupted by CentA and RCB11-2 are highly similar to that of cereba-1 and are also expected to be derived from a recent common *crwydryn* ancestor.

Additional clones were isolated to confirm that the ancestral organization of the *crwydryn* element is maintained in a range of species (Figure 1B). PCR of maize and *A. squarrosa* genomic DNA using oligonucleotides based on conserved integrase and CCS1 motifs gave rise to the products predicted for internal retrotransposon fragments (for example, maize clone UC8.5, Figure 1B), confirming that CentA/RCB11-like LTRs are contiguous with cereba-1 polyprotein homologues in these species, while PCR of rye genomic DNA using oligonucleotides based on a rye B-chromosome-specific *gag* fragment (Langdon *et al.* 2000) and a conserved RT motif gave rise to the expected product (clone UC6.7), confirming the

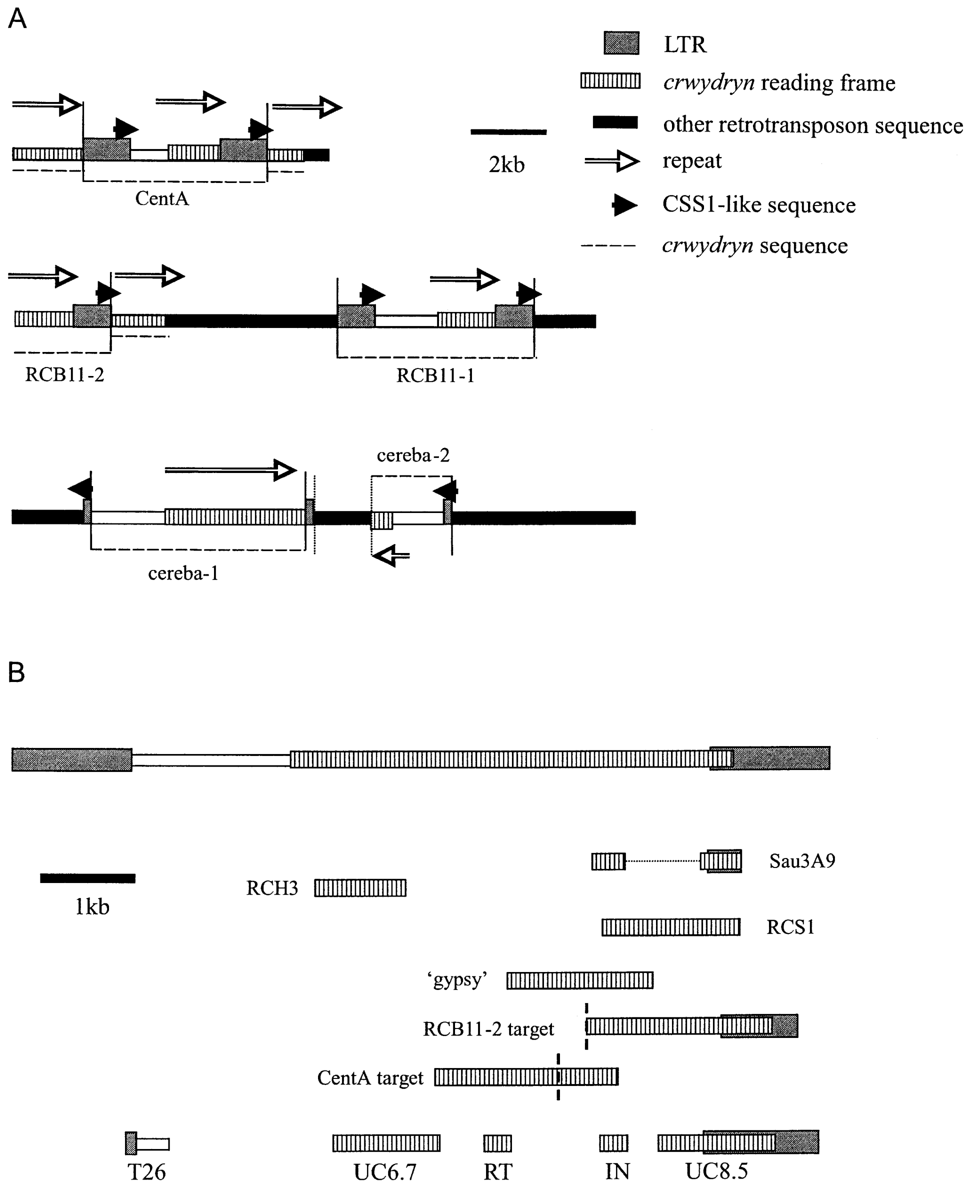

Figure 1.—Relative organization of *crwydryn* elements and fragments. (A) Cereal genomic clones. The CentA repeat of maize is composed of two LTRs flanking a 5′ UTR and partial open reading frame inserted into a second *crwydryn* reading frame. The rice RCB11 clone contains a single CentA homologue (RCB11-1) and a fragmentary CentA homologue (RCB11-2) inserted into a second *crwydryn* reading frame. The cereba "element" of barley contains two inverted repeats. The left-hand repeat contains the initial region of a long conserved polyprotein reading frame; the right appears to be truncated within the *gag* region (indicated by broken vertical line). (B) Composite cereal *crwydryn* element showing the relative positions of diagnostic PCR fragments and RT and IN regions used in the phylogenetic analysis and other *crwydryn* database entries. Details of these sequences are given in Table 2.

presence of contiguous *gag*-RT sequences in rye. Related *gag* sequences were also identified in the rice centromeric clone RCH3 (Dong *et al.* 1998; Figure 2).

Finally, an essential feature of most conventional LTR retrotransposon elements is a primer binding site (PBS), adjacent to the upstream LTR. Although no PBS has been reported for cereal centromeric fragments, a disrupted methionine tRNA complement is present at the appropriate position in a number of clones (Figure 3). Low-stringency PCR carried out on rye genomic DNA allowed recovery of further LTR junctions, among which were sequences containing an intact methionine tRNA PBS (Figure 3). It is likely that this is the PBS present in ancestral sequences and that insertion of additional bases, presumably as a result of aberrant processing of the adjacent LTR junction, represents a subsequent attenuation or inactivation of the PBS (see discussion).

The *crwydryn* family is conserved in Arabidopsis: Database entries for five BAC clones of *Arabidopsis thaliana* were found to contain full-length or relatively intact Ty3-*gypsy* class retrotransposon elements with reading frame sequences very similar to *crwydryn* (Table 2; Figure 4). T9F8 is the least degenerate of these, on the basis of the lack of reading frame deletions and similarity between its LTRs; no deletions have occurred and divergence is ~2%. Intact elements are predicted to be ~6.5 kb long, with 1-kb LTRs, comparable with the size of cereal elements; 5′ UTRs are, however, shorter (Figure 4).

Phylogenetic analysis with the PHYLIP and GCG packages consistently places cereal centromeric sequences with these elements in a lineage that shows no overlap with other database entries (Figure 5), indicating that the *crwydryn* family represents a distinct group whose origin predates monocot-dicot divergence. Comparison of *crwydryn* members with other Ty3-*gypsy* families indicates that all canonical peptide components are present. No regions of unusual divergence were detected, and the most divergent regions are those that are also most

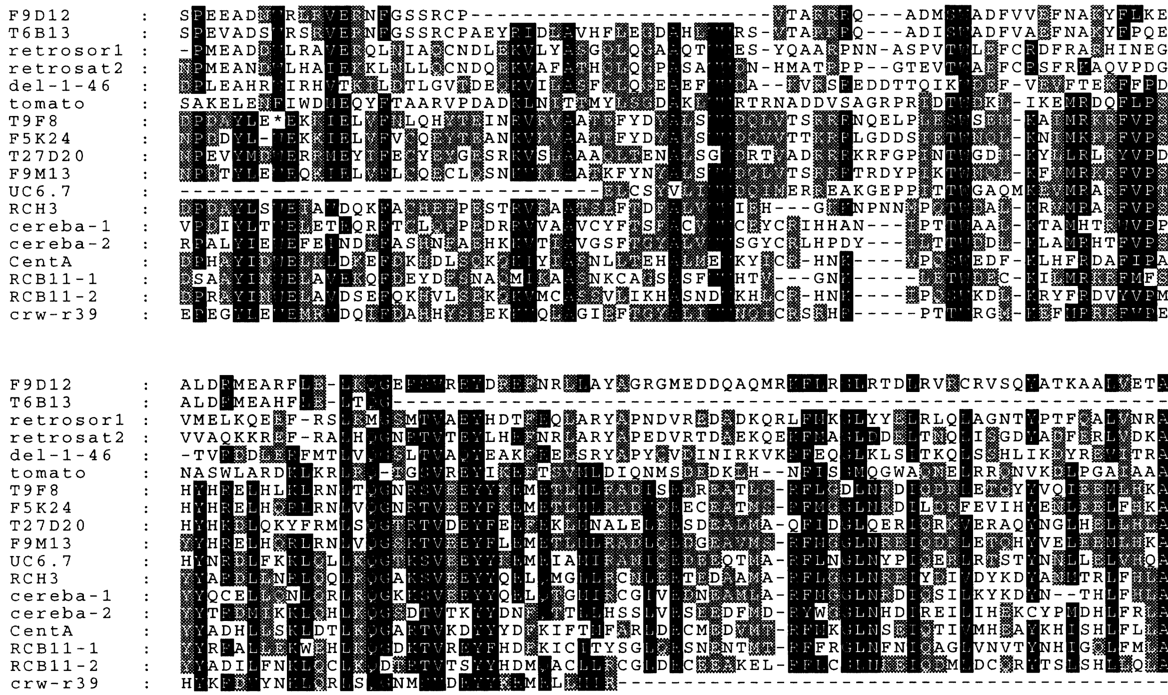

Figure 2.—Reading frames for Ty3-*gypsy* retrotransposon *gag* peptides are present in cereal centromeric clones. The clones from which the *gag* peptide sequences are derived are listed in Table 2. Cereal centromere-specific sequences are represented by clones UC6.7 to RCB11-2, and closely related sequences in Arabidopsis by T9F8 to F9M13. Similar but unrelated plant sequences are represented by clones F9D12 to tomato. T9F8 to cereba-1 are fragments of apparently autonomous elements, and cereba-2 to crw-r39 are fragments of nonautonomous elements. crw-r39 is part of a B-chromosome heterochromatic repeat of rye. Dark shading represents >60% conservation, and light shading >30%.

variable in comparisons between other Ty3 families. Universal characteristics of the family include LTRs defined by terminal TGAT/ATCA inverted repeats, standard tRNAmet$_i$ PBSs, and GGGAG polypurine tracts. The family belongs to branch 1 of Wright and Voytas' classification (Wright and Voytas 1998). A key distinction of this classification is the absence of an *env*-like 3′ reading frame; as the *crwydryn* polyprotein reading frame extends into the 3′ LTR, it is unlikely that such additional genes will be found.

Some Arabidopsis elements display partial or divergent identity with canonical *crwydryn* elements, particularly in LTR regions (Figure 4). For example, T1O16 has acquired LTR sequences that are entirely unrelated to those of T9F8 and also larger (1.2 kb) than normal. The extent of its reading frame divergence indicates that this is a relatively ancient event. A more recent rearrangement has created a chimeric element in T27D20; the 5′ LTR and UTR of a T1O16-like element have been fused to another region of polyprotein and the 3′ LTR of a typical *crwydryn* element (Figure 4). Sequence of unidentified origin lying at the junction of the two includes a second potential PBS. This novel organization may result from a chromosomal rearrangement, but there is also evidence for aberrant processing during retrotransposition, for example, in F5K24, where 40 codons of integrase reading frame are found upstream of a 5′ LTR, and in T1O16 itself, where a PBS is found downstream of the 3′ LTR.

Nonautonomous elements have an ancient origin: In-

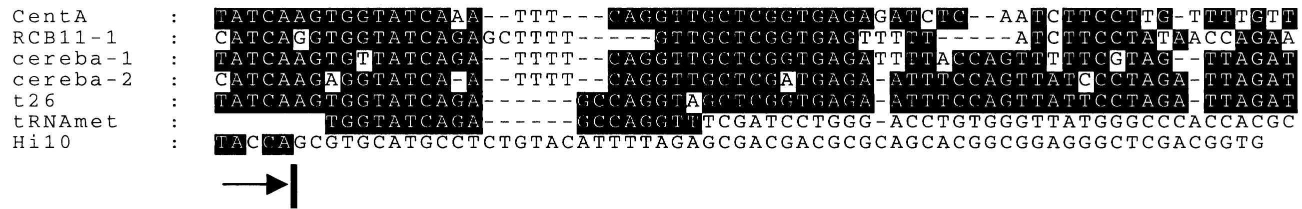

Figure 3.—A rye clone (t26) contains identity with wheat tRNA$_{met}$ at the junction between CCS1 (LTR) and cereba (UTR) homologies. RCB11-1 and the original CCS1 (Hi10) sequences are also shown. Highlighted sequences show significant identities with t26. The limit of conserved CCS1 sequences is indicated by the arrow and vertical bar. The tRNA homologies in CentA, RCB11-1, and the cereba elements are interrupted by additional nucleotides 2 bp downstream of a repeat of the TATCA motif recognized by integrase at the LTR boundary.

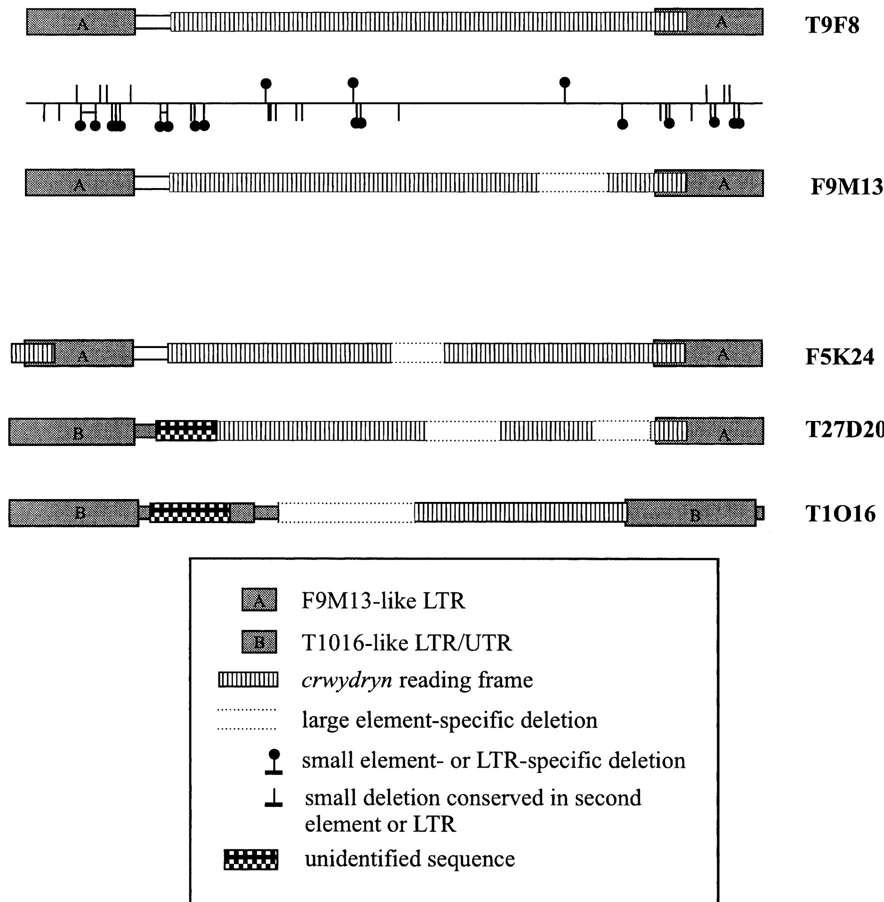

FIGURE 4.—Arabidopsis *crwydryn* elements in BAC database. T9F8 is the least degenerate. F5K24 has an extended 5′ LTR containing integrase sequences. T1016 has a partial but divergent duplication (~70% identical in region marked) that may be part of an independent element inserted in the UTR and a priming site downstream of the conserved 3′ LTR.

ternal deletions of retrotransposon sequences are frequently seen but usually are element specific. The consistency of the CentA/RCB11-2 rearrangement and the coincident truncation of cereba-2, together with striking UTR similarities such as a region of 49-bp identity between cereba-2 and RCB11-2, prompted a closer examination of these variants.

A phylogenetic tree derived from *gag* sequences places all the *crwydryn* elements into the same lineage (Figure 5B). However, elements lacking the *gag* RNA-binding motif (CentA, RCB11-1 and -2, and cereba-2) lie in a separate group to other *crwydryn* members and show greater diversity. Selection for protein function has been maintained, as most pairwise comparisons give K_a:K_s ratios above 2 (*gag* regions diverge faster than other retrotransposon components and so typically show relatively low K_a:K_s ratios). Both the autonomous and nonautonomous groups contain representatives from rice, maize, and barley, indicating that divergence occurred at or before the time of the last common ancestor, at least 60 mya. In addition, neighbor-joining trees for the divergent peptide region place RCB11-2 closer to CentA than to RCB11-1, indicating that the elements' common ancestor arose before the species divergence, some 40 mya. Synonymous substitution rates are high in all interspecific comparisons and for both regions, so that horizontal transfers are unlikely to underlie these relationships. The nonautonomous elements therefore appear to represent divergent members of the same ancient family rather than independent rearrangements.

A further common feature of the nonautonomous elements is a polyadenine stretch in the middle of the 5′ UTR (between 200 and 1000 bp from the LTR). This feature is present in cereb-2 but not cereb-1, although flanking sequences are well conserved between the two; the poly(A) stretch lies within a region upstream of the *gag* gene that is >1 kb and shows >74% identity, while the *gag* region itself shows only 62% identity over <500 bp. This suggests that the stretch may have a functional role that is maintained or regenerated despite pressures of concerted evolution.

The *crwydryn* family maintains a high level of retrotransposition potential in cereals: Most *crwydryn* database entries for cereals are from single genomic clones in which reading frames are degenerate and/or key functional domains are missing. While degenerate retrotransposon elements are frequently seen even for active families (Marillonnet and Wessler 1998), the centromeric domain in which this family is localized may be particularly prone to rearrangement and passive amplification (see below), so that it is not clear to what extent the concentration of *crwydryn* elements represents continuing retrotransposition.

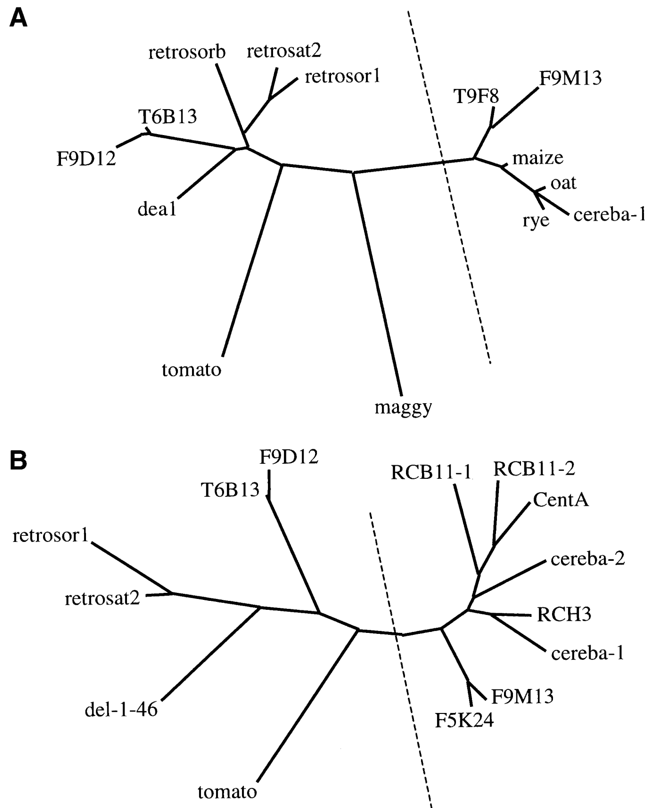

FIGURE 5.—Phylogenetic analyses consistently group together Arabidopsis and cereal *crwydryn* elements. (A) Tree based on one of the most highly conserved retrotransposon genes, reverse transcriptase. *Crwydryn* elements group to the right of the vertical line. The origins of the sequences are described in Table 2. Oat and rye sequences are derived from PCR of reverse transcriptase from genomic DNA and that of maize from the integration target of CentA. (B) Tree based on the most divergent gene, *gag*. *Crwydryn* elements group to the right of the dashed line. Note deep branch lengths between putative autonomous and nonautonomous cereal elements. The identities of the clones are also given in Table 2.

We cloned and sequenced *crwydryn* PCR products from a number of cereal species to characterize their evolution in more detail. In particular, we were interested in the proportion of potentially functional elements and the extent to which individual elements had diverged from the relevant genomic consensus. This information provides a basis for estimating the age of individual insertion events, as nonsynonymous changes are expected to accumulate only in "slave" copies following their generation by retrotransposition from those rare "master" elements in which such changes are absent or nondeleterious. At a minimum, nonsynonymous changes will arise at the same frequency as synonymous changes in host genes; this is expected to be a substantial underestimate as retrotransposition itself is an error-prone process. Thus, the rate of evolution of maize coding genes has been estimated to be $1–4 \times 10^{-9}$/nucleotides/year (PURUGGANAN and WESSLER 1994), while *in vivo* error rates incurred during retroelement replication may be higher than this by at least four orders of magnitude (GABRIEL *et al.* 1996). *Crwydryn* elements that originated in the common ancestor of contemporary species should thus be readily distinguished from recent retrotransposition events.

A total of 45 reverse transcriptase clones were obtained from five species (rye, wheat, oat, sorghum, and rice) and 31 integrase clones were obtained from eight species (rye, wheat, oat, sorghum, and rice, plus maize, millet, and bamboo). The same PCR conditions were used for all species and resulted in a single or predominant product of the expected size in all cases. The only PCR for which no product was obtained was with maize, using reverse transcriptase oligonucleotides. Centromere specificity of clones was confirmed by FISH (see below). It is clear that *crwydryn* sequences have diverged in a species-specific manner, with strong selection for conserved protein function (Figure 6). Most species-specific sequence differences correspond to synonymous or conservative codon replacements and open reading frames were maintained. Surprisingly, there was no evidence for the persistence of ancestral *crwydryn* sequences in these surveys. All clones conformed closely to the relevant species consensus, and total variation was in the range of a few percent (for example, an average of 0.025 nonsynonymous substitutions per nonsynonymous site over a 92-codon region sequenced from 23 rye RT clones). The relative rate of synonymous to nonsynonymous codon changes (K_s:K_a) was very high for interspecific comparisons (6–16:1) and almost equivalent for intraspecific data (*e.g.*, average 1.49:1 for the rye example above), a pattern typical of retrotransposons whose colonization of the genome depends on retrotransposition.

Detailed estimates of the age of individual elements were made by first deriving an approximate mutation rate for the family. We assumed that the close relationship of all the sequences confirms that each domain is derived from a common ancestor in a progenitor of the Poaceae and that the high level of intraspecific similarities indicates that a single consensus may be derived for each species. This is an oversimplification, as functional polymorphisms are apparent in some species. Species consensus sequences were compared to calculate the numbers of changes at synonymous sites; these *K*s values were then combined with published estimates of the date of divergence of various species to provide crude mutation rate estimates. The validity of this approach is supported by the similarity of most estimates, which average 2.08×10^{-8} (reverse transcriptase) or 1.38×10^{-8} (integrase) substitutions per site per year. The dates at which individual elements began to diverge from the species consensus (assumed to be the date of retrotransposition) were then calculated by combining the mutation rate estimate with the number of substitutions calculated to occur at nonsynonymous sites (*K*a) for pairwise comparisons of actual clones *vs.* relevant

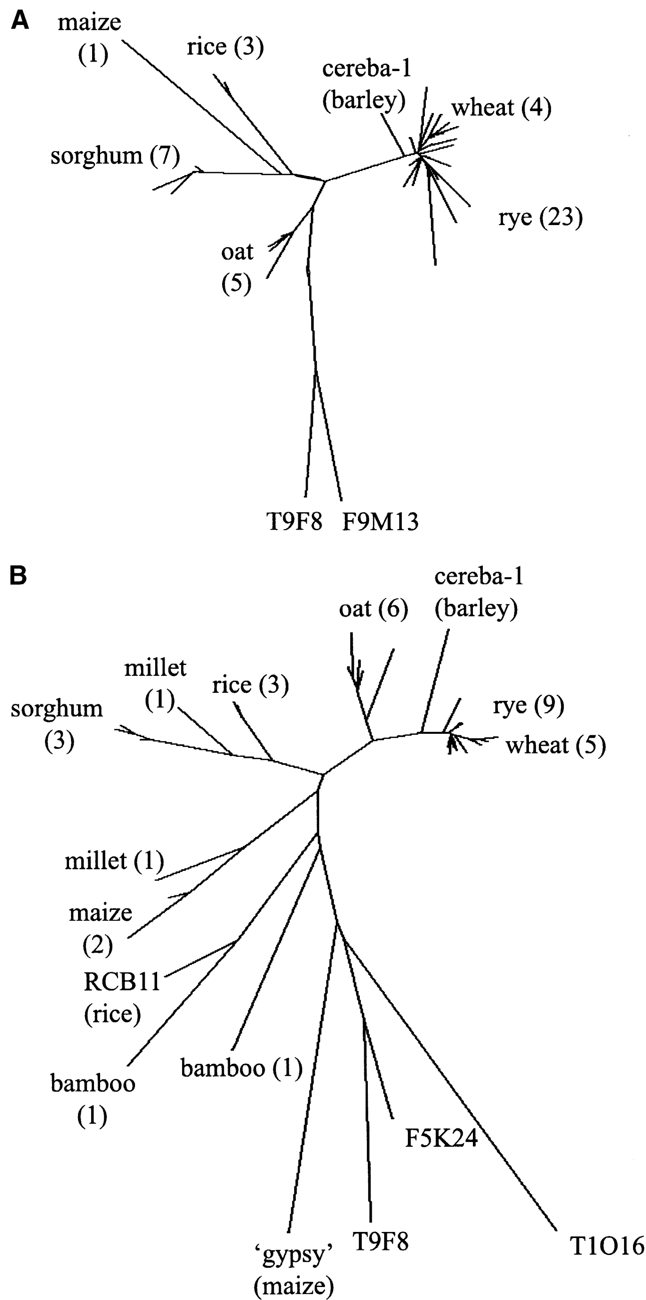

FIGURE 6.—Phylogenetic relationships of cereal *crwydryn* PCR fragments. Relevant regions of other clones are included in the analysis. Figures in parentheses indicate the number of sequences within a lineage. (A) Reverse transcriptase fragments. The single maize sequence is derived from the target of CentA integration (shown in Figure 1). Wheat sequences lie in a single lineage within the rye cluster—one rye sequence is grouped with the wheat lineage. (B) Integrase fragments. The RCB11 sequence indicated is taken from the target of RCB11-2 integration (Figure 1); the three other rice sequences shown include a second rice database entry (RCS1). The maize "gypsy" sequence shown is taken from database entry AF030633.

derived species consensus. In most cases, and for all species, estimates were for dates <1 myr, with 75% of sequences showing <1 myr divergence (see MATERIALS AND METHODS), and the oldest (rye) showing <3.5 myr. These ages are overestimates because nominally "divergent" sites will also include functional subfamilies and PCR errors (note, however, that the nonrandom nature of most intraspecific variation indicates that PCR error rates are relatively low). The recent derivation of most elements from consensus master copies indicates that the family is likely to still be active in most if not all species, while the failure to detect "old" elements implies that either the family is rapidly increasing in abundance at an equivalent rate in each of the divergent species sampled or that ancestral sequences are relatively rapidly removed in their entirety before significant levels of degeneration occur.

***Crwydryn* elements are dispersed throughout rye centromeres:** *Crwydryn*-related sequences have been localized to cereal centromeric regions by FISH to metaphase chromosome preparations (ANANIEV *et al.* 1998; DONG *et al.* 1998; MILLER *et al.* 1998a; PRESTING *et al.* 1998). The resolution of this approach does not allow fine mapping of the elements and leaves open the possibility that *crwydryn* elements are clustered in a few large "islands" or are locally amplified by rearrangements that are frequently found in centromeric regions. Use of fiber-FISH to show short-range (10–20 kb) clustering (DONG *et al.* 1998) does not address these issues. We have used extended rye meiotic prophase chromosome preparations to examine long-range (0.1–1 Mb) organization (Figure 7). Both RT and IN sequences are dispersed throughout centromeric domains, although regional differences in density are apparent in surprisingly regularly sized units. Occasionally we find localization of only one of these sequences in a unit (arrowed in Figure 7). Assuming a chromatin packing ratio of 1400, each unit would be of the order of 100 kb in size, within a domain of at least 4 Mb. KASZAS and BIRCHLER (1998) have demonstrated that the maize B chromosome centromere contains redundant subdomains that are at least 55 kb in size. Each *crwydryn* block may therefore correspond to a functional subdomain of the type identified by KASZAS and BIRCHLER (1998). The occasional concentration of *crwydryn* sequences deficient for key functional components (Figure 7) indicates that the differences in *crwydryn* abundance between blocks may be at least partly a result of passive amplification.

Extinction of *crwydryn* in some Poaceae lineages: PCR screens for the presence of *crwydryn* components were successful in a variety of temperate and tropical species, including Brachyaria and Lolium. The identity of products was confirmed by direct sequencing. There were, however, some species in which no PCR product was obtained for coding regions. In particular, we were unable to amplify reverse transcriptase sequences from a variety of closely related Avena species (designated as potential progenitors of the hexaploid oat C genome, including *Avena eriantha* Cc7056 and *A. ventricosa* Cc7064) or from a number of Festuca species (including

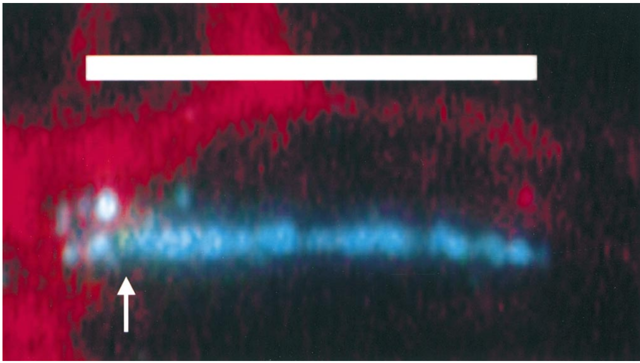

Figure 7.—*Crwydryn* sequences are distributed throughout rye centromeres. Meiotic prophase chromosomes were probed with RT (blue) or IN (green) PCR fragments. A region containing IN but not RT sequences is indicated by the arrow. All other signals represent a combination of both probes. Bar, 10 μm.

Festuca pratensis, F. glaucescens, and *F. mairei*) despite the presence of canonical sequences in more distant relatives (*A. strigosa* Cc7121, *A. canariensis* Cc7041, *A. damascena* Cc7045, *A. prostrata* Cc7060, *A. longiglumis* Cc4851, *A. agadiriana* Cc7433, *A. barbata* Cc4897, *A. macrostachya* Cc7068, and Lolium species). In these lineages, the target sequences appear to be entirely absent rather than variant at priming sites; FISH with appropriate polyploids (*A. sativa* and *F. arundinacea*) demonstrated strong hybridization to the PCR-positive genomes but little or no hybridization to PCR-negative genomes (Figure 8). These data indicate that although the *crwydryn* family colonized the centromeres of a common ancestor before the divergence of the Poaceae and has been maintained intact in most lineages since, there has been occasional recent elimination in some cases. CCS1/LTR sequences from this family may be more universally maintained.

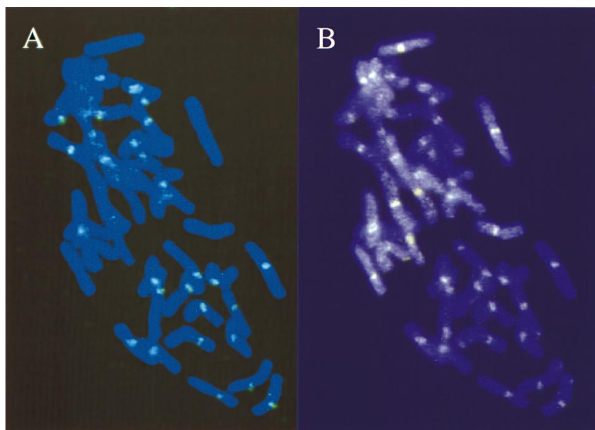

Figure 8.—FISH with *crwydryn* elements. (A) A reverse transcriptase probe derived from *Avena strigosa* (putative A/D genome donor to *A. sativa*) labels the centromeres of the somatic metaphase chromosomes of the A and D but not the C genomes of hexaploid oats (*A. sativa* var. Sun II). (B) An LTR junction probe derived from *A. eriantha* (putative C genome donor to *A. sativa*) labels all the centromeres of the three genomes of *A. sativa* and also contains a fragment of dispersed repeat that specifically "paints" the C genome only.

DISCUSSION

Retrotransposon sequences in Poaceae centromeres: The analyses presented here demonstrate that a single ancient family is the source of all Poaceae centromere-specific retrotransposon sequences reported to date. This family has a conventional organization and its protein components are highly conserved even in Arabidopsis homologues. Within the Poaceae, the family has evolved in a species-specific manner, with selection clearly acting at the protein level to maintain retrotransposition potential. Any explanation for its unusual distribution must therefore include a role for element mobility. Other unusual features are a relatively low mutation rate, a lack of ancient or degenerate progenitors, and a high degree of conservation of LTR sequences. We believe that the simplest explanation for these observations is also the one for which there are the most precedents, namely, that the family has developed an efficient mechanism for targeting retrotransposition to a specialized region of the host genome, in this case the centromere, which is intrinsically unstable. Turnover of sequences within the host region would then rapidly eliminate degenerating sequences, so that only functional elements, which can maintain a high recruitment rate, accumulate.

There are numerous examples of retroelements developing targeting mechanisms, including Ty1, Ty3, and Ty5 retrotransposons of yeast, various subtelomeric or telomere-specific LINEs in insects and algae, and mammalian retroviruses, each of which has different consequences for the elements' distribution in the host genome. The closest parallel for the cereal *crwydryn* distribution appears to be shown by the R elements of arthropods, families of LINEs that use a sequence-specific endonuclease as a mechanism to target integration to rDNA arrays. Although this mechanism is not used by LTR retrotransposons and both retroelement and host domain organization are very different in the two cases, R elements resemble the cereal *crwydryn* in maintaining a high degree of retrotransposition potential and a low mutation rate comparable to that of the host (Eickbush and Eickbush 1995; Eickbush *et al.* 1995, 1997). The major influence is rapid elimination of old elements in their entirety, presumably by random array contraction (Jakubczak *et al.* 1992), but some degree of concerted evolution is also likely to apply, leading to relatively high species homogeneity (Jakubczak *et al.* 1992).

Centromeric domains are typically composed of repetitive sequence arrays that, like rDNA arrays, may be expected to be unstable and unfavorable sites for mobile element targeting. This is demonstrated, for example,

by the Drosophila Dp1187 minichromosome where few insertions within centromeric satellite arrays appear to survive long enough for significant sequence degeneration to occur (Sun *et al.* 1997). The only potential decayed or ancient element in the Dp1187 centromeric region lies at the junction of two different satellite family arrays, *i.e.*, at the point at which ectopic recombination is unlikely to be able to lead to clean excision. However, increasing either the complexity of the centromeric sequences or the rate at which new mobile elements are recruited will reduce the effectiveness of this elimination. We discuss below factors that may have allowed *crwydryn* colonization of cereal centromeres and other features that support this model.

Insertion site choice by the *crwydryn* family: The highly specific distribution of the *crwydryn* family in the Poaceae is reflected in Arabidopsis. Seven BACs of Arabidopsis containing *crwydryn* elements map to within <2 cM of three of the five centromeric map sites, and two additional *crwydryn* elements are found near clusters of centromeric repeats distant from the mapped centromere positions on chromosomes 2 and 4. There is also a frequent association with the Ty3-*gypsy* retrotransposon family Athila, which is particularly abundant in pericentromeric regions and has been found inserted into Arabidopsis centromeric satellite arrays (Pelissier *et al.* 1996). This distribution suggests that centromeric targeting by cereal *crwydryn* may represent a relatively simple refinement of a previously existing regional preference for "silent" or heterochromatin-like domains. Most cereal *crwydryn* insertions are into other *crwydryn* elements, although not at specific sites, raising the possibility that *crwydryn* elements themselves create the conditions, such as chromatin conformation, that direct targeting. It is interesting that the database entries indicate a preferred polarity of insertion (Figure 1). Among the additional LTR junctions that we have recovered, three are within the rye centromere-specific repeat AWRC1, which is itself a degenerate *crwydryn* element (not shown). All three are in reading frames and again in the same orientation as the target.

Evolution of the *crwydryn* family in Poaceae and Arabidopsis genomes: It is striking that the *crwydryn* polyprotein in the two divergent lineages appears to be highly conserved and that an "ancestral" element with a full-length reading frame (T9F8) shows the most recent signs of retrotransposition in Arabidopsis despite the presence in the genome of subfamilies having small common deletions. This suggests that such derivatives have only a short evolutionary life in any genome. There are, however, two differences between *crwydryn* of cereals and Arabidopsis that appear to be significant.

First, the high degree of LTR conservation in cereal *crwydryn* contrasts sharply with the variability in Arabidopsis, where rapid change may occur both by LTR "swaps" as described above and by small incremental changes (F5K24, for example, shares only 54% identity with T9F8 over a 900-bp region of the LTR that includes 16 gaps, but has 76% identity over an ungapped 2-kb region of the reading frame). In cereals, however, the LTR U5 region is sufficiently well conserved to act as a universal centromeric probe (Aragon-Alcaide *et al.* 1996). Given the unusual requirements expected for expression from a centromeric domain, conservation of this region is likely to be at least partly due to functional constraint maintaining promoter elements contained within it. In addition, however, the combination of preferred orientation and frequent interspersion seen in cereals, together with the promiscuous use of neighboring LTRs seen in Arabidopsis, could lead to frequent exchange of LTRs and homogenization of sequences. Gene conversion may also contribute to concerted evolution; the unexpectedly high similarity of cereba-1 and cereba-2 5′ UTRs has already been described.

Second, the nonautonomous but conserved cereal subfamily does not appear to have a direct equivalent in Arabidopsis, where large deletion derivatives all appear to occur at single genomic sites, *i.e.*, to have arisen following integration. Nonautonomous elements must depend on components provided *in trans*, which is known to occur (Avedisov *et al.* 1998) and which may be particularly common in some families (Wright and Voytas 1998), but it appears to be rare for conserved nonautonomous families to emerge. *A priori*, such families can be expected to arise when autonomous elements are highly abundant, both because *trans* components are more readily available and because a greater range of variant elements are produced. The population of *crwydryn* elements in the Arabidopsis genome may be too small to support nonautonomous variants (although the subfamilies with small common deletions may represent attenuated variants that are supported by persistent full-length elements), but it is also surprising that a single nonautonomous family is seen in cereals. We suggest that this reflects an ancient explosion in cereal *crwydryn* activity, which has subsequently been moderated at least in part by the presence of the nonautonomous element itself (see below).

The rise and fall of the *crwydryn* centromeric colonies: We have shown that cereal *crwydryn* elements, despite their presence in the centromeres of many species, behave as conventional retrotransposons and are closely related to elements in Arabidopsis that are not centromere specific. In addition, conserved *crwydryn* components may be absent in the centromeres of some cereal species without apparent effect. Our conclusion is that conservation of *crwydryn* reading frames reflects an ancient optimization, predating the origin of grasses, for function based on retrotransposition rather than centromeric activity. Conservation of noncoding sequences may reflect both a novel functional selection (for centromeric promoter activity) and the consequences of a novel genomic distribution (with homogenization driven by high concentrations of similarly oriented ele-

ments within the centromere). However, the most distinctive and surprising feature of the cereal *crwydryn* distribution is that host centromeres are unusually complex. Moore *et al.* (1997) have recently proposed a scenario that seems to provide an explanation for this, which is that the closest relatives of the Poaceae have holocentric rather than regional centromeres and that emergence of the Poaceae progenitor may have coincided with a large-scale reorganization of these domains. An attractive model is that this reorganization provided a new niche for the *crwydryn* family and that some feature of the new domains corresponded to the integration preference of the family, whether preexisting or variant. Colonization by large numbers of *crwydryn* elements may have reinforced the distinctiveness of particular regions, both for the establishment of centromeric structures and for further retrotransposon "homing," and may indeed have driven the initial reorganization. A precedent for an extremely rapid process has been described for a recently reorganized marsupial centromere (O'Neill *et al.* 1998), while the characterization of various neocentromeres has demonstrated that their establishment is to a large extent dependent on epigenetic effects rather than on specific properties of the underlying DNA (Karpen and Allshire 1997).

The evolution of the nonautonomous family, attenuation of full-length elements by PBS insertions, and extinction of centromeric *crwydryn* in some lineages are all consistent with stabilization of the genomes following this initial explosion. Specifically, the nonautonomous elements, which date from the time of initial amplification, are likely to curb further expansion by the *crwydryn* family as a whole, both because of titration of retrotransposition components and because of specific insertional mutagenesis of autonomous elements. Following this initial restraint, more gradual changes may be expected to make the centromeres more resistant to colonization. In particular, simple repetitive DNA families are expected to evolve, leading to the accumulation of the centromeric satellite arrays seen in most other organisms. Such species-specific arrays have been reported, for example, in maize and sorghum (Ananiev *et al.* 1998; Miller *et al.* 1998b) and will increase the rate at which new insertions will be removed by ectopic recombination. The cumulative effect of these processes will be to drive extinction of active *crwydryn* elements in some lineages, followed by elimination or fragmentary incorporation into tandem arrays, as appears to have occurred relatively recently in the oat C genome progenitors and some Fescue species (Figure 8). Alternatively, active members of the family may acquire new specificities, as may have occurred in maize where noncentromeric *crwydryn* LTRs have been found (Aragon-Alcaide *et al.* 1996). In either case, satellite arrays may eventually oust *crwydryn* elements from all Poaceae centromeres. Under such circumstances the more modest copy numbers and less-specialized targets of the Arabidopsis elements may provide a more secure long-term evolutionary future.

We thank Dr. Steve Reader and Dr. Terry Miller for kindly supplying Aegilops species and wheat/rye addition lines. We also thank Dr. Graham Moore and members of his laboratory for helpful discussions and John Harper for assistance with FISH. This work was funded by Biotechnology and Biological Sciences Research Council grant PO1643 to G.J., J.W.F., and R.N.J., a University of Wales, Aberystwyth, Institute of Biological Sciences studentship to C.S.

LITERATURE CITED

Abbo, S., R. P. Dunford, T. N. Foote, S. M. Reader, R. B. Flavell *et al.*, 1995 Organization of retro-element and stem-loop repeat families in the genomes and nuclei of cereals. Chrom. Res. **3:** 5–15.

Ananiev, E. V., R. L. Phillips and H. W. Rines, 1998 Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc. Natl. Acad. Sci. USA **95:** 13073–13078.

Aragon-Alcaide, L., T. Miller, T. Schwarzacher, S. Reader and G. Moore, 1996 A cereal centromeric sequence. Chromosoma **105:** 261–268.

Avedisov, S. N., E. S. Zelentsova and Y. V. Ilyin, 1998 Transmobilization of deletion copies of the mdg3 retrotransposon in cultured cells of *Drosophila*. Genetika **34:** 335–342.

Bennetzen, J. L., and M. Freeling, 1993 Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet. **9:** 259–261.

Bennetzen, J. L., and M. Freeling, 1997 The unified grass genome: synergy in synteny. Genome Res. **7:** 301–306.

Bennetzen, J. L., and E. A. Kellogg, 1997 Do plants have a one-way ticket to genomic obesity? Plant Cell **9:** 1509–1514.

Bennetzen, J. L., P. San Miguel, M. S. Chen, A. Tikhonov, M. Francki *et al.*, 1998 Grass genomes. Proc. Natl. Acad. Sci. USA **95:** 1975–1978.

Boeke, J. D., and S. E. Devine, 1998 Yeast retrotransposons: finding a nice quiet neighborhood. Cell **93:** 1087–1089.

Dong, F. G., J. T. Miller, S. A. Jackson, G. L. Wang, P. C. Ronald *et al.*, 1998 Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc. Natl. Acad. Sci. USA **95:** 8135–8140.

Eickbush, D. G., and T. H. Eickbush, 1995 Vertical transmission of the retrotransposable elements R1 and R2 during the evolution of the *Drosophila melanogaster* species subgroup. Genetics **139:** 671–684.

Eickbush, D. G., W. C. R. Lathe, M. P. Francino and T. H. Eickbush, 1995 R1 and R2 retrotransposable elements of *Drosophila* evolve at rates similar to those of nuclear genes. Genetics **139:** 685–695.

Eickbush, T. H., W. D. Burke, D. G. Eickbush and W. C. R. Lathe, 1997 Evolution of R1 and R2 in the rDNA units of the genus *Drosophila*. Genetica **100:** 49–61.

Gabriel, A., M. Willems, E. H. Mules and J. D. Boeke, 1996 Replication infidelity during a single cycle of Ty1 retrotransposition. Proc. Natl. Acad. Sci. USA **93:** 7767–7771.

Grandbastien, M.-A., 1992 Retroelements in higher plants. Trends Genet. **8:** 103–108.

Heslop-Harrison, J. S., T. Schwarzacher, K. Anamthawat-Jonsson, A. R. Leitch and I. J. Leitch, 1991 *In situ* hybridisation with automated chromosome denaturation. Technique **3:** 109–116.

Higashiyama, T., Y. Noutoshi, M. Fujie and T. Yamada, 1997 Zepp, a LINE-like retrotransposon accumulated in the *Chlorella* telomeric region. EMBO J. **16:** 3715–3723.

Jakubczak, J. L., M. K. Zenni, R. C. Woodruff and T. H. Eickbush, 1992 Turnover of R1 (type I) and R2 (type II) retrotransposable elements in the ribosomal DNA of *Drosophila melanogaster*. Genetics **131:** 129–142.

Jiang, J. M., S. Nasuda, F. G. Dong, C. W. Scherrer, S. S. Woo *et al.*, 1996 A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc. Natl. Acad. Sci. USA **93:** 14210–14213.

Karpen, G. H., and R. C. Allshire, 1997 The case for epigenetic

effects on centromere identity and function. Trends Genet. **13:** 489–496.

Kaszas, E., and J. A. Birchler, 1998 Meiotic transmission rates correlate with physical features of rearranged centromeres in maize. Genetics **150:** 1683–1692.

Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Prudhomme *et al.*, 1994 Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **91:** 1285–1289.

Langdon, T., C. Seago, R. N. Jones, H. Thomas, H. Ougham *et al.*, 2000 *De novo* evolution of satellite DNA on the rye B chromosome. Genetics **154:** 869–884.

Lee, C., R. Wevrick, R. B. Fisher, M. A. Fergusonsmith and C. C. Lin, 1997 Human centromeric DNAs. Hum. Genet. **100:** 291–304.

Leitch, I. J., and J. S. Heslop-Harrison, 1992 Detection of digoxygenin labelled DNA probes hybridized to plant chromosomes *in situ*, pp. 18–25 in *Methods in Molecular Biology: Protocols for Nucleic Acid Analysis by Non-radioactive Techniques*, edited by P. G. Isaac. Humana Press, Clifton, NJ.

Marillonnet, S., and S. R. Wessler, 1998 Extreme structural heterogeneity among the members of a maize retrotransposon family. Genetics **150:** 1245–1256.

Miller, J. T., F. G. Dong, S. A. Jackson, J. Song and J. M. Jiang, 1998a Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics **150:** 1615–1623.

Miller, J. T., S. A. Jackson, S. Nasuda, B. S. Gill, R. A. Wing *et al.*, 1998b Cloning and characterization of a centromere-specific repetitive DNA element from *Sorghum bicolor.* Theor. Appl. Genet. **96:** 832–839.

Moore, G., L. Aragon-Alcaide, M. Roberts, S. Reader, T. Miller *et al.*, 1997 Are rice chromosomes components of a holocentric chromosome ancestor? Plant Mol. Biol. **35:** 17–23.

Nonomura, K.-I., and N. Kurata, 1999 Organization of the 1.9-kb repeat unit RCE1 in the centromeric region of rice chromosomes. Mol. Gen. Genet. **261:** 1–10.

Noutoshi, Y., Y. Ito, S. Kanetani, M. Fujie, S. Usami *et al.*, 1998 Molecular anatomy of a small chromosome in the green alga *Chlorella vulgaris.* Nucleic Acids Res. **26:** 3900–3907.

O'Neill, R. J. W., M. J. O'Neill and J. A. M. Graves, 1998 Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature **393:** 68–72.

Pan, W. H., A. Houben and R. Schlegel, 1992 Highly effective cell synchronization in plant roots by hydroxyurea and amiprophosmethyl or colchicine. Genome **36:** 387–390.

Pardue, M. L., O. N. Danilevskaya, K. Lowenhaupt, F. Slot and K. L. Traverse, 1996 *Drosophila* telomeres: new views on chromosome evolution. Trends Genet. **12:** 48–52.

Pearce, S. R., U. Pich, G. Harrison, A. J. Flavell, J. S. Heslop-Harrison *et al.*, 1996 The Ty1-copia group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. Chrom. Res. **4:** 357–364.

Pelissier, T., S. Tutois, S. Tourmente, J. M. Deragon and G. Picard, 1996 DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in *Athila* retroelement sequences. Genetica **97:** 141–151.

Petrov, D. A., 1997 Slow but steady reduction of genome size through biased mutation. Plant Cell **9:** 1900–1901.

Presting, G. G., L. Malysheva, J. Fuchs and I. Z. Schubert, 1998 A TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. **16:** 721–728.

Purugganan, M. D., and S. R. Wessler, 1994 Molecular evolution of the plant R-regulatory gene family. Genetics **138:** 849–854.

San Miguel, P., B. S. Gaut, A. Tikhonov, Y. Nakajima and J. L. Bennetzen, 1998 The paleontology of intergene retrotransposons of maize. Nat. Genet. **20:** 43–45.

Song, S. U., T. Gerasimova, M. Kurkulos, J. D. Boeke and V. G. Corces, 1994 An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. Genes Dev. **8:** 2046–2057.

Sun, X. P., J. Wahlstrom and G. Karpen, 1997 Molecular structure of a functional *Drosophila* centromere. Cell **91:** 1007–1019.

Suoniemi, A., K. Anamthawat-Jonsson, T. Arna and A. H. Schulman, 1996a Retrotransposon BARE-1 is a major, dispersed component of the barley (*Hordeum vulgare* L) genome. Plant Mol. Biol. **30:** 1321–1329.

Suoniemi, A., A. Narvanto and A. H. Schulman, 1996b The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol. Biol. **31:** 295–306.

Suoniemi, A., D. Schmidt and A. H. Schulman, 1997 BARE-1 insertion site preferences and evolutionary conservation of RNA and cDNA processing sites. Genetica **100:** 219–230.

Waugh, R., K. McLean, A. J. Flavell, S. R. Pearce, A. Kumar *et al.*, 1997 Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol. Gen. Genet. **253:** 687–694.

Wright, D. A., and D. F. Voytas, 1998 Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. Genetics **149:** 703–715.

Zhong, X-B., P. F. Fransz, J. Wennekes Van Eden, P. Zabel, A. van Kammen *et al.*, 1996 High resolution mapping on pachytene chromosomes and extended DNA fibres by fluorescence *in situ* hybridisation. Plant Mol. Biol. Rep. **14:** 232–242.

Communicating editor: C. S. Gasser