A New *Ac*-Like Transposon of Arabidopsis Is Associated With a Deletion of the *RPS5* Disease Resistance Gene

Adam D. Henk, Randall F. Warren and Roger W. Innes

Department of Biology, Indiana University, Bloomington, Indiana 47405

Manuscript received November 11, 1998
Accepted for publication December 30, 1998

ABSTRACT

The *RPS5* and *RFL1* disease resistance genes of Arabidopsis ecotype Col-0 are oriented in tandem and are separated by 1.4 kb. The Ler-0 ecotype contains *RFL1*, but lacks *RPS5*. Sequence analysis of the *RPS5* deletion region in Ler-0 revealed the presence of an *Ac*-like transposable element, which we have designated *Tag2*. Southern hybridization analysis of six Arabidopsis ecotypes revealed 4–11 *Tag2*-homologous sequences in each, indicating that this element is ubiquitous in Arabidopsis and has been active in recent evolutionary time. The *Tag2* insertion adjacent to *RFL1* was unique to the Ler-0 ecotype, however, and was not present in two other ecotypes that lack *RPS5*. DNA sequence from the latter ecotypes lacked a transposon footprint, suggesting that insertion of *Tag2* occurred after the initial deletion of *RPS5*. The deletion breakpoint contained a 192-bp insertion that displayed hallmarks of a nonhomologous DNA end-joining event. We conclude that loss of *RPS5* was caused by a double-strand break and subsequent repair, and cannot be attributed to unequal crossing over between resistance gene homologs.

THE recognition of pathogens by plants is often mediated by dominant or semidominant disease resistance (R) genes, which are thought to encode receptors for pathogen-derived ligands (Hammond-Kosack and Jones 1997). Because of the diversity of potential pathogens, plant genomes are expected to possess large numbers of R genes with differing specificities. Consistent with this prediction, several dozen R gene homologs have been identified in Arabidopsis via sequencing of random cDNAs (Botella *et al.* 1997). This work has led to an estimate of ~100 R genes in the Arabidopsis genome (Botella *et al.* 1997).

A major question relative to R gene function is how R gene ligand specificity evolves, particularly in light of potentially rapid evolution among pathogens. Recent molecular data have shown that R genes often occur in tandem arrays of two or more R gene homologs (Martin *et al.* 1993; Jones *et al.* 1994; Whitham *et al.* 1994; Dixon *et al.* 1996; Ellis *et al.* 1997; Jia *et al.* 1997; Parker *et al.* 1997; Parniske *et al.* 1997; Song *et al.* 1997; Thomas *et al.* 1997; Warren *et al.* 1998). The tandem array structure of R gene clusters is thought to have major implications relative to the evolution of R genes (Pryor and Ellis 1993; Hulbert 1997; Parniske *et al.* 1997). For example, unequal crossing over between different homologs of a cluster could give rise to the rapid expansion and/or contraction of the cluster. More significantly, recombination between members of a cluster could give rise to R genes with novel specificities. Detailed genetic analyses of the *Rp1* complex of maize indicates that new specificities have arisen via such unequal crossovers (Richter *et al.* 1995). Molecular analyses of the *Cf4*/*Cf9* gene cluster in two tomato species indicate that recombination between family members has occurred, and furthermore, that some of the recombinant products likely have specificities that differ from the progenitor alleles (Parniske *et al.* 1997).

We recently described an R gene cluster on chromosome I of Arabidopsis ecotype Col-0 consisting of two genes, *RFL1* and *RPS5*, separated by 1.4 kb (Warren *et al.* 1998). *RPS5* confers resistance to strains of the bacterium *Pseudomonas syringae* that carry the avirulence gene *avrPphB*. The specificity of *RFL1* is unknown, but its coding region is 74% identical at the DNA level to *RPS5*. Southern hybridization analysis revealed that the Arabidopsis ecotype Landsberg *erecta* (Ler-0) contains *RFL1*, but completely lacks *RPS5* (Warren *et al.* 1998). The relatively high level of sequence divergence between *RFL1* and *RPS5* indicated that the lack of *RPS5* in Ler-0 was most likely the result of a recent deletion in Ler-0 rather than a duplication in Col-0. We wished to determine whether this deletion could be attributed to an unequal crossover event between *RPS5* and *RFL1*, as would be predicted by current models of R gene cluster evolution. We therefore isolated and sequenced this region from Ler-0 and from two additional ecotypes that lack *RPS5*. Here we describe a previously unknown *Ac*-like transposon that was found associated with the *RPS5* deletion site in Ler-0, but not in the latter two ecotypes. Contrary to expectations, the loss of *RPS5* appears to be the result of a double-strand break fol-

Corresponding author: Roger W. Innes, Department of Biology, Jordan Hall 142, Indiana University, Bloomington, IN 47405.
E-mail: rinnes@bio.indiana.edu

lowed by nonhomologous end joining and cannot be attributed to recombination between R gene homologs.

MATERIALS AND METHODS

Sources of bacteria and seed: *P. syringae* pathovar *tomato* strain DC3000 was obtained from D. Cuppels (Agricultural Canada-Research Center, London, Ontario, Canada). Construction of strains DC3000(*avrB*::Ω) and DC3000(*avrPphB*) (formerly called *avrPph3*) has been described previously (Innes *et al.* 1993; Simonich and Innes 1995). Seeds of Arabidopsis ecotypes were obtained from Dr. Brian Staskawicz (University of California at Berkeley; Col-0 and Ler-0) and from A. Kranz (Arabidopsis Information Service seed bank; Bch-1, Chi-0, Mt-0, and Ws-0). These ecotypes can now be obtained from the Arabidopsis Biological Resource Center at Ohio State University (Columbus, OH).

Screening cosmid libraries: A cosmid library of ecotype Ler-0 genomic DNA made by C. Lister and C. Dean (John Innes Centre, Norwich, UK) was obtained from the laboratory of J. Jones (Sainsbury Laboratory, Norwich, UK; Parker *et al.* 1997). The library was provided as 68 DNA pools, where each pool represented clones from a single 384-well microtiter plate. Pools containing a clone with the *RFL1* gene were identified by using *RFL1*-specific primers and PCR amplification. Cosmid DNA from positive pools was then used to transform *Escherichia coli* strain DH5-α, and colonies containing *RFL1* clones were identified by colony hybridization (Sambrook *et al.* 1989). Cosmid DNA was purified from *E. coli* using Qiagen Tip-100 columns (Qiagen, Chatsworth, CA). The presence of *RFL1* was confirmed by DNA sequence analysis.

DNA sequencing: Sequencing was performed using an ABI Dye Terminator FS kit protocol (Perkin Elmer, Foster City, CA) on an ABI Prism 377 DNA sequencer. DNA sequence from ecotype Ler-0 was determined by primer walking, using cosmid DNA as a template. For ecotypes Mt-0 and Bch-1, templates were prepared by PCR amplification from genomic DNA (see below). To avoid errors produced by *Taq* polymerase, the products of four independent PCR reactions were pooled, purified by filtration (Ultrafree-MC filter unit, 30,000-D cutoff; Millipore, Bedford, MA), and sequenced directly. Products of reverse transcriptase (RT-PCR) reactions (see below) were prepared and sequenced similarly. Evaluation of sequencing data and construction of sequence contigs was performed with the Sequencher software package for the Power Macintosh (GeneCodes Corporation, Ann Arbor, MI). Homology searches of the GenBank database were performed using the BLAST2 algorithm (Altschul *et al.* 1997), and alignment of sequences was performed using the GAP program of the Genetics Computer Group (Madison, WI) Wisconsin Package version 9.1.

RT-PCR: Total RNA was isolated from Arabidopsis ecotype Ler-0 rosette tissue using the RNeasy kit (Qiagen), and then used directly as template in RT-PCR reactions following the protocol of Kawasaki and Wang (1989). First-strand cDNA synthesis was primed with the *Tag2*-specific antisense primer 5′-TCGGAATGCTTAAGATATCAC-3′, and the subsequent PCR amplification was performed with this primer and the sense primer 5′-GATGTCTCAACCCGCTGGAA-3′, which flank a putative intron in *Tag2*. RT-PCR products were separated on an agarose gel and found to contain two predominant bands. One corresponded in size to full-length *Tag2* sequence (*i.e.*, unspliced). The second product was ~170 bp smaller, consistent with removal of the putative intron. The latter band was excised from the gel and reamplified by PCR using the same primers and then sequenced directly.

Southern hybridization: Genomic Arabidopsis DNA was prepared using the DNeasy kit (Qiagen). Approximately 1 μg of DNA of each Arabidopsis ecotype was digested with *Hin*dIII restriction enzyme and separated on 1.0% agarose gels. After electrophoresis, DNA was denatured and transferred to Hybond-N nylon membrane (Amersham, Arlington Heights, IL). DNA templates for probes were prepared by PCR amplification from a cosmid clone containing Ler-0 genomic DNA spanning the *Tag2* transposon. The primers used to generate the 143-bp *Tag2* end-probe were 5′-CATGGTCGGCCCGTAAAGAA-3′ and 5′-GCCGAGGGAGAGAGAAGAG-3′. Primers used to generate the 595-bp internal *Tag2* probe were 5′-TGGAGAGC TTTAACTGTTGA-3′ and 5′-GTTCTGCTCTTTCCCACTCC-3′. PCR products were labeled with [^{32}P]-dATP using a random primer protocol (Feinberg and Vogelstein 1983). Hybridization and wash conditions were as described by Ashfield *et al.* (1998). Radioactivity was visualized by exposing membranes to X-ray film (Fuji film RX, Fisher Scientific, Pittsburgh, PA).

Identification of Arabidopsis ecotypes that lack *RPS5*: Ecotypes lacking *RPS5* were identified by first screening for those that lacked *RPS5* function [resistance to *P. syringae* strain DC3000(*avrPphB*)], and then analyzing these ecotypes by Southern hybridization for the presence of *RPS5*. Resistance to DC3000(*avrPphB*) was assayed by immersing whole rosettes in a suspension of $\sim 2 \times 10^8$ cfu of DC3000(*avrPphB*) per ml ($OD_{600} = 0.2$) as described by Innes *et al.* (1993). Plants were scored for presence of disease symptoms (water-soaked pits and chlorosis) 4 and 5 days after inoculation. Southern hybridizations were performed using the hybridization and wash conditions described by Ashfield *et al.* (1998). Lines that displayed no hybridization with a probe that spanned the central third (850 bp) of the *RPS5* coding region were selected. The *RPS5* deletion region from these lines (Bch-1 and Mt-0) was amplified by PCR using the primers 5′-TGTGAGTGTTT TAGAGAAGGAG-3′ and 5′-GGGAAGAGGAGTAACGGAGA-3′ and sequenced directly as described above.

RESULTS

Ler-0 contains an *Ac*-like transposon insertion in the place of *RPS5*: In ecotype Col-0, *RFL1* and *RPS5* are oriented in the same direction, with the start codon of *RPS5* located 1.4 kb 3′ of the *RFL1* stop codon (Figure 1; Warren *et al.* 1998). In ecotype Ler-0, the entire *RPS5* open reading frame (ORF) appeared to be absent, but we did not know the endpoints of this deletion (Warren *et al.* 1998). To precisely define these endpoints, we isolated a Ler-0 cosmid clone containing *RFL1* (see materials and methods) and then, starting near the 3′ end of *RFL1*, used primer walking to determine the DNA sequence. This analysis revealed a 4.0-kb deletion in Ler-0 relative to Col-0 that started 114 bp 3′ of *RFL1* and included all of *RPS5*, ending 45 bp 3′ of *RPS5* (Figure 1). This was not a simple deletion, however, as we found 3298 bp of new sequence not present in Col-0 (Figure 2).

A search of the GenBank database with this new sequence revealed significant similarity to several transposable elements in the hAT family (*hobo*, *Ac*, and *Tam3*), with the highest similarity being to the *Ac* element of maize. As shown in Figure 2, translation of the 3297-bp insertion revealed two long ORFs that could be joined by excision of a putative 166-bp intron pre-

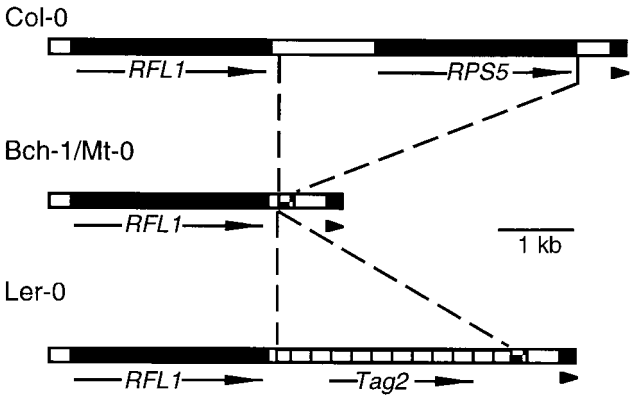


Figure 1.—Schematic representation of the *RFL1* region from four ecotypes of Arabidopsis. The *RPS5* gene is absent in ecotypes Bch-1, Mt-0, and Ler-0. In Bch-1 and Mt-0, *RPS5* has been replaced by a 192-bp sequence that is not found in Col-0 (checkered box; see Figure 5). In Ler-0, the *Tag2* transposon has inserted immediately adjacent to this sequence. The black box on the far right indicates the 5′ end of an ORF that overlaps with an Arabidopsis EST clone (GenBank accession no. N65072).

dicted by the NetPlantGene v. 2.1 program (Hebsgaard *et al.* 1996). We confirmed the presence of this intron by RT-PCR analysis (see materials and methods). Figure 3 shows that the resulting protein sequence shares several large blocks of identity with the Ac transposase, including a highly conserved domain present in the C terminus of all members of the hAT family (Calvi *et al.* 1991; Liu and Crawford 1998). We have named this putative transposase Tag2 [Tag1 was the first member of the hAt family identified in Arabidopsis (Tsay *et al.* 1993; Liu and Crawford 1998)]. The putative Tag2 transposase is more similar to the Ac transposase (33% identity) than are any other member of the hAT family, including Tag1, which cannot be confidently aligned with Ac outside the conserved C-terminal region.

In addition to the similarity in their transposases, members of the hAT family are characterized by the presence of short terminal inverted repeats and the formation of 8-bp target site duplications upon insertion (Saedler and Gierl 1996). As shown in Figure 2, *Tag2* also displays these characteristics. The terminal inverted repeats are 18 bp long. Multiple copies of the sequence TGGGC and TGGAC in both direct and inverted orientation are found immediately internal to the inverted repeats. Similar repeats are found in all hAT family members and are thought to function as transposase binding sites (reviewed in Saedler and Gierl 1996). The total length of the *Tag2* element is 3094 bp.

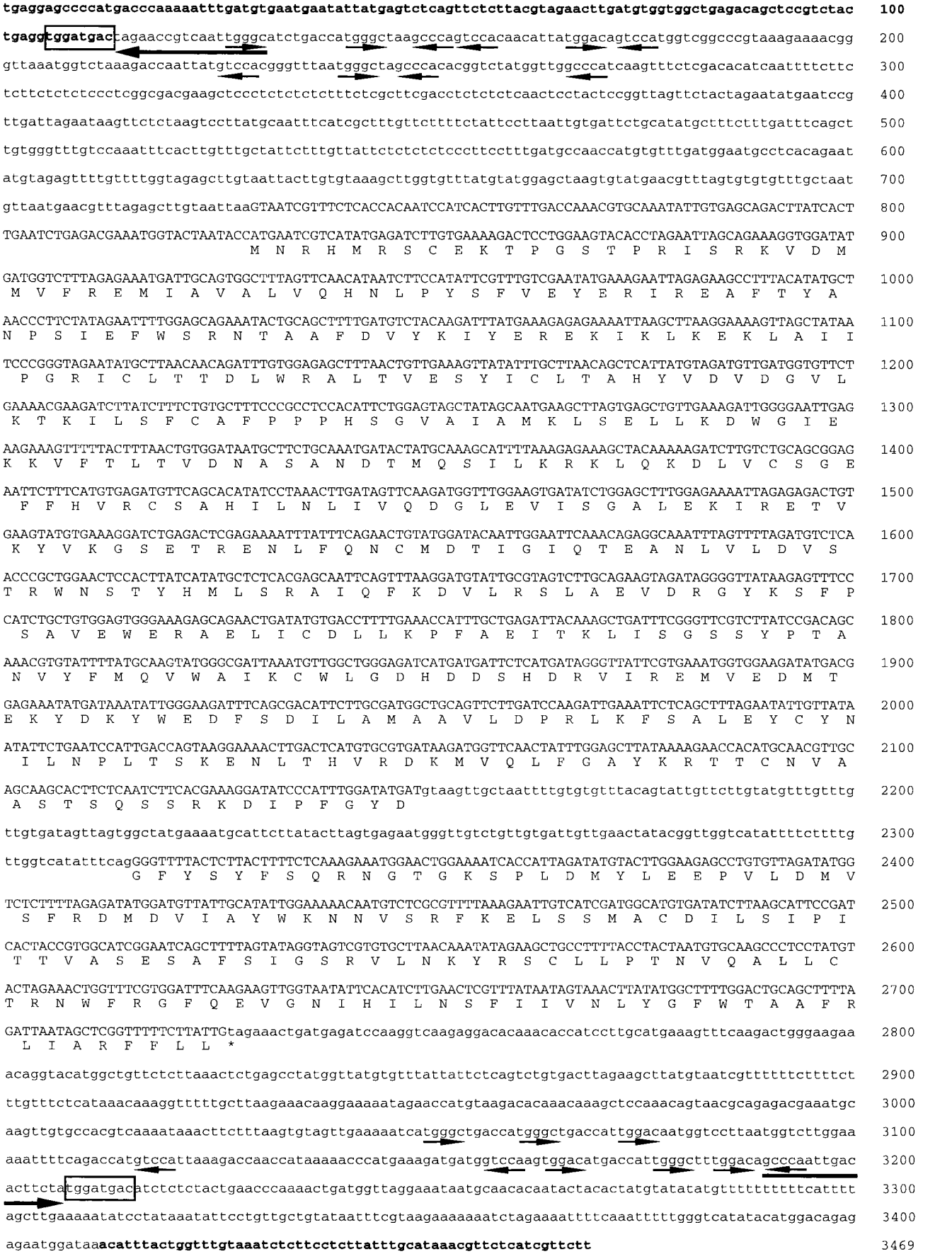
***Tag2* is ubiquitous among Arabidopsis ecotypes:** To assess the copy number and distribution of *Tag2* among Arabidopsis ecotypes we selected a set of six ecotypes for analysis by Southern hybridization. These six ecotypes were chosen because they are commonly used by the research community (Col-0, Ler-0, and Ws-0), or lack *RPS5* function (Bch-1, Chi-0, Ler-0, and Mt-0; see below). Genomic DNA was restricted with *Hin*dIII, blotted, and hybridized with a 143-bp probe from one of the ends of the *Tag2* element (see materials and methods). This probe was used because it spans most of the putative transposase binding sites that should be present even in defective (internally deleted) *Tag2* elements. In addition, because of its small size, this probe should rarely detect more than one band per element. Figure 4 shows that *Tag2*-homologous sequences are present in all six ecotypes, with copy numbers ranging from ~4 in Mt-0 to ~11 in Col-0 and Chi-0. Many of the bands were faint, suggesting some sequence divergence among these copies. Hybridization with a probe derived from the Tag2 ORF revealed a similar number and pattern of bands (data not shown).

Searches of the GenBank DNA sequence database confirmed that Col-0 contains multiple *Tag2*-homologous sequences. Table 1 lists the sequences that have >80% identity with *Tag2* over a minimum of 100 bp. Two of these sequences may represent nonoverlapping portions of the same element as the bacterial artificial chromosome clones from which they were derived map to the same region of chromosome I. The remaining five sequences represent independent insertions on chromosomes I, II, III, and V. The two elements on chromosome I, located at positions 6 and 39 on the Lister-Dean recombinant inbred map (http://genome-www.stanford.edu/Arabidopsis/ww/Feb98RImaps/index.html/), flank the *RPS5* gene, which is located at position 10 (Simonich and Innes 1995).

Alignment of the T10F20 BAC end sequence (accession no. B29529) with *Tag2* revealed an ~900-bp deletion in the transposase coding region of the T10F20 element (data not shown), indicating that this element is defective. Similarly, the elements on chromosomes III and V have suffered internal deletions ranging from 150 to 650 base pairs, and the element on chromosome II appears to represent only a fragment from one end of *Tag2* (data not shown).

The cutoff of 80% identity used for Table 1 was arbitrary; there were many additional sequences that fell below this threshold, but had significant similarity to *Tag2* (probability values lower than 10^{-6} as determined by the BLAST2 algorithm; Altschul *et al.* 1997). These latter sequences likely represent additional members of the hAT transposable element family distinct from *Tag2*, as the homology was confined to the transposase coding region (data not shown).

Insertion of *Tag2* adjacent to *RFL1* occurred after deletion of *RPS5*: The location of *Tag2* adjacent to one end of the deletion breakpoint in Ler-0 (Figure 1) suggested that *Tag2* might be causally related to the deletion of *RPS5*. To investigate this possibility further we identified two additional Arabidopsis ecotypes that lacked the *RPS5* gene, Bch-1 and Mt-0 (see materials and methods). The region containing the *RPS5* dele-


tgaggagccccatgacccaaaaatttgatgtgaatgaatattatgagtctcagttctcttacgtagaacttgatgtggtggctgagacagctccgtctac 100
tgaggtggatgactagaaccgtcaattgggcatctgaccatgggctaagcccagtccacaacattatggacagtccatggtcggcccgtaaagaaaacgg 200
gttaaatggtctaaagaccaattatgtccacgggtttaatgggctagcccacacggtctatggttggcccatcaagtttctcgacacatcaatttttcttc 300
tcttctctctccctcggcgacgaagctccctctctctctttctcgcttcgacctctctctcaactcctactccggttagttctactagaatatgaatccg 400
ttgattagaataagttctctaagtccttatgcaatttcatcgctttgttcttttctattccttaattgtgattctgcatatgctttctttgatttcagct 500
tgtgggtttgtccaaatttcacttgtttgctattctttgttattctctctctcccttccttgatgccaaccatgtgtttgatggaatgcctcacagaat 600
atgtagagttttgttttggtagagcttgtaattacttgtgtaaagcttggtgtttatgtatggagctaagtgtatgaacgtttagtgtgtgtttgctaat 700
gttaatgaacgtttagagcttgtaattaaGTAATCGTTTCTCACCACAATCCATCACTTGTTTGACCAAACGTGCAAATATTGTGAGCAGACTTATCACT 800
TGAATCTGAGACGAAATGGTACTAATACCATGAATCGTCATATGAGATCTTGTGAAAAGACTCCTGGAAGTACACCTAGAATTAGCAGAAAGGTGGATAT 900
M N R H M R S C E K T P G S T P R I S R K V D M
GATGGTCTTTAGAGAAATGATTGCAGTGGCTTTAGTTCAACATAATCTTCCATATTCGTTTGTCGAATATGAAAGAATTAGAGAAGCCTTTACATATGCT 1000
M V F R E M I A V A L V Q H N L P Y S F V E Y E R I R E A F T Y A
AACCCTTCTATAGAATTTTGGAGCAGAAATACTGCAGCTTTTGATGTCTACAAGATTTATGAAAGAGAGAAAATTAAGCTTAAGGAAAAGTTAGCTATAA 1100
N P S I E F W S R N T A A F D V Y K I Y E R E K I K L K E K L A I I
TCCCGGGTAGAATATGCTTAACAACAGATTTGTGGAGAGCTTTAACTGTTGAAAGTTATATTTGCTTAACAGCTCATTATGTAGATGTTGATGGTGTTCT 1200
P G R I C L T T D L W R A L T V E S Y I C L T A H Y V D V D G V L
GAAAACGAAGATCTTATCTTTCTGTGCTTTCCCGCCTCCACATTCTGGAGTAGCTATAGCAATGAAGCTTAGTGAGCTGTTGAAAGATTGGGGAATTGAG 1300
K T K I L S F C A F P P P H S G V A I A M K L S E L L K D W G I E
AAGAAAGTTTTTACTTTAACTGTGGATAATGCTTCTGCAAATGATACTATGCAAAGCATTTTAAAGAGAAAGCTACAAAAAGATCTTGTCTGCAGCGGAG 1400
K K V F T L T V D N A S A N D T M Q S I L K R K L Q K D L V C S G E
AATTCTTTCATGTGAGATGTTCAGCACATATCCTAAACTTGATAGTTCAAGATGGTTTGGAAGTGATATCTGGAGCTTTGGAGAAAATTAGAGAGACTGT 1500
F F H V R C S A H I L N L I V Q D G L E V I S G A L E K I R E T V
GAAGTATGTGAAAGGATCTGAGACTCGAGAAAATTTATTTCAGAACTGTATGGATACAATTGGAATTCAAACAGAGGCAAATTTAGTTTTAGATGTCTCA 1600
K Y V K G S E T R E N L F Q N C M D T I G I Q T E A N L V L D V S
ACCCGCTGGAACTCCACTTATCATATGCTCTCACGAGCAATTCAGTTTAAGGATGTATTGCGTAGTCTTGCAGAAGTAGATAGGGGTTATAAGAGTTTCC 1700
T R W N S T Y H M L S R A I Q F K D V L R S L A E V D R G Y K S F P
CATCTGCTGTGGAGTGGGAAAGAGCAGAACTGATATGTGACCTTTTGAAACCATTTGCTGAGATTACAAAGCTGATTTCGGGTTCGTCTTATCCGACAGC 1800
S A V E W E R A E L I C D L L K P F A E I T K L I S G S S Y P T A
AAACGTGTATTTTATGCAAGTATGGGCGATTAAATGTTGGCTGGGAGATCATGATGATTCTCATGATAGGGTTATTCGTGAAATGGTGGAAGATATGACG 1900
N V Y F M Q V W A I K C W L G D H D D S H D R V I R E M V E D M T
GAGAAATATGATAAATATTGGGAAGATTTCAGCGACATTCTTGCGATGGCTGCAGTTCTTGATCCAAGATTGAAATTCTCAGCTTTAGAATATTGTTATA 2000
E K Y D K Y W E D F S D I L A M A A V L D P R L K F S A L E Y C Y N
ATATTCTGAATCCATTGACCAGTAAGGAAAACTTGACTCATGTGCGTGATAAGATGGTTCAACTATTTGGAGCTTATAAAAGAACCACATGCAACGTTGC 2100
I L N P L T S K E N L T H V R D K M V Q L F G A Y K R T T C N V A
AGCAAGCACTTCTCAATCTTCACGAAAGGATATCCCATTTGGATATGATgtaagttgctaattttgtgtgtttacagtattgttcttgtatgtttgtttg 2200
A S T S Q S S R K D I P F G Y D
ttgtgatagttagtggctatgaaaatgcattcttatacttagtgagaatgggttgtctgttgtgattgttgaactatacggttggtcatatttttcttttg 2300
ttggtcatatttcagGGGTTTTACTCTTACTTTTCTCAAAGAAATGGAACTGGAAAATCACCATTAGATATGTACTTGGAAGAGCCTGTGTTAGATATGG 2400
G F Y S Y F S Q R N G T G K S P L D M Y L E E P V L D M V
TCTCTTTTAGAGATATGGATGTTATTGCATATTGGAAAAACAATGTCTCGCGTTTTAAAGAATTGTCATCGATGGCATGTGATATCTTAAGCATTCCGAT 2500
S F R D M D V I A Y W K N N V S R F K E L S S M A C D I L S I P I
CACTACCGTGGCATCGGAATCAGCTTTTAGTATAGGTAGTCGTGTGCTTAACAAATATAGAAGCTGCCTTTTACCTACTAATGTGCAAGCCCTCCTATGT 2600
T T V A S E S A F S I G S R V L N K Y R S C L L P T N V Q A L L C
ACTAGAAACTGGTTTCGTGGATTTCAAGAAGTTGGTAATATTCACATCTTGAACTCGTTTATAATAGTAAACTTATATGGCTTTTGGACTGCAGCTTTTA 2700
T R N W F R G F Q E V G N I H I L N S F I I V N L Y G F W T A A F R
GATTAATAGCTCGGTTTTTCTTATTGtagaaactgatgagatccaaggtcaagaggacacaaacaccatccttgcatgaaagtttcaagactgggaagaa 2800
L I A R F F L L *
acaggtacatggctgttctcttaaactctgagcctatggttatgtgtttattattctcagtctgtgacttagaagcttatgtaatcgttttttcttttct 2900
ttgtttctcataaacaaaggtttttgcttaagaaacaaggaaaaatagaaccatgtaagacacaaacaaagctccaaacagtaacgcagagacgaaatgc 3000
aagttgtgccacgtcaaaataaacttctttaagtgtagttgaaaaatcatgggctgaccatgggctgaccattggacaatggtccttaatggtcttggaa 3100
aaattttcagaccatgtccattaaagaccaaccataaaaacccatgaaagatgatggtccaagtggacatgaccattgggctttggacagcccaattgac 3200
acttctatggatgacatctctctactgaacccaaaactgatggttaggaaataatgcaacacaatactacactatgtatatatgttttttttttcatttt 3300
agcttgaaaaatatcctataaatattcctgttgctgtataatttcgtaagaaaaaaatctagaaaattttcaaatttttgggtcatatacatggacagag 3400
agaatggataaacatttactggtttgtaaatctcttcctcttatttgcataaacgttctcatcgttctt 3469

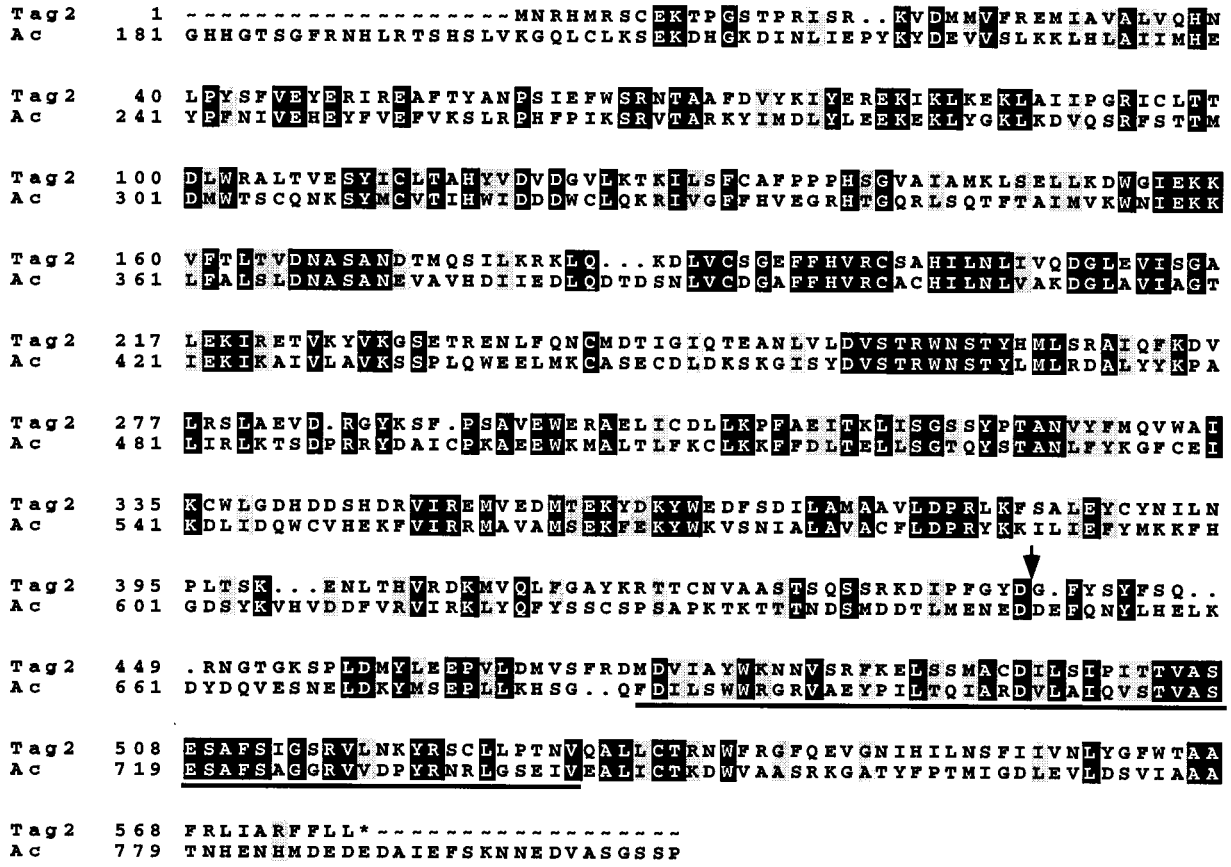


Figure 3.—Alignment of the putative Tag2 transposase with Ac. The position of the single intron is indicated by an arrowhead. Solid boxes indicate identical residues and stippled boxes indicate conservative substitutions. The highly conserved region found in the C-terminal region of all hAT transposases is underlined (Calvi *et al.* 1991; Liu and Crawford 1998). Overall, the two sequences are 33% identical.

tion was amplified from these two ecotypes using primers that flanked the deletion (and *Tag2*) in Ler-0. Both ecotypes yielded PCR products that were ~3 kb smaller than those obtained from Ler-0, indicating that the structure of the deletion area differed. The DNA sequences of the Bch-1 and Mt-0 products were identical to each other (Figure 5) and nearly identical to the sequence present in Ler-0, except that *Tag2* was absent (Figure 1). This sequence identity indicates that the loss of *RPS5* in Ler-0, Bch-1, and Mt-0 represents a single deletion event.

The absence of *Tag2* in the *RPS5* deletion region of Bch-1 and Mt-0 raised the question of whether *Tag2* had excised in these two ecotypes, or had not been present in the first place. We looked for a footprint (*e.g.*, a remnant of the target site duplication) at the position corresponding to the *Tag2* insertion in Ler-0, which would be indicative of an excision event. No such footprint was present (compare Figures 2 and 5). Less than 3% of *Ac* excision events in maize and Arabidopsis remove the target site duplication precisely (Rinehart *et al.* 1997), thus the Mt-0/Bch-1 DNA sequence suggests that *Tag2* inserted after the *RPS5* deletion event in Ler-0.

The *RPS5* deletion event displays hallmarks of nonhomologous DNA end joining: The Mt-0/Bch-1 DNA sequence contained a 192-bp insertion of DNA that was not present in ecotype Col-0 (Figure 5). This insertion makes it highly unlikely that the *RPS5* deletion event was the result of homologous recombination between DNA sequences that flanked *RPS5*. An alternative mechanism for deletion formation is nonhomologous end joining (Gorbunova and Levy 1997). In this mechanism, a double-strand break is followed by exonuclease

Figure 2.—The *Tag2* transposon insertion in Arabidopsis ecotype Ler-0. The terminal inverted repeats are indicated by heavy arrows, and the 8-bp target site duplication is boxed. Subterminal TGGGC and TGGAC repeats are underlined with thin arrows. Uppercase letters indicate two ORFs separated by a single intron. Flanking DNA sequence present in ecotype Col-0 is indicated in bold. The tga codon at position 1 corresponds to the stop codon of *RFL1*. This sequence has been deposited as GenBank accession number AF120335.

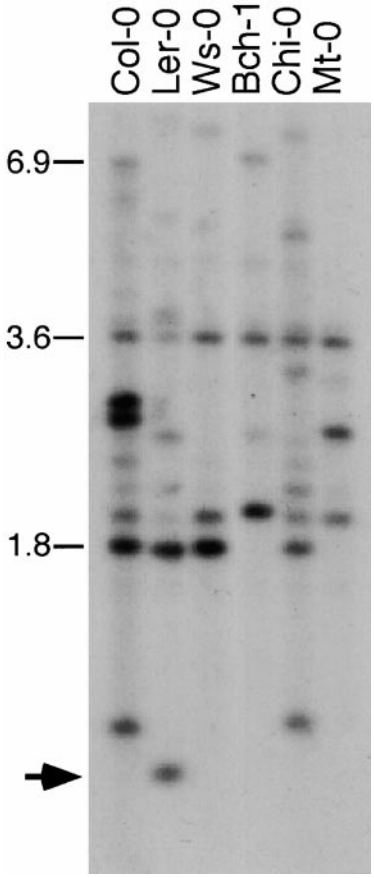


Figure 4.—Presence of *Tag2*-homologous sequences in diverse Arabidopsis ecotypes. Arabidopsis genomic DNA was restricted with *Hin*dIII and analyzed by Southern blot hybridization. The probe used was a 143-bp fragment from the left end of *Tag2* (nucleotides 176–318 in Figure 2). The approximate sizes of common bands are given in kilobase pairs. The band in ecotype Ler-0 expected to hybridize to the probe is indicated by the arrow. The additional bands represent other insertions of *Tag2*-homologous elements.

digestion of the ends. The free ends are subsequently rejoined by a mechanism that is unclear. In plants, the latter step must involve transient invasion of nonhomologous duplex DNA and DNA synthesis, because the products often display large (>20 bp) insertions of "filler DNA," which is frequently derived from DNA sequences physically linked (within a few kilobases) to the deletion point (Wessler *et al.* 1990; Doseff *et al.* 1991; Gorbunova and Levy 1997). The majority of double-strand breaks in plant cells appear to be repaired via such nonhomologous end joining [Gorbunova and Levy (1997) and references therein].

To ascertain the possible origins of the filler DNA present in Bch-1, Ler-0, and Mt-0, we searched for similarities to *RPS5* and *RFL1*. This analysis revealed the presence of a perfect 17-bp inverted repeat (Figure 5). One copy of this repeat is present at the 3′ end of *RFL1* in Col-0, but is disrupted by an insertion of 16 bp relative to the Bch-1, Ler-0, and Mt-0 sequences. It is plausible

TABLE 1

***Tag2*-homologous sequences in Arabidopsis ecotype Col-0**

Accession no./clone	Sequence type	Genetic map position[a]
B08666/T28I21	BAC end	I-6
B08547/T12L3	BAC end	I-6
B29529/T10F20	BAC end	I-39
AC004747/T19L18	Whole BAC	II-49
X98130	81-kb λ contig	III-40[b]
AB018121/MXE2	Whole P1	III-47
AB016892/MXF12	Whole P1	V-82

Sequences listed contain at least one region longer than 100 bp that is >80% identical to *Tag2*.

[a] Approximate position on the Arabidopsis genetic map was determined by locating given BAC clones on contigs containing genetically mapped markers. The assignment of BAC clones to contigs was taken from physical map data available from the *Arabidopsis thaliana* Genome Center (http://genome.bio.upenn.edu/ATGCUP.html/). Roman numerals indicate chromosome number and Arabic numerals indicate the position in centimorgans on the February, 1998 Lister/Dean recombinant inbred map (http://genome-www.stanford/edu/Arabidopsis/ww/Feb98RImaps/index.html/).

[b] Contig contains GAP-A marker (Quigley *et al.* 1996).

that this 16-bp gap represents the true endpoint of the original deletion. The additional 94 bp beyond this point that are present in all four ecotypes would then represent filler DNA that was copied from an intact copy on a sister chromatid or homologous chromosome. Regardless, the presence of the 17-bp repeat suggests duplication of this region during the repair process. We also identified a 30-bp region that is 83% identical to a region within the *RPS5* coding region (Figure 5),

```
1    TGAGGAGCCCCATGACCCAAAAA.....................TtTGATGTGAA
35   tGAATATTATGAGTCTCAGTTCTCTTAcGTAGAacTTGATGTGGTGGcTG
85   AGACAGCTCCGTCTACTGAGGTgGATGACATCTctctactgaacccaaaa
135  ctgatggttaggaaataatgcaacacaatactacactatgtatatatgtt
185  ttttttttcattttagcttgaaaaaaatcctacaaatattcctgttgctg
235  tataatttcgtaagaaaaaaatctagaaaatttttcaaatttttgggtcat
285  atacatggacagagagaatggataaACATTTaCtGGTtTGTAAATCTCTT
```

Figure 5.—DNA sequence from the *RPS5* deletion region of ecotypes Bch-1 and Mt-0. Capital letters indicate DNA sequences shared with ecotype Col-0 that define the endpoints of the deletion. Lower case letters indicate sequences not found in this region of Col-0. The deleted region relative to Col-0 starts 111 bp 3′ of the *RFL1* stop codon (solid underline) and ends 45 bp 3′ of the *RPS5* stop codon (see Figures 1 and 2), spanning a total of 3990 bp. The dots indicate a 16-base deletion in the Bch-1/Mt-0 sequence relative to Col-0. Arrows indicate an inverted repeat. The dashed line indicates a region that is 83% identical (25 of 30 bp) to part of the *RPS5* coding region. The 8-bp sequence duplicated upon insertion of *Tag2* (Figure 2) is boxed.

suggesting that *RPS5* may have also served as template during the repair event.

DISCUSSION

We initiated this study to gain insight into the evolution of the *RPS5* disease resistance gene cluster. The discovery of a transposable element at the *RPS5* deletion site in ecotype Ler-0 made us critically examine the potential role of this transposon in the deletion event. Analysis of other *RPS5*-minus ecotypes suggested that the *Tag2* transposon inserted after the initial deletion of *RPS5* and may not be causally related to the deletion. However, this analysis also indicated that *RPS5* was likely deleted during the repair of a double-strand break, and transposons are well-known inducers of such breaks (Saedler and Gierl 1996).

It is plausible that *Tag2* originally inserted into *RPS5* and then caused its deletion upon transposition to a nearby linked site. A subsequent transposition then brought it back to the site of the deletion. Although this model is highly speculative, it is consistent with the documented behavior of the *Ac* element in maize (Moreno *et al.* 1992). In this example, an insertion of *Ac* was isolated in the *P* gene, and several unstable revertants were selected in which the *Ac* had transposed only a few kilobases away. These revertant lines gave rise to new insertions into the *P* gene at a high frequency. This model is also consistent with the map positions of *Tag2* elements in the Col-0 ecotype; the B08666 element (Table 1) is located only 4 cM from *RPS5*.

A second possibility is that *RPS5* was deleted during the insertion of *Tag2*, and then *Tag2* subsequently excised precisely in ecotypes Mt-0 and Bch-1, removing the target site duplication. Although such excision is rare for the *Ac* element of maize, it does occur in ~3% of excision events (Rinehart *et al.* 1997). Furthermore, other plant transposons, such as *Mu1*, appear to have a much higher frequency of precise excision (Doseff *et al.* 1991). The location of *Tag2* immediately adjacent to one of the deletion breakpoints in the Ler-0 ecotype is consistent with this hypothesis (Figure 5).

Regardless of the original cause of the double-strand break, the subsequent repair event appears to have been accomplished by nonhomologous DNA end joining, rather than homologous recombination. Nonhomologous end joining often results in the insertion of short patches of DNA from adjacent regions, which presumably serve as templates during the repair process (Wessler *et al.* 1990; Doseff *et al.* 1991; Gorbunova and Levy 1997). The junctions between such filler DNA and the breakpoints usually contain very short stretches of homology (2–11 bp), suggesting that the repair process may be similar to gene-conversion events in which the broken end must first base pair with the ectopic template before DNA synthesis proceeds. Because nonhomologous end joining is the predominant mechanism by which double-strand breaks are repaired in plants (Gorbunova and Levy 1997), it may represent an important generator of sequence polymorphism.

Deletion of R genes via a nonhomologous end-joining mechanism may be common. The *RPM1* gene present in Arabidopsis and Brassica has been lost on several independent occasions, and the breakpoints associated with each deletion event contain filler DNA indicative of nonhomologous end joining (Grant *et al.* 1998). Interestingly, like the *RPS5* deletion event, the *RPM1* deletion event in Arabidopsis was confined to just the R gene and did not extend into the flanking genes (Grant *et al.* 1998).

In contrast to *RPS5* and *RPM1*, the *RPP8* gene of Arabidopsis present in ecotype Ler-0 appears to have been eliminated from ecotype Col-0 by an unequal crossover event (McDowell *et al.* 1998). This recombination event may have been promoted by the high similarity of the two participating R genes (>90% identical). The similarity between *RPS5* and *RFL1* is only 74%, which makes homologous recombination less likely. *RPM1* is not a member of a complex, precluding unequal crossing over as a deletion mechanism (Grant *et al.* 1998).

The above studies of *RPS5*, *RPM1*, and *RPP8* illustrate that deletion of R genes is a frequent occurrence. This observation has given rise to speculation that R genes may have a fitness cost in the absence of selective pathogen pressure (Grant *et al.* 1998). Since activation of R genes triggers multiple physiological changes, including localized cell death, it is tempting to speculate that even low level induction of these responses in the absence of pathogens could be counteradaptive. There may thus be an adaptive limit to the total number of R genes present in a plant genome. This may, in part, explain why Arabidopsis appears to have only 100–200 R genes (Botella *et al.* 1997), even though the likely number of pathogen molecules encountered must be much greater.

Recent molecular analyses of R gene clusters have revealed a patchwork pattern of similarity between individual members of a cluster (Parniske *et al.* 1997; McDowell *et al.* 1998), indicative of numerous intracluster recombination events. Although unequal crossing over during meiosis could be one cause of such recombination events, premeiotic repair of double-strand breaks should also be considered. Double-strand breaks could induce unequal crossover events, gene conversion events, and nonhomologous end-joining events. The latter in particular may explain the large number of small deletions and insertions that are observed when comparing members of an R gene complex (McDowell *et al.* 1998) or R gene alleles (Ellis *et al.* 1997) and could also contribute to the patchwork pattern of homologies observed.

Although there are likely many causes of double-strand breaks in plants, transposons are clearly one cause. Somatic excision of the *Mu1* transposable ele-

ment in maize appears to induce deletions at a high frequency, and the breakpoints often contain filler DNA derived from flanking genes, consistent with nonhomologous end-joining repair. Recent work on the *L6* disease resistance gene of flax has documented several small insertions and deletions caused by the native flax element *d Lute* (Luck *et al.* 1998). These examples point out the potential of transposons to contribute significantly to the evolution of disease resistance genes.

The impact of *Tag2* on the Arabidopsis genome in general, and R gene clusters in particular, depends on its past and present activity. The diversity in number and size of restriction fragments detected by the *Tag2* endprobe (Figure 4) indicates that *Tag2* has been active during the diversification of Arabidopsis ecotypes. In addition, we recovered a *Tag2* cDNA by RT-PCR, which suggests that the Tag2 transposase is being expressed and that *Tag2* may therefore be an active transposable element. Quantitation of the germinal transposition activity of *Tag2*, however, awaits further analysis. The relative contributions of transposon-induced changes *vs.* other types of DNA repair and recombination are not yet known. The discovery of *Tag2* will facilitate further investigations into transposon activity within Arabidopsis.

This work was funded by grant R01 GM-46451 from the Institute of General Medical Sciences of the National Institutes of Health to R.W.I.

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Ashfield, T., J. R. Danzer, D. Held, K. Clayton, P. Keim *et al.*, 1998 *Rpg1*, a soybean gene effective against races of bacterial blight, maps to a cluster of previously identified disease resistance genes. Theor. Appl. Genet. **96:** 1013–1021.

Botella, M., M. Coleman, D. Hughes, M. Nishimura, J. Jones *et al.*, 1997 Map positions of 47 Arabidopsis sequences with sequence similarity to disease resistance genes. Plant J. **12:** 1197–1211.

Calvi, B. R., T. J. Hong, S. D. Findley and W. M. Gelbart, 1991 Evidence for a common evolutionary origin of inverted repeat transposons in Drosophila and plants. Cell **66:** 465–471.

Dixon, M. S., D. A. Jones, J. S. Keddie, C. M. Thomas, K. Hanson *et al.*, 1996 The tomato *Cf-2* disease resistance locus comprises two functional genes encoding leucine rich repeats. Cell **84:** 451–459.

Doseff, A., R. Martienssen and V. Sundaresan, 1991 Somatic excision of the *Mu1* transposable element of maize. Nucleic Acids Res. **19:** 579–584.

Ellis, J., G. Lawrence, M. Ayliffe, P. Anderson, N. Collins *et al.*, 1997 Advances in the molecular genetic analysis of the flax-flax rust interaction. Annu. Rev. Phytopathol. **35:** 271–291.

Feinberg, A. P., and B. Vogelstein, 1983 A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

Gorbunova, V., and A. A. Levy, 1997 Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res. **25:** 4650–4657.

Grant, M. R., J. M. McDowell, A. G. Sharpe, M. T. Zabala, D. J. Lydiate *et al.*, 1998 Independent deletions of a pathogen resistance gene in Brassica and Arabidopsis. Proc. Natl. Acad. Sci. USA **95:** 15843–15848.

Hammond-Kosack, K. E., and J. D. G. Jones, 1997 Plant disease resistance genes. Annu. Rev. Plant Physiol. Plant Mol. Biol. **48:** 575–607.

Hebsgaard, S. M., P. G. Korning, N. Tolstrup, J. Engelbrecht, P. Rouze *et al.*, 1996 Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. Nucleic Acids Res. **24:** 3439–3452.

Hulbert, S. H., 1997 Structure and evolution of the *rp1* complex conferring rust resistance in maize. Annu. Rev. Phytopathol. **35:** 293–310.

Innes, R. W., S. R. Bisgrove, N. M. Smith, A. F. Bent, B. J. Staskawicz *et al.*, 1993 Identification of a disease resistance locus in *Arabidopsis* that is functionally homologous to the *Rpg1* locus of soybean. Plant J. **4:** 813–820.

Jia, Y., Y.-T. Loh, J. Zhou and G. B. Martin, 1997 Alleles of *Pto* and *Fen* occur in bacterial speck-susceptible and fenthion-insensitive tomato cultivars and encode active protein kinases. Plant Cell **9:** 61–73.

Jones, D. A., C. M. Thomas, K. E. Hammond-Kosack, P. J. Balint-Kurti and J. D. G. Jones, 1994 Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science **266:** 789–792.

Kawasaki, E. S., and A. M. Wang, 1989 Detection of gene expression, pp. 89–97 in *Detection of Gene Expression*, edited by H. A. Erlich. Stockton Press, New York.

Liu, D., and N. M. Crawford, 1998 Characterization of the putative transposase mRNA of *Tag1*, which is ubiquitously expressed in Arabidopsis and can be induced by Agrobacterium-mediated transformation with *dTag1* DNA. Genetics **149:** 693–701.

Luck, J. E., G. J. Lawrence, E. J. Finnegan, D. A. Jones and J. G. Ellis, 1998 A flax transposon identified in two spontaneous mutant alleles of the *L6* rust resistance gene. Plant J. **16:** 365–370.

Martin, G. B., S. H. Brommonschenkel, J. Chunwongse, A. Frary, M. W. Ganal *et al.*, 1993 Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science **262:** 1432–1436.

McDowell, J. M., M. Dhandaydham, T. A. Long, M. G. M. Aarts, S. Goff *et al.*, 1998 Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of Arabidopsis. Plant Cell **10:** 1861–1874.

Moreno, M. A., J. Chen, I. Greenblatt and S. L. Dellaporta, 1992 Reconstitutional mutagenesis of the maize *P* gene by short-range *Ac* transpositions. Genetics **131:** 939–956.

Parker, J. P., M. J. Coleman, V. Szabo, L. N. Frost, R. Schmidt *et al.*, 1997 The Arabidopsis downy mildew resistance gene *RPP5* shares similarity to the Toll and interleukin-1 receptors with *N* and *L6*. Plant Cell **9:** 879–894.

Parniske, M., K. E. Hammond-Kosack, C. Golstein, C. M. Thomas, D. A. Jones *et al.*, 1997 Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. Cell **91:** 821–832.

Pryor, T., and J. Ellis, 1993 The genetic complexity of fungal resistance genes in plants. Adv. Plant Pathol. **10:** 281–305.

Quigley, F., P. Dao, A. Cottet and R. Mache, 1996 Sequence analysis of an 81 kb contig from *Arabidopsis thaliana* chromosome III. Nucleic Acids Res. **24:** 4313–4318.

Richter, T. E., T. J. Pryor, J. L. Bennetzen and S. H. Hulbert, 1995 New rust resistance specificities associated with recombination in the *Rp1 complex* in maize. Genetics **141:** 373–381.

Rinehart, T. A., C. Dean and C. F. Weil, 1997 Comparative analysis of non-random DNA repair following *Ac* transposon excision in maize and *Arabidopsis*. Plant J. **12:** 1419–1427.

Saedler, H., and A. Gierl, 1996 *Transposable Elements. Current Topics in Microbiology and Immunology.* Springer-Verlag, Berlin.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Simonich, M. T., and R. W. Innes, 1995 A disease resistance gene in *Arabidopsis* with specificity for the *avrPph3* gene of *Pseudomonas syringae* pv. *phaseolicola*. Mol. Plant-Microbe Interact. **8:** 637–640.

Song, W.-Y., L.-Y. Pi, G.-L. Wang, J. Gardner, T. Holsten *et al.*, 1997 Evolution of the rice Xa21 disease resistance gene family. Plant Cell **9:** 1279–1287.

Thomas, C. M., D. A. Jones, M. Parniske, K. Harrison, P. J. Balint-Kurti *et al.*, 1997 Characterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that

determine recognitional specificity in Cf-4 and Cf-9. Plant Cell **9:** 2209–2224.

Tsay, Y.-F., M. J. Frank, T. Page, C. Dean and N. M. Crawford, 1993 Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. Science **260:** 342–344.

Warren, R. F., A. Henk, P. Mowery, E. Holub and R. W. Innes, 1998 A mutation within the leucine rich repeat domain of the Arabidopsis disease resistance gene *RPS5* partially suppresses multiple bacterial and downy mildew resistance genes. Plant Cell **10:** 1439–1452.

Wessler, S., A. Tarpley, M. Purugganan, M. Spell and R. Okagaki, 1990 Filler DNA is associated with spontaneous deletions in maize. Proc. Natl. Acad. Sci. USA **87:** 8731–8735.

Whitham, S., S. P. Dinesh-Kumar, D. Choi, R. Hehl, C. Corr *et al.*, 1994 The product of the tobacco mosaic virus resistance gene *N*: similarity to Toll and the interleukin-1 receptor. Cell **78:** 1101–1115.

Communicating editor: V. Sundaresan