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 (barley retrotransposon) 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...
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 (pSaun3A9; Jang 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 Tyl-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 at R.N.J. at UWA; Chinese Spring (CS; wheat); SunII (oat); maize, sorghum, grass, and Avena species from germ plasm maintained at IGCR; 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 (326 clone) was carried out with rye genomic DNA and oligonucleotide Hi10R (CGRTYGCTAGGCCTGA); 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 (GAGATCMWGCGTCARATWCAGAAATNCT) and BOTYR (GGCATGCAAGCCACTGCT); 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 (TRGCGAGAAGGGCGATG GAG) and INTR2 (TTTGTCCCATGTYTGNNRGT), conditions as for RT. The insert in clone UC6.7 was derived from PCR of rye DNA amplified with AW37 (TATGTKTRATHT 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 127 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 (CCTCAGTCMGA TGGMCARACNGA) and UNIHI (AGGKGC CCGATCCTTTGGY) (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 digoxigenin-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 antidigoxigenin 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 with 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 usingPILEUP and adjusted by eye. Sequence alignments were displayed using GeneDoc (K.B. Nicholas and H.B. Nicholas, distributed by the authors). K/K ratios were calculated using the GCG program Diverge. Phylogenetic analyses were carried out using the PHYLIIP 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 cry8/dyn elements in Arabidopsis bacterial artificial chromosomes (BACs) T9F8, F5K24, F9M13, T1A11, 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 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/ubiquitous 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>
<thead>
<tr>
<th>TABLE 1</th>
<th>Mutation rates of RT and IN in the Poaceae using species consensus or least divergent pairs of sequences (in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT (in millions of years)</td>
</tr>
<tr>
<td>Wheat/sorghum</td>
<td>1.48 × 10^{-3}</td>
</tr>
<tr>
<td>Wheat/rice</td>
<td>1.44 × 10^{-3}</td>
</tr>
<tr>
<td>Wheat/oat</td>
<td>1.41 × 10^{-3}</td>
</tr>
<tr>
<td>Oat/sorghum</td>
<td>1.40 × 10^{-3}</td>
</tr>
<tr>
<td>Oat/rice</td>
<td>1.39 × 10^{-3}</td>
</tr>
<tr>
<td>Oat/millet</td>
<td>1.38 × 10^{-3}</td>
</tr>
<tr>
<td>Sorghum/maize</td>
<td>1.37 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/sorghum</td>
<td>1.35 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/maize</td>
<td>1.34 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/oat</td>
<td>1.33 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/millet</td>
<td>1.32 × 10^{-3}</td>
</tr>
<tr>
<td>Millet/bamboo</td>
<td>1.31 × 10^{-3}</td>
</tr>
<tr>
<td>Rice/sorghum</td>
<td>1.30 × 10^{-3}</td>
</tr>
<tr>
<td>Rice/maize</td>
<td>1.29 × 10^{-3}</td>
</tr>
<tr>
<td>Rice/millet</td>
<td>1.28 × 10^{-3}</td>
</tr>
<tr>
<td>Millet/maize</td>
<td>1.27 × 10^{-3}</td>
</tr>
<tr>
<td>Millet/bamboo</td>
<td>1.26 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/millet</td>
<td>1.25 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/maize</td>
<td>1.24 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/maize</td>
<td>1.23 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/oat</td>
<td>1.22 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/millet</td>
<td>1.21 × 10^{-3}</td>
</tr>
<tr>
<td>Millet/bamboo</td>
<td>1.20 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/millet</td>
<td>1.19 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/maize</td>
<td>1.18 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/maize</td>
<td>1.17 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/oat</td>
<td>1.16 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/millet</td>
<td>1.15 × 10^{-3}</td>
</tr>
<tr>
<td>Millet/bamboo</td>
<td>1.14 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/millet</td>
<td>1.13 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/maize</td>
<td>1.12 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/maize</td>
<td>1.11 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/oat</td>
<td>1.10 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/millet</td>
<td>1.09 × 10^{-3}</td>
</tr>
<tr>
<td>Millet/bamboo</td>
<td>1.08 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/millet</td>
<td>1.07 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/maize</td>
<td>1.06 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/maize</td>
<td>1.05 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/oat</td>
<td>1.04 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/millet</td>
<td>1.03 × 10^{-3}</td>
</tr>
<tr>
<td>Millet/bamboo</td>
<td>1.02 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/millet</td>
<td>1.01 × 10^{-3}</td>
</tr>
<tr>
<td>Bamboo/maize</td>
<td>1.00 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/maize</td>
<td>0.99 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/oat</td>
<td>0.98 × 10^{-3}</td>
</tr>
<tr>
<td>Maize/millet</td>
<td>0.97 × 10^{-3}</td>
</tr>
<tr>
<td>Millet/bamboo</td>
<td>0.96 × 10^{-3}</td>
</tr>
</tbody>
</table>
| Bambo...
TABLE 2

Details of the clones used in this analysis

<table>
<thead>
<tr>
<th>Clone</th>
<th>Species</th>
<th>PCR primers</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN</td>
<td>(Integrase) SORGF1 + INR2</td>
<td>This work</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>(Reverse transcriptase) S14F + BOTYR</td>
<td>This work</td>
<td></td>
</tr>
<tr>
<td>UC8.5</td>
<td>Aegilops squarrosa</td>
<td>REPQS + UNIHI</td>
<td>This work</td>
</tr>
<tr>
<td>dea1</td>
<td>A. comosus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F9D12</td>
<td>Arabidopsis thaliana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T6B13</td>
<td>A. thaliana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T9F8</td>
<td>A. thaliana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F5K24</td>
<td>A. thaliana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T27D20</td>
<td>A. thaliana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F9M13</td>
<td>A. thaliana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1016</td>
<td>A. thaliana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T17A11</td>
<td>A. thaliana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCS1</td>
<td>Brachypodium sylvaticum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hi10</td>
<td>B. sylvaticum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cereba</td>
<td>Hordeum vulgare</td>
<td></td>
<td></td>
</tr>
<tr>
<td>del-1-46</td>
<td>Lilium henryi</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tomato</td>
<td>Lycopersicon esculentum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>maggy</td>
<td>Magnaportha grisea</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCB11</td>
<td>Oryza sativa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCS1</td>
<td>O. sativa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCH3</td>
<td>O. sativa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>retrosat</td>
<td>O. sativa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t26</td>
<td>Secale cereale</td>
<td>Hi10R</td>
<td>This work</td>
</tr>
<tr>
<td>UC6.7</td>
<td>S. cereale</td>
<td>AW37 + BOTYR</td>
<td>This work</td>
</tr>
<tr>
<td>crw-r39</td>
<td>S. cereale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sau3A9</td>
<td>Sorghum bicolor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>retrosor1</td>
<td>S. bicolor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>retrosorb</td>
<td>S. bicolor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CentA</td>
<td>Zea mays</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UC7.12</td>
<td>Z. mays</td>
<td>REPQS + UNIHI</td>
<td>This work</td>
</tr>
<tr>
<td>gypsy</td>
<td>Z. mays</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The clones have been listed alphabetically by species.

Expected 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 cereal gypsy 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 cereal gypsy ancestor.

Additional clones were isolated to confirm that the ancestral organization of the cereal gypsy 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 homologs 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
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
variable in comparisons between other Ty3 families. Universal characteristics of the family include LTRs defined by terminal TGAT/ATCA inverted repeats, standard tRNAMet, PBs, and GGAG 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 extended 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 extended 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 extended 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 extended 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 extended 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 extended 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 extended 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 extended into the 3' LTR, it is unlikely that such additional genes will be found.
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.
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 non-synonymous site over a 92-codon region sequenced 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 × 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.

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.
Retrotransposon Evolution in Plants

These ages are overestimates because nominally "diver-
gent" 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 im-
plies that either the family is rapidly increasing in abun-
dance at an equivalent rate in each of the divergent
species sampled or that ancestral sequences are rela-
tively rapidly removed in their entirety before significant
levels of degeneration occur.

**Crwydryn elements are dispersed throughout rye cent-
tromeres:** Crwydryn-related sequences have been local-
ized 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 "is-
lands" 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) organi-
zation (Figure 7). Both RT and IN sequences are dis-
persed throughout centromeric domains, although regi-
Onal 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 cen-
tromere contains redundant subdomains that are at
least 55 kb in size. Each crwydryn block may therefore
correspond to a functional subdomain of the type identi-
fi ed 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 prod-
ucts was confirmed by direct sequencing. There were,
however, some species in which no PCR product was
obtained for coding regions. In particular, we were un-
able 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 6](image-url)

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 frag-
ments. 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 (RC31). The maize
“gypsy” sequence shown is taken from database entry
AF036033.
occasional recent elimination in some cases. CCS1/LTR sequences from this family may be more universally maintained.

**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 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 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 degenera-
tion to occur (Sun et al. 1997). The only potential de-
cayed 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 se-
quences or the rate at which new mobile elements are
recruited will reduce the effectiveness of this elimina-
tion. We discuss below factors that may have allowed
craydryn colonization of cereal centromeres and other
features that support this model.

Insertion site choice by the craydryn family: The
highly specific distribution of the craydryn family in the
Poaceae is reflected in Arabidopsis. Seven BACs of Ara-
bidopsis containing craydryn elements map to within ~2
cM of three of the five centromeric map sites, and two
additional craydryn elements are found near clusters of
centromeric repeats distant from the mapped centro-
mere positions on chromosomes 2 and 4. There is also a
frequent association with the Ty3-gypsy retrotransposon
family Athila, which is particularly abundant in pericen-
tromeric regions and has been found inserted into Ara-
bidopsis centromeric satellite arrays (Pelissier et al.
1996). This distribution suggests that centromeric tar-
geting by cereal craydryn may represent a relatively sim-
ple refinement of a previously existing regional prefer-
ence for “silent” or heterochromatin-like domains. Most
cereal craydryn insertions are into other craydryn ele-
ments, although not at specific sites, raising the possibil-
ity that craydryn elements themselves create the condi-
tions, such as chromatin conformation, that direct
targeting. It is interesting that the database entries indi-
cate 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 craydryn element
(not shown). All three are in reading frames and again
in the same orientation as the target.

Evolution of the craydryn family in Poaceae and Arabi-
dopsis genomes: It is striking that the craydryn polypro-
tein 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 craydryn of cere-
als and Arabidopsis that appear to be significant.

First, the high degree of LTR conservation in cereal
craydryn contrasts sharply with the variability in Arabi-
dopsis, 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 pre-
ferred orientation and frequent interspersion seen in
cereals, together with the promiscuous use of neigh-
boring LTRs seen in Arabidopsis, could lead to frequent
exchange of LTRs and homogenization of sequences.
Gene conversion may also contribute to concerted evol-
uation; 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 ap-
ppear 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 (Avvedisov 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 fami-
lies 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
craydryn elements in the Arabidopsis genome may be
too small to support nonautonomous variants (although
the subfamilies with small common deletions may repre-
sent 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
craydryn activity, which has subsequently been moder-
atated at least in part by the presence of the nonautono-
ous element itself (see below).

The rise and fall of the craydryn centromeric colonies:
We have shown that cereal craydryn elements, despite
their presence in the centromeres of many species, be-
have as conventional retrotransposons and are closely
related to elements in Arabidopsis that are not centro-
mere specific. In addition, conserved craydryn compo-
ments may be absent in the centromeres of some cereal
species without apparent effect. Our conclusion is that
conservation of craydryn reading frames reflects an an-
cient optimization, predating the origin of grasses, for
function based on retrotransposition rather than cen-
tromeric activity. Conservation of noncoding sequences
may reflect both a novel functional selection (for centro-
mere 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 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 curtail 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 outst *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**


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


Communicating editor: C. S. Gasser