The Plant Cell, Vol. 6, 1177–1186, August 1994 © 1994 American Society of Plant Physiologists

Integration and Nonrandom Mutation of a Plasma Membrane Proton ATPase Gene Fragment within the *Bs1* Retroelement of Maize

Young-Kwan Jin and Jeffrey L. Bennetzen[1]

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907-1392

Retrotransposons are a class of mobile DNA sequences in eukaryotes that transpose through a reverse-transcribed RNA intermediate. Retrotransposons containing long terminal repeats have many of the attributes of retroviruses in animals but have not been previously observed to acquire a portion of a cellular gene as RNA tumor viruses do with oncogenes. We have found homology to plasma membrane proton ATPase genes within the *Bs1* retrotransposon of maize, and this homology led us to clone the maize plasma membrane proton ATPase gene, which we have named *Mha1*. The sequence of *Mha1* confirmed that 654 bp of this ATPase gene are present in *Bs1*; this segment represents the last amino acid of exon 4, all of exons 5 to 9, and part of exon 10. All introns have been removed from this acquired DNA, whereas 81 single base pair substitutions and a deletion of 183 bp in *Bs1* differentiate these contiguous segments. The secondary mutations led to fewer changes in the derived *Bs1* protein sequence than predicted for neutral events, suggesting that the acquired *Mha1* DNA performs a selected function within *Bs1*. These data indicate that a retrotransposon can incorporate and transmit a portion of a standard nuclear gene transcript within its genetic material. Alternatively, these results suggest that *Bs1* may represent a defective version of a plant retrovirus.

INTRODUCTION

In eukaryotes, several classes of mobile elements amplify their copy number via reverse transcription of a transcribed RNA intermediate. Among these retroelements are the retrotransposons and retroviruses, both of which are distinctive for the long terminal repeats (LTRs) found in direct orientation in their integrated forms. The retroviruses, unlike the retrotransposons, encode proteins that allow their assembly into infectious virions that are released from the host cell. Although it seems reasonable that retroviruses have evolved from retrotransposons (Temin, 1980), the converse is also possible (Finnegan, 1983). Some animal viruses have been observed to acquire parts of normal cellular genes. Commonly, these genes serve an oncogenic function that conditions host cells for high levels of metabolism and mitotic activity that are favorable to retrovirus production (Varmus, 1984). Like hepadnaviruses in animal systems (Summers and Mason, 1982), double-stranded DNA caulimoviruses in plants are known to replicate by reverse transcription without integrating into the host chromosome (Pfeiffer and Hohn, 1983). However, retroviruses have not been observed in plants, and no retrotransposon has yet been observed to acquire a portion of a normal cellular gene.

Several retrotransposons have been observed in plants (reviewed by Grandbastien, 1992), and the first of these was the *Bs1* element of maize (Johns et al., 1985). *Bs1* was identified as an inactivating insertion within exon 9 of the maize alcohol dehydrogenase 1 (*Adh1*) gene. Relative to other classes of mobile DNA in maize, *Bs1*-like elements were found to be present at a very low copy number of one to five per haploid genome (Johns et al., 1985). Our first sequence analyses of *Bs1* (Jin and Bennetzen, 1989) indicated that this 3203-bp element had the LTRs and primer binding sites essential in *cis* for retrotransposon function but did not show impressive homology with any other nucleic acid sequence recorded in the EMBL or GenBank data bases (Devereux et al., 1984). This included a lack of convincing homology with the reverse transcriptase moiety so highly conserved in retroelements from animals, fungi, and plants. Hence, we concluded that the *Bs1* element is a defective derivative of an "autonomous" retrotransposon and can only transpose with the provision of reverse transcriptase activity in *trans*. Some cellular pseudogenes, those lacking promoters and introns but having poly(A) at their 3′ ends, are presumed to have a similar etiology.

In this study, we present evidence that *Bs1* is indeed a defective retroelement that has acquired a portion of a standard maize gene. The integrated fragment from the maize plasma membrane proton ATPase gene has been significantly conserved despite numerous secondary mutations subsequent to its acquisition by *Bs1*. These data indicated either that a

[1] To whom correspondence should be addressed.

retrotransposon can acquire a portion of a cellular gene, as retroviruses often do, or that *Bs1* is actually a defective version of a yet-to-be-discovered plant retrovirus.

RESULTS

Cloning and Sequencing of a Plasma Membrane Proton ATPase Gene from Maize

A comparison of the *Bs1* sequence with all DNA and protein sequences in the GenBank and EMBL data bases in 1988 yielded no strong nucleotide or predicted amino acid homology with any other entry, except weak homology to various fungal and animal cation-transporting ATPases. Careful searches for highly conserved features in other retroelements did identify a 5′ tRNA homology that could be used to prime minus strand DNA synthesis, a 3′ polypurine stretch that would facilitate the priming of plus strand DNA synthesis, and plausibly significant homologies to the conserved domains of protease and nucleic acid binding proteins (Jin and Bennetzen, 1989). A subsequent data base search in 1990 indicated an interrupted 654-bp stretch in Arabidopsis plasma membrane proton ATPase (PMPA) genes, *Aha1* (Harper et al., 1989) and *Aha3* (Pardo and Serrano, 1989). *Aha1* and *Aha3* had 69% nucleotide and ~70% predicted amino acid sequence identity with the *Bs1* region that has weak homology to various cation-transporting ATPases (data not shown). These observations suggested that the *Bs1* element contains a highly conserved ATPase function, unlike any other known retroelement, and/or that *Bs1* had acquired PMPA sequences from a maize gene.

Gel blot hybridization analysis of genomic maize DNA with probes from various sections of *Bs1* indicated several restriction fragments that cross-hybridized with the element (Figure 1). One of these fragments was unique in hybridizing only to the region of *Bs1* that contained the proton ATPase gene homology (Figure 1C).

We isolated several clones with homology to the entire *Bs1* element from a library of recombinant λ phage containing maize DNA that had been partially digested with restriction enzyme Sau3A. Three of these recombinant clones exhibited homology only to the ATPase-conserved region of *Bs1*. The inserts in these three phage contain 8.1 kb of overlap and together cover a contiguous stretch of ~20 kb of the maize genome (Figure 2).

We sequenced a contiguous 9285 bp in and around the region of homology between *Bs1* and the Arabidopsis PMPA gene from the C18 phage clone (Figure 2). More than 1.4 kb of overlapping sequence between the different clones was determined and yielded identical results. The sequence data obtained indicated that the cloned segment contains a maize plasma membrane proton ATPase gene (Figure 3), which we have named *Mha1*. Comparisons to PMPA gene sequences from other plant species demonstrated that *Mha1* is most similar to *Pma1* (Perez et al., 1992) and *Pma2* (Boutry et al., 1989), which are PMPA genes from tobacco. *Pma1* and *Mha1* contain introns in 19 identical positions within the protein-coding region, but *Pma1* has an additional intron near its 3′ end. Overall, the introns are smaller in the maize gene than in *Pma1* and show no detected primary sequence conservation between the monocot and dicot genes. *Mha1* has the standard GT (5′) and AG (3′) splice sites at all of its introns except the sixth, which has the rarer GC (5′) and AG (3′) junctions. This lone exception and its placement are also seen in *Pma1* and in the

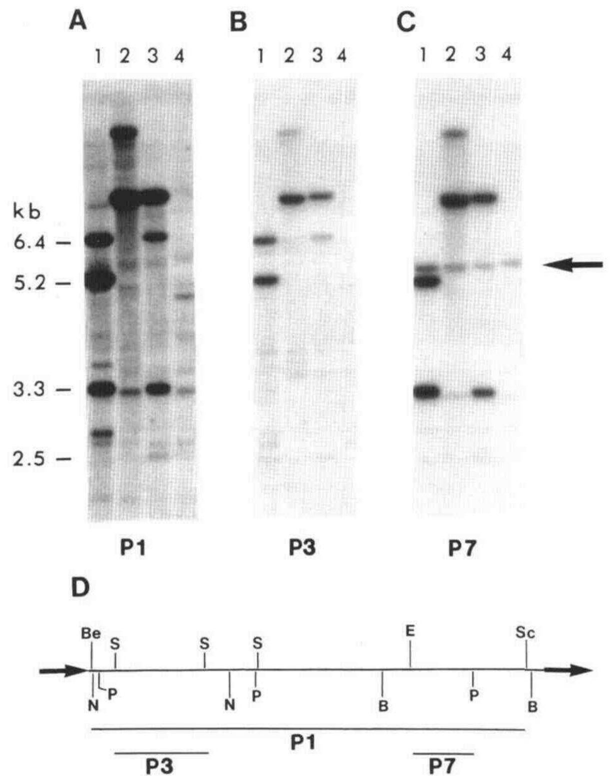

Figure 1. Hybridization of *Bs1* DNA Probes to Different Maize Lines.

The results of sequential hybridization of the indicated probes to BstEII- and ScaI-digested total maize DNA from four different maize lines (Sg18, Kp39, SA24, and Hy in lanes 1 to 4, respectively) are shown in **(A)**, **(B)**, and **(C)**. The arrow in **(C)** indicates a 5.6-kb BstEII-ScaI fragment that was later cloned and found to contain most of a plasma membrane proton ATPase gene. Molecular length markers are given at left in kilobases.

(A) P1 is a 2477-bp BstEII-ScaI fragment from *Bs1* used as a probe in this filter hybridization experiment.

(B) P3 is a 531-bp SalI fragment from *Bs1* used as a probe in this filter hybridization experiment.

(C) P7 is a 370-bp EcoRI-PstI fragment from *Bs1* used in this filter hybridization experiment.

(D) Positions of these probes within the 3203-bp *Bs1* element, with the LTRs of *Bs1* indicated by the bold arrows. B, BamHI; Be, BstEII; E, EcoRI; N, NcoI; P, PstI; S, SalI; Sc, ScaI.

partial genomic sequence of *Pma2* (Perez et al., 1992). The rarer GC (5′) splice site was previously observed in one other plant gene (Katinakis and Verma, 1985) and has been documented in a number of animal genes (Jackson, 1991).

Mha1 Transcript and Linkage Mapping

Because the C-terminal 21 amino acid sequence of *Mha1* is different from that of other plant PMPA genes (Figure 3), we have used 3′ rapid amplification of cDNA ends (3′ RACE) (Frohman et al., 1988) to determine its polyadenylation site and to show that *Mha1* is expressed at a low level in developing seedlings (data not shown). From the DNA sequence of a polymerase chain reaction (PCR) product obtained by 3′ RACE of seedling mRNA, the poly(A) tail of *Mha1* was found at 118 or 119 nucleotides downstream from the stop codon. The cDNA sequence also confirmed the absence of a last intron in *Mha1* like that seen in *Pma1* (Perez et al., 1992). This region of the *Mha1* genomic sequence has a simple sequence repeat of 5′-AT-3′ (34 tandem copies), and the polyadenylation site is located after the first four such repeats (Figure 2). The same region also provided a PCR-detected polymorphism that allowed us to map the *Mha1* gene in a standard set of recombinant inbreds (Burr et al., 1988). By using this analysis, we mapped *Mha1* at approximately the same position as UMC36A and NPI294A on the long arm of chromosome 2 (data not shown). Gel blot analysis indicated only one strong band in both maize and sorghum, which is a close maize relative, when *Mha1* or its flanking DNA sequences were used as a hybridization probe (data not shown).

Within the protein-encoding regions of its exons, the predicted 949–amino acid sequence of *Mha1* shows 85.5% (92%), 85.8% (91%), 85.4% (92%), 84.8% (93%), and 76.7% (88%) amino acid identity (similarity) to the tobacco *Pma1*, potato *Pha1*, tomato *Lha1*, rice *Osa1*, and Arabidopsis *Aha3* PMPA genes, respectively (Figure 3). All four known tobacco PMPA genes (*Pma1*, *Pma2*, *Pma3*, and *Pma4*) (Perez et al., 1992; GenBank data base, accession number X66737) show higher amino acid identity or similarity to *Mha1* than to any of the three Arabidopsis PMPA genes (*Aha1*, *Aha2*, and *Aha3*) (Harper et al., 1989, 1990; Pardo and Serrano, 1989). At the DNA sequence level, the *Mha1* coding sequence shows 74.2, 74.1, 74.3, 76.5, and 69.4% identity to *Pma1*, *Pha1*, *Lha1*, *Osa1*, and *Aha3*, respectively. These genes exhibit no apparent homology either 5′ or 3′ to the coding regions or within known introns.

Insertion of *Mha1* Sequences into *Bs1*

Comparison of the sequence of *Mha1* to *Bs1* indicates a 654-bp region of homology, starting with the last amino acid of exon 4, continuing through all of exons 5 to 9, and ending with the twenty-fourth amino acid of exon 10 (Figure 4). *Mha1* intron sequences are not found in *Bs1*. The precise nucleotide junction is not clear at the 5′ end of the insert, because the five-nucleotide sequence of *Bs1* at this presumptive junction

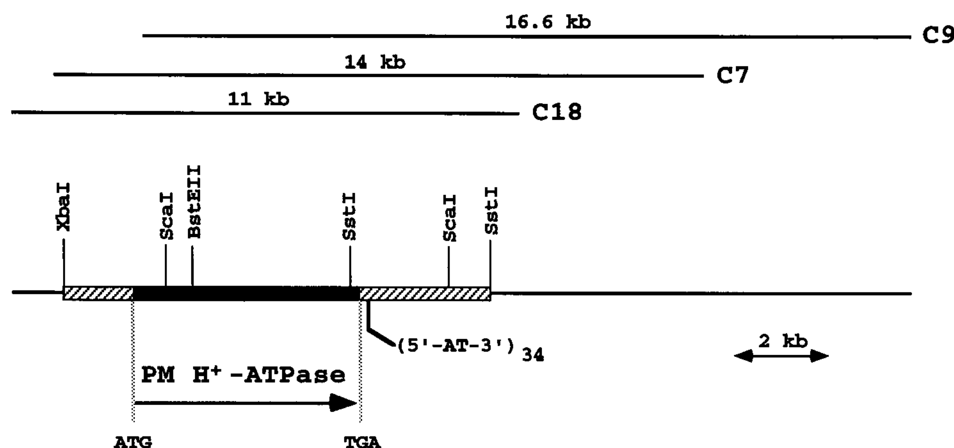

Figure 2. Overlap and Sequenced Regions of Clones Carrying a Plasma Membrane Proton ATPase Gene from Maize.

Three clones (C7, C9, and C18) with DNA sequence homology to the P7 probe (Figure 1), but not to the *Bs1* LTRs or other internal probes, were selected from a Sau3A partial library of maize DNA inserted into Charon 40. The upper lines show the length and overlap of the inserts in these three clones. The bottom line indicates the extent of the maize genome cloned in these phage, and the hatched rectangle shows the 9285-bp contiguous region sequenced from the C18 insert. The black rectangle indicates the exons and introns between the ATG at the predicted translation start site and the TGA at the predicted translation stop site of the maize plasma membrane (PM) H^+-ATPase gene. The 5′ end of the RNA specified by this gene is not known, but the 3′ end terminates within the fifth repeat of the simple sequence repeat designated $(5'\text{-AT-}3')_{34}$. The 9285-bp sequence of this maize plasma membrane H^+-ATPase gene, *Mha1*, and flanking DNA has been submitted to GenBank as accession number U09989.

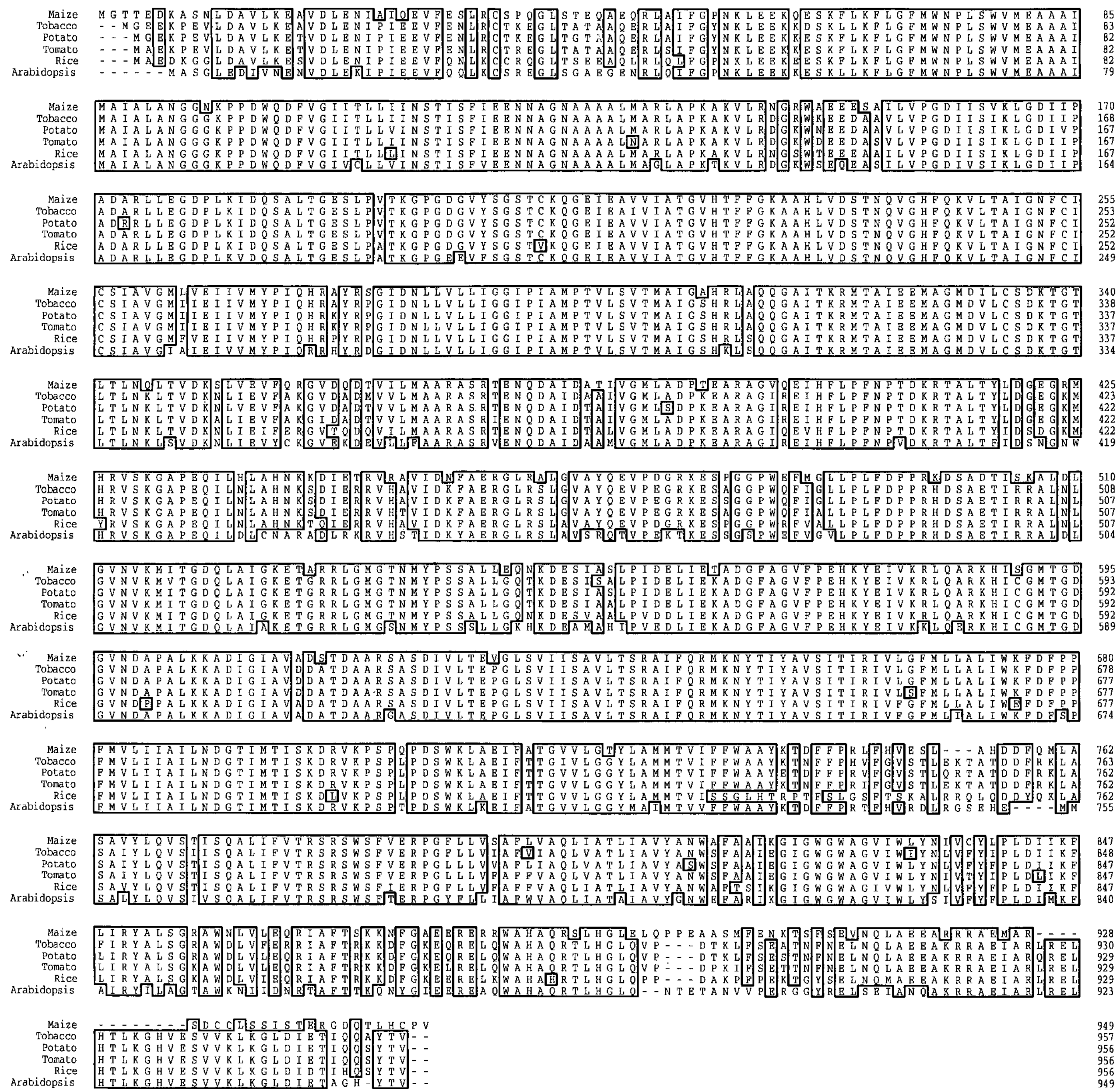

Figure 3. Comparison of Predicted Amino Acid Sequences for Several PMPA Genes.

Boxes are drawn only to show comparisons in which five or more of the sequences of these six genes predict the same amino acid. Sequences are shown for the entire predicted proteins from the initiating methionine to the C-terminal amino acid. The maize gene is *Mha1*; the tobacco gene is *Pma1*; the potato gene is *Pha1*; the tomato gene is *Lha1*; the rice gene is *Osa1*; and the Arabidopsis gene is *Aha3*. Dashes indicate the absence of corresponding amino acids in each gene.

(5′-TA/AAG-3′; underlined in Figure 4) is identical to the last five-nucleotide sequence of the fourth intron of *Mha1* (5′-TAAAG/-3′). Nevertheless, amino acid sequence identity begins with the last amino acid of exon 4. The proton ATPase insertion in *Bs1* maintains the *Mha1* reading frame within *Bs1*'s longest open reading frame (ORF). Hence, translation of *Bs1* RNA would yield a 740-amino acid ORF1 product (Jin and Bennetzen, 1989) that terminates with 12 additional amino acids after the 218 amino acids of *Mha1* sequence.

In the region of shared homology between *Bs1* and *Mha1*, the *Bs1* sequence exhibits 72.6% nucleotide identity with the tobacco *Pma1* gene and 87.6% nucleotide identity with the

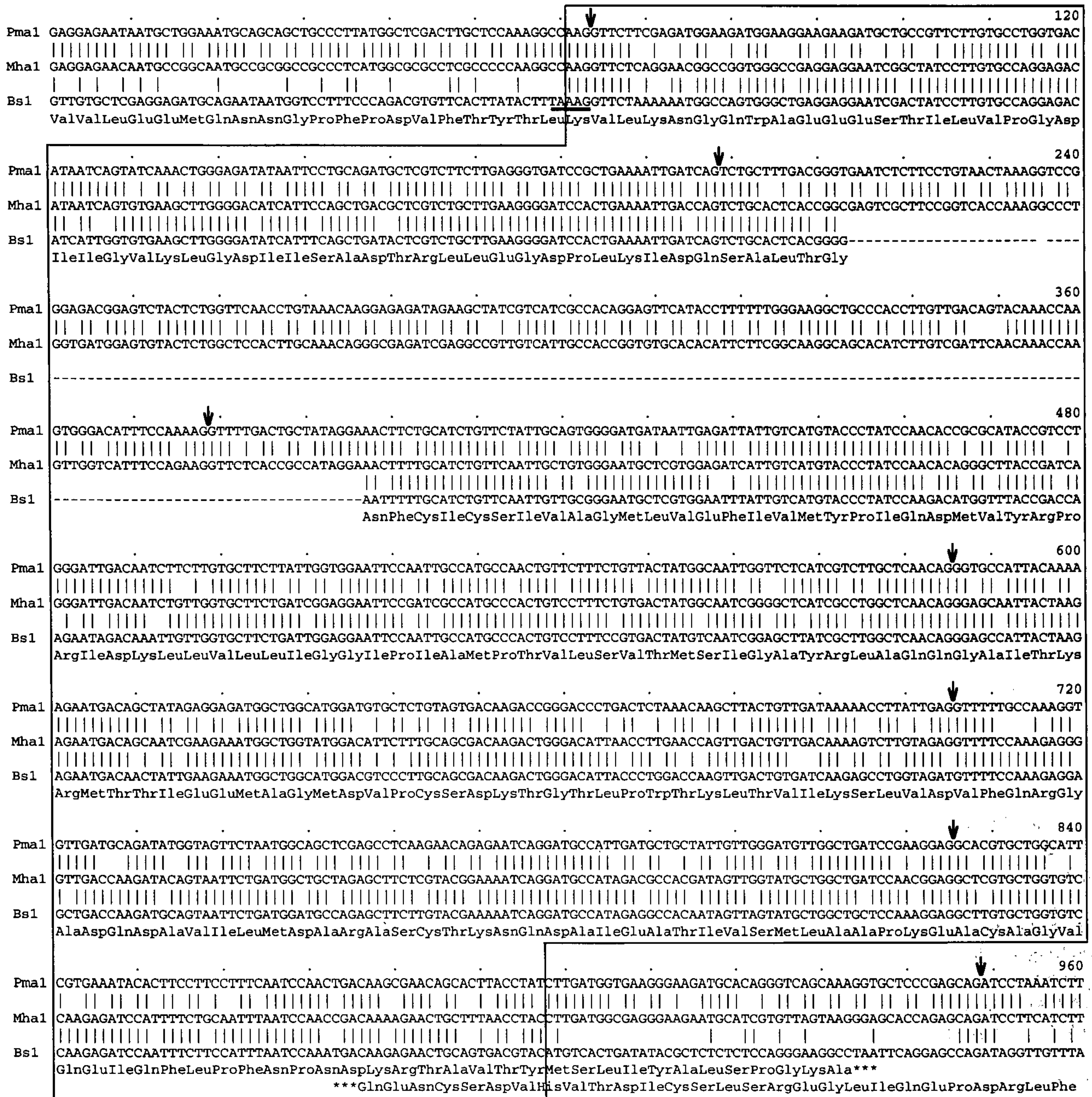

Figure 4. Extent and Sequence Evolution of *Mha1* Sequences Acquired by *Bs1*.

Comparison of the nucleotide sequences of homologous regions in *Bs1* and the exons of *Mha1* and *Pma1*. Homologous regions are contained within the box. Vertical lines designate nucleotide identity; the horizontal dashed line indicates the 183-bp deletion of *Mha1* sequences within *Bs1*. Vertical arrows show the positions of seven introns in this region in both *Pma1* and *Mha1*. Underlined nucleotides indicate the ambiguous 5′ junction of the *Mha1* insert in *Bs1*. Asterisks indicate nonsense codons.

Mha1 sequence. In this same region, *Pma1* and *Mha1* exhibit 76.5% nucleotide identity (Figure 4). Hence, the data support the conclusion that the PMPA-encoding sequences that are inserted into *Bs1* were derived from a portion of *Mha1* long after the divergence of maize and tobacco. Moreover, the absence of sequences homologous to the maize ATPase introns indicates either that the insertion into *Bs1* was derived from transcript sequences via an RNA recombination event or,

more likely, that intron sequences within the acquired *Mha1* sequence were subsequently lost upon RNA-mediated transposition of a *Bs1* ancestor (Bishop, 1983).

Mha1 Sequence Evolution within *Bs1*

Our sequence of *Mha1* indicates 82 mutational events that differentiate the homologous 654-bp regions in *Bs1* and *Mha1*. Most of these changes (81) were single nucleotide substitutions, but a single deletion of 183 bp is also observed (Figures 4 and 5). Eight of the *Bs1* codons exhibit two nucleotide changes, close to the 9.5 codons that would be predicted to have two changes if all changes were a result of independent, single nucleotide events. Therefore, compensating small (e.g., one or two nucleotide) insertions and deletions are not likely to have occurred. None of these 82 changes altered the frame of the *Bs1*-encoded polypeptide or created a nonsense codon. Hence, any essential protein product(s) encoded 3′ to the insertion would still be properly translated.

Of the 81 single nucleotide changes observed, 51 were transition events and 30 were transversion events. Over half of these changes (43 of 81) occurred in the third codon position. These 81 nucleotide changes led to 39 amino acid changes. Calculations based on the likelihood of mutation at any given codon position and of the frequencies of transition versus transversion events indicated that 54 amino acid changes would have been predicted with this level of nucleotide change (81 of 654). Hence, as with the maintenance of frame and the bias of mutations to the third codon position, the nucleotide changes observed were selected to lower changes in the amino acid sequence of the inserted segment.

When the same 218 amino acid sequences of 11 other plant PMPA genes are compared to the *Mha1* sequence, *Pma1*, *Pma3*, and *Osa1* show the lowest number of amino acid changes (23), and the two Arabidopsis PMPA genes (*Aha2* and *Aha3*) show the highest number of amino acid changes (37 and 40, respectively). However, when compared to the respective 24 and 29 conservative amino acid changes in *Aha2* and *Aha3*, only 10 of the 39 amino acid changes in *Bs1* are conservative. When the 218 amino acid sequences of all 12 known PMPA genes including *Mha1* are aligned, 164 invariable amino acid sites and 54 variable amino acid sites are found. However, 25 of the 39 amino acid changes in *Bs1* are located at invariable sites. Hence, although *Mha1* sequences within *Bs1* have been conserved, they have not been as conserved as these same sequences are during the evolution of PMPA genes. This could be a result of the *Mha1* sequences in *Bs1* contributing a less precisely defined or somewhat different function in *Bs1* than they do in a PMPA gene.

The portion of *Mha1* that has been acquired by *Bs1* contains the two highly conserved regions specifying the phosphatase and phosphorylation/transduction domains but not the highly conserved domains that bind ATP or the region that specifies kinase activity (Sussman and Surowy, 1987; Serrano, 1989). The 183-bp, 61–amino acid deletion that interrupts the contiguous stretch of *Bs1*/*Mha1* homology is located between the phosphatase and phosphorylation domains but removes two amino acids of the conserved phosphatase amino acid motif (TGES to TG). Hence, the portion of *Mha1* in *Bs1* contains only the well-conserved transmembrane anchoring sequences and the phosphorylation domain (CSDKTGTL) conserved in all plasma membrane cation-transporting ATPases in eukaryotes (Sussman and Surowy, 1987; Serrano, 1989).

DISCUSSION

Plasma Membrane Proton ATPase Genes

Several plasma membrane proton ATPase (PMPA) genes have now been cloned from plants: *Aha1*, *Aha2*, and *Aha3* from Arabidopsis (Harper et al., 1989, 1990; Pardo and Serrano, 1989); *Pma1*, *Pma2*, *Pma3*, and *Pma4* from tobacco (Perez at al., 1992; GenBank data base, accession number X66737); *Pha1* and *Pha2* from potato (GenBank data base, accession numbers X76536 and X76535, respectively); *Lha1* and part of *Lha2* from tomato (Ewing et al., 1990); *Osa1* from rice (Wada et al., 1992); and *Mha1* from maize. These genes are believed to be responsible for controlling intracellular and extracellular pH, and for generating the proton gradients across the plasma

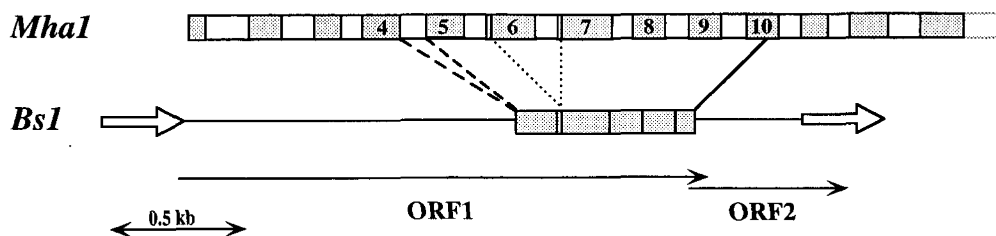

Figure 5. Schematic Drawing of the Structure of *Mha1* Compared to That of *Bs1*.

Exons are shaded and exons 4 through 10 are numbered. The lines connecting *Mha1* with *Bs1* show the end points of the regions of homology: the dashed lines indicate the two possible junction sites at this end (see text), and dotted lines show the end points of the 183-bp deletion in *Bs1*. The open arrows indicate the *Bs1* LTRs, and the thin arrows below the *Bs1* diagram indicate its two longest open reading frames.

membrane essential to nutrient uptake and cell elongation (Serrano, 1989). The higher level of homology between *Mha1* and the tobacco *Pma1* gene than with rice or Arabidopsis PMPA genes (Figure 3) is probably a result of the fact that *Pma1* and *Mha1* represent the same class of PMPA gene. When we employed exon-containing portions of *Mha1* as a hybridization probe in gel blot analysis of maize DNA, we generally observed one strong band and two or three weak bands. We expect that these weak bands are other PMPA genes that are likely to have higher homology to the cloned Arabidopsis and rice genes. We have used an *Mha1* probe to identify one strong and two or three weak cross-hybridizing bands in a close relative of maize, sorghum. One of the weakly *Mha1*-homologous bands was polymorphic and mapped in sorghum by DNA marker technology (Melake-Berhan et al., 1993).

Beyond higher sequence homology with *Pma1* than with other plant PMPA genes, *Mha1* also shares with *Pma1* an additional five introns not seen in the Arabidopsis PMPA genes, *Aha2* and *Aha3*. Compared to *Aha3*, *Aha2* is missing one intron (intron 14), and *Mha1* is missing the last intron present in both Arabidopsis and tobacco PMPA genes. The 14 introns shared by all four PMPA genomic sequences investigated to date (*Aha2*, *Aha3*, *Pma1*, and *Mha1*) vary somewhat in size and greatly in sequence, but all are positioned at identical sites and have high conservation of sequences near the intron-exon junctions. The five extra introns found in *Mha1* and *Pma1* are also identically placed in the genes, and one of these (number 6) has an unusual GC/AG 5′/3′ junction conserved in both species. These results provide further support for the generation of PMPA gene families long before the divergence of monocots and dicots (some 200 million years ago) (Wolfe et al., 1989) and suggest subsequent differential loss of introns among PMPA gene families in higher plants.

Acquisition of *Mha1* Gene Homology by *Bs1*

A high level of similarity between *Mha1* and a contiguous 654-bp segment within *Bs1* indicates that a portion of this maize PMPA gene was acquired by *Bs1*. Because this homologous region in *Bs1* lacks the introns from *Mha1*, this segment could have been acquired by RNA recombination (Lai, 1992) with an ancestral *Bs* element transcript. This would have required either a conversion or double recombination event. More likely, the *Mha1*-homologous sequence in *Bs1* might have been acquired by an ancestral *Bs* element by the same template switch mechanism proposed for transduction by oncogenic retroviruses (Swain and Coffin, 1992; Zhang and Temin, 1993).

Cloning and gel blot hybridization analysis indicated that several other *Bs1*-like elements with internal *Mha1*-homologous sequences exist in maize, the teosintes, and *Tripsacum* species (*T. dactyloides*, *T. jalapense*, and *T. zopilotense*) (Y.-K. Jin and J.L. Bennetzen, unpublished results). This suggests that an ancestral *Bs* element acquired *Mha1* sequence prior to the divergence of these species, over one million years ago (Gaut and Clegg, 1991, 1993).

When *Bs1* was first cloned by Johns et al. (1985), their hybridization experiments suggested that the *Bs1* element might be present in a broad range of species. Although this still may be true, we have found that at least some of the hybridization they observed is a result of the *Mha1*-homologous region in *Bs1*.

Conservation and Possible Function of *Mha1* Sequences in *Bs1*

The 82 changes that differentiate the homologous regions in *Bs1* and *Mha1* have maintained the same reading frame and have led to fewer amino acid substitutions than expected from neutral events. Even the 183-bp deletion within exons 6 and 7 did not alter reading frame and primarily removed the region between two PMPA functional domains. Frameshift events are not seen, and most of the single base pair substitutions detected were limited to the third codon position (43 of 81). Moreover, most of the single nucleotide changes observed are transitions (51 of 81), which cause amino acid changes less commonly than do transversions or frameshifts. The overall amino acid sequence identity of *Mha1*-homologous regions in *Bs1* and *Mha1* (82%) is similar to that of *Mha1* to other plant PMPA genes (76 to 86%). Two functional regions of *Mha1*, the transmembrane and phosphorylation domains, are particularly well conserved in the *Mha1*-homologous region of a predicted *Bs1* ORF1 peptide. The high degree of conservation of these two functional domains in *Bs1* implies that they provide some selective advantage to the element.

Bs1 differs from many retrotransposons in not encoding any peptide with convincing homology to reverse transcriptase. For this reason, we have concluded that *Bs1* is a defective retrotransposon (Jin and Bennetzen, 1989). Conversely, Johns et al. (1989) had postulated that *Bs1* might be a highly compact retrotransposon that potentiates multiple translational frameshifts and/or RNA splicing events that could create what would be, at best, a highly divergent reverse transcriptase. Our analysis of *Mha1* homology in *Bs1* indicates that this "best possible reverse transcriptase" region is probably located too close (just 5′) to the transmembrane component specified by *Mha1* to allow the production of a functional polymerase.

The transduction of cellular genes by retroviruses has been documented in several animal systems. In most cases, the acquired gene has an oncogenic function that provides a selective advantage for retrovirus production. Most of these oncogenic retroviruses are replication defective because parts of their genomes have been replaced by the oncogenes (Bishop, 1983; Varmus, 1984). The infectious replication and resultant survival of these defective RNA tumor viruses require the presence of a helper virus that provides the missing viral replication function in *trans* (Bishop, 1983; Varmus, 1984). *Bs1* appears to be similar in this regard, because *Mha1* homology has been inserted in a part of the retrotransposon where replication

functions are normally encoded. Because *Bs1* was first identified as a de novo insertion in the *Adh1* gene of maize (Johns et al., 1985), we postulated that the absence of reverse transcriptase coding potential implies that replication functions must have been provided in *trans* by another retroelement (Jin and Bennetzen, 1989). Although reverse transcriptase could, potentially, be provided by an unrelated retroelement, it could also be provided by an "autonomous" *Bs* element.

The selective advantage of the *Mha1* function(s) acquired by *Bs1* is not clear. Selection could act at the level of the whole organism (i.e., maize) or on *Bs1* itself as a "selfish" DNA (Orgel and Crick, 1980). The most obvious potential added to a *Bs1* polypeptide by the acquisition of *Mha1* sequences is the ability to associate with the plasma membrane. There is, as yet, no reason to believe such a membrane association would be of value to a retrotransposon, but essential plasma membrane localization of the *env* gene product is an attribute associated with retroviruses (Varmus and Brown, 1989). Moreover, the fact that no retrotransposon had been previously observed to acquire part of a cellular gene, as retroviruses often do, suggests that *Bs1* may actually represent a defective proviral form of a retrovirus. Although no plant retrovirus has yet been observed in nature, there is no a priori reason why this class of virus should not exist. Viruses present at very low levels are often difficult to detect, and a *Bs1*-related retrovirus might now be extinct in its fully functional form. However, in our recent studies of maize genomic sequences with homology to *Bs1*, we have found one element that is larger than *Bs1* in size but has no homology to *Mha1* (Y.-K. Jin and J.L. Bennetzen, unpublished observations). Further studies will include functional and structural analysis of this retroelement.

METHODS

Materials

Maize lines Hy, Kp39, SA24, and Sg18 were provided by B. Ashman (Purdue University, West Lafayette, IN). Restriction enzymes were purchased from Bethesda Research Laboratories, New England Biolabs, and Promega and used according to the conditions specified by the manufacturer. A maize genomic DNA library from a *Mutator* maize line (D3L) (Bennetzen et al., 1988), consisting of total DNA partially digested with Sau3A and inserted into the BamHI site of Charon 40 (Dunn and Blattner, 1987), was prepared and provided by S. Hulbert (Kansas State University, Manhattan, KS).

Gel Blot Hybridization Analysis

Maize DNA was prepared by the method of Saghai-Maroof et al. (1984). Approximately 6 to 8 μg of total DNA was digested with the indicated restriction enzyme and electrophoretically resolved through a 0.6% agarose gel. The gel was blotted onto a nylon filter (Micron Separations Inc., Westboro, MA) and baked for 2 hr at 80°C under vacuum. Hybridization and washing were performed as previously described (Bennetzen, 1984). The hybridization probes used were specific segments of *Bs1*, as described in Figure 1.

Cloning

The 2477-bp BstEII-ScaI fragment of *Bs1* (Jin and Bennetzen, 1989) was used as a hybridization probe to screen a recombinant library consisting of total maize DNA partially digested with Sau3A and inserted into the BamHI site of Charon 40. Eighteen clones were identified and purified from a screening of $\sim 5 \times 10^6$ recombinant phage. Subsequent hybridizational screening with a 330-bp fragment containing primarily *Bs1* long terminal repeat (LTR) sequences and with a 370-bp fragment containing only the region of homology with plasma membrane proton ATPase (PMPA) genes in *Bs1* identified three clones that had no LTR homology but did have PMPA homology. These three clones (C7, C9, and C18) were restriction mapped and found to have an 8.1-kb region of overlap (Figure 2).

DNA Sequencing and Sequence Analysis

Overlapping and adjacent regions from clones C7, C9, and C18 were variously subcloned into pUC18 or pUC19 (Yanisch-Perron et al., 1985). Large subclones were deleted to sizes suitable for DNA sequencing either by further restriction enzyme digestion or by exonuclease III and mung bean nuclease deletion (Maniatis et al., 1982). Double-stranded plasmid DNAs were used for universal primer-directed dideoxy sequencing (Sanger et al., 1977). Both dGTP and dITP reactions were run for regions that exhibited gel compression problems. Dideoxy DNA sequencing reactions were resolved on 43-cm-long, 0.4-mm-thick gels composed of 5% (w/v) polyacrylamide and 7 M urea. All regions of the C18 clone were sequenced on both strands, and 1.4 kb of overlapping DNA sequence was derived from clones C7 and C9. This overlapping sequence was identical between the three independent clones. DNA sequences were read and entered by hand and were analyzed on the University of Wisconsin Genetics Computer Group software programs (Devereux et al., 1984) run through Purdue's AIDS Center Laboratory for Computational Biochemistry VAX cluster.

3′ RACE Analysis

3′ Rapid Amplification of cDNA Ends (3′ RACE) analysis (Frohman et al., 1988) of *Mha1* was performed with cDNA prepared from mRNA of 4-day-old maize seedlings. The 34 base oligo(dT) primer used for cDNA synthesis was 5′-GACTCGAGTCGACATCGTTTTTTTTTTTTTTTTT-3′. The first polymerase chain reaction (PCR) amplification of the cDNA was performed with a 3′-specific primer from exon 19 (5′-CTCATCCGGTACGCATT-3′) and an adapter primer (5′-GACTCGAGTCGACATCG-3′). The second amplification was performed with a 3′-specific primer from exon 20 (5′-CAGTGAAGTGAACCAGCTC-3′) and the same adapter primer. The 3′ RACE product purified from a 10% polyacrylamide gel was directly sequenced.

ACKNOWLEDGMENTS

We thank Benjamin Burr for assistance in the placement of *Mha1* on the maize genetic map; Scot Hulbert for providing a recombinant λ

library containing maize DNA; and Phillip SanMiguel for assistance in computer searches, *Mha1* RNA analysis, and PCR-based mapping of *Mha1*. This research was supported by National Science Foundation Grant No. DMB-8552557, U.S. Department of Agriculture National Research Initiative Competitive Grants Program Grant No. 91-37300-6498, and by National Institutes of Health Grant No. AI27713 in support of Purdue's AIDS Center Laboratory for Computational Biochemistry.

Received March 30, 1994; accepted June 6, 1994.

REFERENCES

Bennetzen, J.L. (1984). Transposable element *Mu1* is found in multiple copies only in *Mutator* maize lines. J. Mol. Appl. Genet. **2,** 519–524.

Bennetzen, J.L., Qin, M.M., Ingels, S., and Ellingboe, A.H. (1988). Allele-specific and *Mutator*-associated instability at the *Rp1* disease resistance locus of maize. Nature **332,** 369–370.

Bishop, J.M. (1983). Cellular oncogenes and retroviruses. Annu. Rev. Biochem. **52,** 301–354.

Boutry, M., Michelet, B., and Goffeau, A. (1989). Molecular cloning of a family of plant genes encoding a protein homologous to plasma membrane H⁺-translocating ATPases. Biochem. Biophys. Res. Commun. **162,** 567–574.

Burr, B., Burr, F.A., Thompson, K.A., Albertson, M.C., and Stuber, C.W. (1988). Gene mapping with recombinant inbreds in maize. Genetics **118,** 519–526.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. **12,** 387–395.

Dunn, I.S., and Blattner, F.R. (1987). Charons 36 to 40: Multienzyme, high capacity, recombination deficient replacement vectors with polylinkers and polystuffers. Nucl. Acids Res. **15,** 2677–2698.

Ewing, N.N., Wimmers, L.E., Meyer, D.J., Chetelat, R.T., and Bennett, A.B. (1990). Molecular cloning of tomato plasma membrane H⁺-ATPase. Plant Physiol. **94,** 1874–1881.

Finnegan, D.J. (1983). Retoviruses and transposable elements–which came first? Nature **302,** 105–106.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNA from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85,** 8998–9002.

Gaut, B.S., and Clegg, M.T. (1991). Molecular evolution of alcohol dehydrogenase 1 in members of the grass family. Proc. Natl. Acad. Sci. USA **88,** 2060–2064.

Gaut, B.S., and Clegg, M.T. (1993). Molecular evolution of the *Adh1* locus in the genus *Zea.* Proc. Natl. Acad. Sci. USA **90,** 5095–5099.

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

Harper, J.F., Surowy, T.K., and Sussman, M.R. (1989). Molecular cloning and sequence of cDNA encoding the plasma membrane proton pump (H⁺-ATPase) of *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. USA **86,** 1234–1238.

Harper, J.F., Manney, L., DeWitt, N.D., Yoo, M.H., and Sussman, M.R. (1990). The *Arabidopsis thaliana* plasma membrane H⁺-ATPase multigene family. J. Biol. Chem. **265,** 13601–13608.

Jackson, I.J. (1991). A reappraisal of non-consensus mRNA splice sites. Nucl. Acids Res. **19,** 3795–3798.

Jin, Y.-K., and Bennetzen, J.L. (1989). Structure and coding properties of *Bs1*, a maize retrovirus-like transposon. Proc. Natl. Acad. Sci. USA **86,** 6235–6239.

Johns, M.A., Mottinger, J., and Freeling, M. (1985). A low copy number *copia*-like transposon in maize. EMBO J. **4,** 1093–1102.

Johns, M.A., Babcock, M.S., Fuerstenberg, S.M., Fuerstenberg, S.I., Freeling, M., and Simpson, R.B. (1989). An unusually compact retrotransposon in maize. Plant Mol. Biol. **12,** 633–642.

Katinakis, P., and Verma, D.P.S. (1985). Nodulin-24 gene of soybean codes for a peptide of the peribacteroid membrane and was generated by tandem duplication of a sequence resembling an insertion element. Proc. Natl. Acad. Sci. USA **81,** 2733–2737.

Lai, M.M.C. (1992). RNA recombination in animal and plant viruses. Microbiol. Rev. **56,** 61–79.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Melake-Berhan, A., Hulbert, S.H., Butler, L.G., and Bennetzen, J.L. (1993). Structure and evolution of the genomes of *Sorghum bicolor* and *Zea mays.* Theor. Appl. Genet. **86,** 598–604.

Orgel, L.E., and Crick, F.H.C. (1980). Selfish DNA: The ultimate parasite. Nature **284,** 604–607.

Pardo, J.M., and Serrano, R. (1989). Structure of a plasma membrane H⁺-ATPase gene from the plant *Arabidopsis thaliana.* J. Biol. Chem., **264,** 8557–8562.

Perez, C., Michelet, B., Ferrant, V., Bogaerts, P., and Boutry, M. (1992). Differential expression within a three-gene subfamily encoding a plasma membrane H⁺-ATPase in *Nicotiana plumbaginifolia.* J. Biol. Chem. **267,** 1204–1211.

Pfeiffer, P., and Hohn, T. (1983). Involvement of reverse transcription in the replication of cauliflower mosaic virus: A detailed model and test of some aspects. Cell **33,** 781–789.

Saghai-Maroof, M.A., Soliman, K.M., Jørgensen, R.A., and Allard, R.W. (1984). Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA **81,** 8014–8018.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74,** 5463–5467.

Serrano, R. (1989). Structure and function of plasma membrane ATPase. Annu. Rev. Plant Physiol. Plant Mol. Biol. **40,** 61–94.

Summers, J., and Mason, W.S. (1982). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. Cell **29,** 403–415.

Sussman, M.R., and Surowy, T.K. (1987). Physiology and molecular biology of membrane ATPases. Oxf. Surv. Plant Mol. Cell Biol. **4,** 47–70.

Swain, A., and Coffin, J.M. (1992). Mechanism of transduction by retroviruses. Science **255,** 841–845.

Temin, H.M. (1980). Origin of retroviruses from cellular moveable genetic elements. Cell **21,** 599–600.

Varmus, H.E. (1984). The molecular genetics of cellular oncogenes. Annu. Rev. Genet. **18**, 553–612.

Varmus, H.E., and Brown, P. (1989). Retroviruses. In Mobile DNA, D.E. Berg and M.M. Howe, eds (Washington, DC: American Society for Microbiology), pp. 53–108.

Wada, M., Takano, M., and Kasamo, K. (1992). Nucleotide sequence of a complementary DNA encoding plasma membrane H^+-ATPase from rice (*Oryza sativa* L.). Plant Physiol. **99**, 794–795.

Wolfe, K.H., Gouy, M., Yang, Y.-W., Sharp, P.M., and Li, W.-H. (1989). Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86**, 6201–6205.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**, 103–119.

Zhang, J., and Temin, H.M. (1993). Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science **259**, 234–238.