nature genetics

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize

Michele Morgante[1], Stephan Brunner[2], Giorgio Pea[2,3], Kevin Fengler[2], Andrea Zuccolo[1,3] & Antoni Rafalski[2]

We report a whole-genome comparison of gene content in allelic BAC contigs from two maize inbred lines. Genic content polymorphisms involve as many as 10,000 sequences and are mainly generated by DNA insertions. The termini of eight of the nine genic insertions that we analyzed shared the structural hallmarks of helitron rolling-circle transposons[1–3]. DNA segments defined by helitron termini contained multiple gene-derived fragments and had a structure typical of nonautonomous helitron-like transposons. Closely related insertions were found in multiple genomic locations. Some of these produced transcripts containing segments of different genes, supporting the idea that these transposition events have a role in exon shuffling and the evolution of new proteins. We identified putative autonomous helitron elements and found evidence for their transcription. Helitrons in maize seem to continually produce new nonautonomous elements responsible for the duplicative insertion of gene segments into new locations and for the unprecedented genic diversity. The maize genome is in constant flux, as transposable elements continue to change both the genic and nongenic fractions of the genome, profoundly affecting genetic diversity.

The widespread DNA-sequence noncollinearity among maize inbred lines[4,5] has been attributed to recent long-terminal-repeat (LTR) retrotransposon insertions and to differential presence of genes or gene fragments[6]. The molecular mechanism responsible for the differences in genic sequence content has not been elucidated, but it has been suggested[6] that insertions rather than deletions are involved. Among five genomic regions sequenced from inbred lines B73 and Mo17, as well as the bronze1 locus in inbred lines McC and B73 (ref. 4), 38% of the 72 identified genic segments were not shared in total and 34% were not shared between lines B73 and Mo17 (ref. 6). To estimate the frequency of nonshared genes or genic fragments in the maize genome, we took advantage of the availability of BAC contig maps of lines B73 (from the Arizona Genomics Institute) and Mo17 (ref. 7). We hybridized nearly 15,000 oligonucleotide probes selected from different maize genes simultaneously to the two maps[8], allowing us to identify allelic pairs of contigs covering ~67% of the total map size. In each pair of contigs, we identified shared and nonshared genes on the basis of the probe hybridization using a conservative criterion to minimize the effect of false positive and false negative hybridizations. Of the 20,656 qualifying hybridization instances, 79% were shared between the two lines, 11% were found only in line B73 and 9% were found only in line Mo17 (**Table 1**). We concluded that a large fraction (20%) of genome segments hybridizing with gene-derived probes was not shared between inbred lines B73 and Mo17, in agreement with observations from fully sequenced genomic regions[4–6]. Assuming that maize has at least 40,000 functional shared genes (on the basis of previous estimates for rice[9] and maize[10,11]), ~10,000 genes or gene fragments are not shared between the two lines. The fraction of nonshared sequences varies widely among different contigs, even when considering only large contigs (**Fig. 1**). We observed no substantial skewing toward regions with or without a high proportion of nonshared genes, but 8% of the contigs shared all genes.

Analysis of the nonshared genic segments in the five sequenced regions showed that they correspond to fragments of genes (pseudogenes) usually present in clusters. Gene fragments in clusters tend to have the same orientation with respect to the direction of transcription[6]. The fragments seemed to result from insertion events, because, unlike most shared genes, they were consistently not found in the homologous rice genomic region[6]. Clusters of gene fragments that were part of a larger duplicated segment were previously observed in the *Rp1* region[12] and proposed to result from the general instability of the region, perhaps originating through double-strand DNA-break repair. The high frequency of nonshared genic fragments raises the question of the molecular mechanism of insertion. We compared all

Table 1 Estimates of frequency of shared and nonshared genes and gene fragments between inbred lines B73 and Mo17

	Total	Shared	Unique to B73	Unique to Mo17
Absolute number	20,656	16,408	2,353	1,895
Proportion (%)	100.0	79.4	11.4	9.2

[1]Dipartimento di Scienze Agrarie ed Ambientali, Universita' di Udine, Via delle Scienze 208, 33100 Udine, Italy. [2]DuPont Crop Genetics Research, DuPont Experimental Station Building E353, Wilmington, Delaware 19880-353, USA. [3]Present addresses: Università degli Studi di Milano, Dipartimento di Scienze Biomolecolari e Biotecnologie, Via Celoria 26, 20133 Milano, Italy (G.P.); Arizona Genomics Institute, University of Arizona, Tucson, Arizona 85721, USA (A.Z.). Correspondence should be addressed to M.M. (michele.morgante@uniud.it).

Published online 31 July 2005; doi:10.1038/ng1615

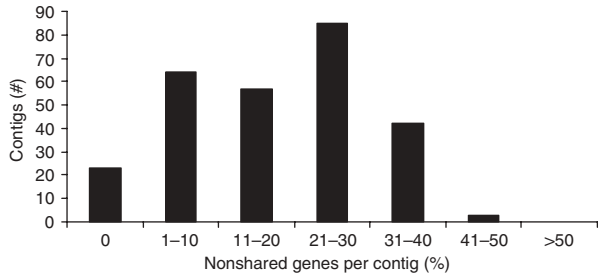

Figure 1 Genome analysis of nonshared genic regions in maize. Distribution of the proportion of nonshared genes and gene fragments between inbred lines B73 and Mo17 in 274 large contigs containing at least 20 shared overgo probes.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

genic insertion termini in the five genomic regions among themselves and with their insertion sites. Eight of nine shared the same features: an A↓T target site, lack of target site duplication, a 5′ TCT end and a 3′ CTAG end preceded by an 18- to 25-bp region capable of forming a hairpin structure (**Fig. 2**). These features could not be recognized at the ends of the insertion involving the RS genes at locus 9009. Unlike the others, this insertion has target site duplication and short inverted repeats and contains a full-length gene.

These features are those of helitrons, a recently described type of DNA transposon[1–3]. Helitrons have been identified computationally in *Caenorhabditis elegans*, *Arabidopsis thaliana* and rice[1] as transposable elements encoding the signature protein (HEL) composed of the rolling-circle replication initiator and DNA helicase domains necessary for transposition. Plant elements also contain a replication protein A (RPA)-like protein with putative single-stranded DNA–binding activity[1]. On the basis of their structural features, a rolling-circle replicative transposition mechanism has been postulated[1]. Unlike most traditional DNA transposons[13], helitrons do not have terminal inverted repeats or form target site duplications. In maize, two nonautonomous helitron insertions carrying different gene fragments are responsible for a loss-of-function mutation in the shrunken2 gene[3] and for the reference barren stalk mutated allele[14], and multiple putative nonautonomous helitron insertions have been reported[14].

The sequences around the bronze1 locus from inbred lines Mo17 and McC indicate that the four genes identified as exclusive to McC originated from two independent insertion events into an Mo17-like haplotype, each flanked by helitron termini (**Fig. 3**) and separated by 893 bp. One insertion carried the G, H and I gene fragments; the other carried the J gene fragment alone. The putative predicted gene I (hypro3)[4] crosses the insertion element boundaries.

Assuming the five sequenced regions are representative of the genome as a whole, our characterization of eight new helitrons carrying multiple gene fragments indicates that nonautonomous helitron elements are widespread in the maize genome and provides strong circumstantial evidence that they are responsible for the frequent genic insertion polymorphisms.

Given the insertion structures, it is evident that the enzymatic activities supporting their replicative transposition must have been provided in *trans*. We searched for putative autonomous helitrons encoding the HEL and RPA proteins in maize genomic sequences and identified 11 sequences with substantial homology to the HEL protein, all but one of which contained all the conserved helicase and replication initiator domains (**Supplementary Table 1** online). Five of these also showed homology to the RPA protein, previously described in plants for putative autonomous elements of the same class[1]. In two of them, the expected helitron 5′ and 3′ termini could be identified (**Supplementary Fig. 1** online). Both elements are interrupted by large insertions and are probably inactive. cDNA sequences with very high similarity to 8 of the 11 putative autonomous elements identified have been found in cDNA collections, further supporting the possibility that active elements exist in the maize genome (**Supplementary Table 1** online).

The nonautonomous helitrons we observed carry fragments from one or several genes. It has been suggested that autonomous helitron elements have captured the gene encoding RPA in plants[1]. Helitron elements have also been predicted to be able to incorporate genome-resident genes into nonautonomous helitron-like structures[15]. The molecular mechanism by which the genes are captured has not yet been explained. Genes in the nonautonomous helitrons are usually incomplete. We examined in detail the exon-intron structure of 18 nonshared gene fragments in five helitrons[6]. We detected single exons for 7 nonshared gene fragments, whereas 11 nonshared gene fragments were represented by two or more exons with conserved intron-exon boundaries (**Supplementary Tables 2** and **3** online).

Under high-stringency conditions, we obtained correctly sized PCR products from genomic DNA of lines B73 and Mo17 for all primer sets targeting individual exons within helitrons (**Table 2** and **Supplementary Fig. 2** online). In contrast, each targeted helitron occurred in

Locus	Line with insertion	Inserted genes	GenBank accession	5′ start site	Insertion sequence termini and genes contained	3′ end site
bz1	McC	GHI	AF391808	TTTAAGTGCA*A* - (139589) -	**TCT**CTACTACTCTATAAGG---**IHG**---TTTTACTCCCGTCGCAACGCACGGGCACTCAC**CTAG** - (133732) -	*T*GGATTTTGAA
bz1	McC	J	AF391808	GACATCTAGA*A* - (140482) -	**TCT**AAAGTATTAAAACTTG----**J**----CCGATAGGGCGCCCGTATGGGCGCCCCATGTT**CTAG** - (143193) -	*T*ATGAGTAATT
9002	B73	GHIJKLM	AY664413	CAAATGGCCT*A* - (180460) -	**TCT**GTATCTATATCTATAT-**GHIKLM**--CATGGAAACCGTAGCAACGCACGGGCATTCAA**CTAG** - (199786) -	*T*TGAGCACCAT
9002	B73	NOPQ	AY664413	TGCTTAATAA*A* - (210885) -	**TCT**CTACTACTCCTTAAGC--**QPON**-- ATAGGTCCCCATTGCAACGCACGGGCACTCAC**CTAG** - (205938) -	*T*AGTACTCTTA
9002	B73	RST	AY664413	AACTACTACA*A* - (240549) -	**TCT**ATACTACCTATTAAGA---**RST**---TTTTACTCCCGTCGCAACGCACGTGCACTCAC**CTAG** - (259755) -	*T*AGAACTATAG
9002	B73	U	AY664413	CTAACCTTAG*A* - (7748) -	**TCC**CTACTACTAATTAAGA----**U**----TTTAACTTACGTAGCGAAGCACGGGCACATAC**CTAG** - (5070) -	*T*AGAAAAATTA
9008	B73	HI	AY664414	TTCATGCTATT - (89533) -	**TCT**CTATTAATCCTTAAGG---**IH**----ACATGTTCCCGTTGCAACGCACGGGCACTCAC**CTAG** - (81613) -	*T*ATGAAAATAT
9009	Mo17	TUVW	AY664419	GTGCCCTCTA*A* - (238234) -	**TCT**CTTCTATTAAAACCTG---**TUVW**--CAGCTGAGCGCCCGATAGGGCGCTTTCCTATT**CTAG** - (240687) -	*T*ATCATCTCCA
9009	Mo17	RS	AY664419	ACCTTAGGAG*A* - (108803) -	CACTACTACAGTTCATGCT----**RS**---AAGATGTATGTCTTAAATAAGCATATTTCTAGTAGT - (113892) -	*GAGA*GATCAGA

Figure 2 Sequence features of gene fragment–containing polymorphic insertions in five genomic regions of maize. Ten bases are shown on either side of the AT target site, which is shown in bold italic letters. The 5′ (19 bases) and 3′ (36 bases) termini of each insertion are also shown. Conserved 5′ and 3′ termini are in bold letters; 3′ terminal hairpins are underlined. Gene fragments contained in each insertion are shown, using letters as in ref. 6. For the RS9009 insertion, only the putative target site duplications are in italic. Seven of the eight insertions that show conserved features occurred in single- or low-copy regions, one (NOPQ9002) in a repetitive sequence represented by a LTR retrotransposon[6]. Insertions occurred in regions of low GC content: the average over the eight loci for the 20 bases that are centered on the AT target site is 32.5%, compared with an average for the genome of 47% (ref. 30).

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

either B73 or Mo17, but not in both, indicating that at least one close homolog of the targeted exon sequences existed elsewhere in the genome of the inbred line that lacked the helitron insertion. We analyzed oat-maize chromosome addition lines representing each of the ten maize chromosomes[16,17] with the primer pairs targeting single exons and found that individual exons representative of those found in helitrons were usually present on multiple maize chromosomes (**Table 2**). We obtained PCR products of the expected sizes from both inbred lines with primer pairs targeting clusters of exons identified in helitrons, indicating that clusters, which are unique to either B73 or Mo17 in specific regions, occurred in both inbred lines at different locations (**Table 2** and **Supplementary Fig. 2** online). We screened oat-maize lines with the same primer pairs and found that there were fewer chromosomes hosting such complex sequences than hosting single exons (**Table 2**). The chromosomes that were positive for a specific gene cluster formed a subset of those that were positive for the constituent gene fragments. Therefore, gene clusters are less common in the genome than are single gene fragments. The possibilities that some of the negative PCR results may originate from SNPs in one of the lines analyzed and that the results on location and copy number observed in Seneca60, the maize donor of the addition lines, may differ from those in other maize lines cannot be ruled out, but the overall trend will not be affected.

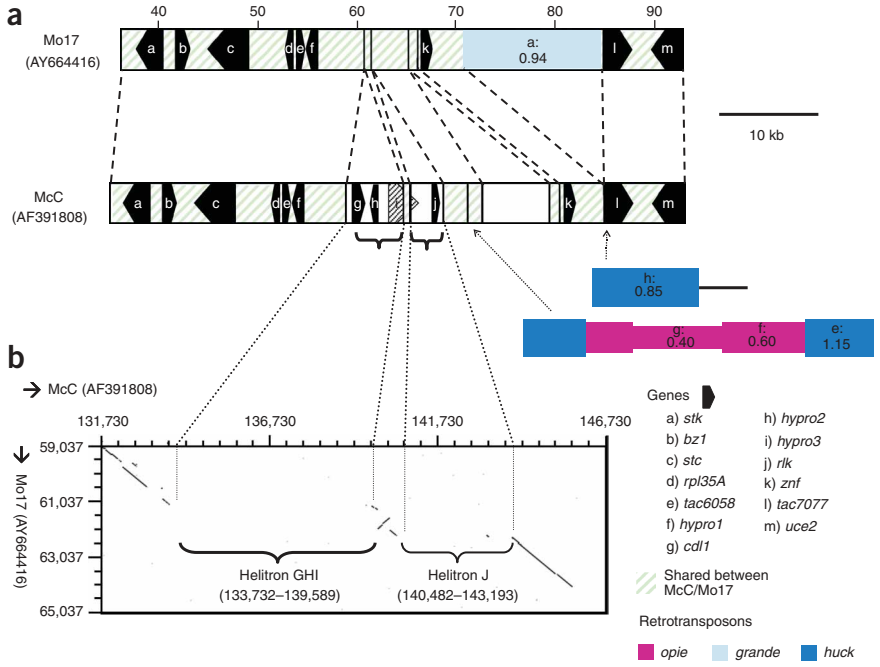

Figure 3 Two helitron-like insertions in the bronze1 region[4,6] are responsible for the observed genic differences. (**a**) Comparison of the Mo17 and the McC sequenced regions. Putative genes are shown as black arrows; LTR retrotransposons, as colored boxes (the estimated time of insertions is shown inside the box). Nonshared single-copy regions are in white; shared ones are dashed. Dashed lines connect corresponding positions in the two sequences. (**b**) Dot-plot analysis of the region surrounding the four identified nonshared gene fragments. The position of the two helitron-like elements in the McC sequence is shown.

To identify the intact genes from which the gene fragments present in helitron 9002NOPQ were derived, we first identified maize cDNAs highly homologous to the four genic fragments. Amplifications in maize-oat addition lines using PCR primer sets at the 5′ and 3′ ends of each cDNA indicated that each of the four intact genes was located on a single chromosome. The intact genes were also amplified from both B73 and Mo17 inbred lines (data not shown).

These data strongly support a mechanism by which the helitron element acquires fragments of multiple genes from different genomic locations, leaving the original copies intact, and produces gene fragment duplications by inserting in a new genomic location. Presence and distribution of single exons and clusters of exons are consistent with gene fragment cluster–carrying nonautonomous helitrons further inserting into multiple locations in the genome using a rolling-circle replicative transposition mechanism. The finding of single exons in more locations than the corresponding exon clusters supports the conclusion that the same genic fragment may occur in combination with different gene fragments in multiple nonautonomous helitrons.

RT-PCR analysis of mRNAs from inbred lines B73 and Mo17 and their reciprocal hybrids showed expression of individual gene fragments present in nonshared helitrons (or of their homologs elsewhere in the genome) in both inbred lines (**Table 3** and **Supplementary Fig. 3** online). The smaller size of some RT-PCR amplicons, relative to the genomic PCR products, is consistent with mRNA splicing. RT-PCR analysis across neighboring gene fragments in helitrons was positive in all four helitrons, indicating that clusters of exons derived from different genes might produce a single transcript. Some clusters are expressed in both inbred lines, and others are expressed in only one of two inbred lines, even though genomic PCR indicates the presence of homologous clusters in different genomic locations in both inbred lines. We conclude that some, but not all, of the nonautonomous helitrons produce transcripts extending across exons derived from different genes.

We have not been able to determine when the insertions of the helitron elements occurred. The fact that specific helitron elements are present in some but not other inbred lines indicates that their transposition occurred relatively recently. Two of the eight identified helitron elements (GHIJKLM9002 and RST9002) are interrupted by other transposons, a LTR retroelement and a DNA transposon of the En/Spm class, that carry target site duplications. Therefore, they probably inserted after the helitrons transposed to the present locations. The retroelement insertion into the helitron has been dated to 0.8 million years ago[6]. One of the helitron elements (NOPQ9002) has inserted into a LTR retrotransposon whose insertion is estimated to be 2.04 million years old[6]. This limited information seems to indicate that gene fragment–carrying helitrons have moved in the maize genome in the same time period as other maize transposons.

A search using both ends (**Fig. 2**) of the eight characterized helitron-like elements as a BLASTN query against maize sequences in GenBank identified four additional putative helitron-like nonautonomous elements (**Supplementary Table 4** online), all containing gene fragments. Their analysis provides additional support to our findings. One of them was previously undetected in the 9009 region sequenced by our group[6] and is present in both B73 and Mo17

Table 2 Summary of PCR amplification results for nonshared genic fragments

	Parental lines		Oat/maize chromosome addition lines											
Gene cluster	B73	Mo17	Oat	Maize	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10
GHIJKLM (B73-9002)														
H9002	+	+	–	+	+	–	–	–	+	–	+	+	–	–
J9002	+	+	–	+	+	+	+	–	–	+	+	–	–	+
K9002	+	+	–	+	+	–	–	–	+	–	–	+	–	+
L9002	+	+	–	+	+	–	–	–	+	+	–	–	–	–
M9002	+	+	–	+	+	–	–	–	–	–	–	–	–	–
K9002 + L9002	+	–	–	+	+	–	–	–	–	–	–	–	–	–
L9002 + M9002	+	+	–	+	+	–	–	–	–	–	–	–	–	–
NOPQ (B73-9002)														
O9002	+	+	–	+	+	–	–	–	+	–	+	+	–	–
P9002	+	+	–	+	+	+	+	+	+	+	–	–	–	+
Q9002	+	+	–	+	+	–	–	–	–	–	–	–	–	–
N9002 + O9002	+	+	–	+	+	–	–	–	–	–	–	–	–	–
O9002 + Q9002	+	+	–	+	+	–	–	–	–	–	–	–	–	–
P9002 + Q9002	+	+	–	+	+	–	–	–	–	–	–	–	–	–
RST (B73-9002)														
S9002	+	+	–	+	+	+	–	–	–	+	–	+	–	–
T9002	+	+	–	+	–	+	+	–	–	+	–	–	–	–
S9002 + T9002	+	+	–	+	–	+	–	–	–	+	–	–	–	–
HI (B73-9008)														
H9008	+	+	–	+	–	+	–	–	–	+	+	+	+	–
I9008	+	+	–	+	+	+	–	–	–	–	–	+	–	–
H9008 + I9008	+	+	–	+	–	+	–	–	–	–	–	+	–	–
RS (Mo17-9009)														
R9009	+	+	–	+	–	+	+	+	–	–	+	–	–	+
S9009	+	+	–	+	–	–	–	–	–	–	–	–	–	–
R9009 + S9009	+	+	–	+	–	–	–	–	–	–	–	–	–	–

+, positive amplification; –, negative amplification. Primer pairs and their locations are listed in **Supplementary Table 2** online. Primer sets across neighboring clustered genes are underlined.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

sequences. The Mo17 sequence is interrupted by a Ji LTR retrotransposon insertion that was dated to 0.03 million years ago[6]. The sequence divergence between the two copies of the helitron element corresponds to a divergence time of only ~0.04 million years. Two other elements are found in regions that were compared with rice and sorghum[18,19] and contain predicted genes (two and one, respectively) not found in the orthologous sequences of either rice or sorghum. One of them contains a predicted gene with similarity to a 40S ribosomal protein S8 (ref. 19). Notably, a shorter genic fragment seemingly derived from the same gene is present in the NOPQ9002 helitron element (N gene fragment), confirming that fragments of the same gene can be incorporated into different nonautonomous helitron elements.

The gene fragment duplications and insertions mediated by helitron elements seem to disrupt the pattern of gene collinearity between maize and related species[18–20]. When these pseudogenes are disregarded, however, the collinearity is better than previously thought[6]. Complex mechanisms have been invoked to explain the presence or absence of specific genes in orthologous regions of cereals[21], including gene movement[21] in conjunction with the whole-genome duplication event that occurred in an ancestor of maize. Helitron-mediated gene duplications suggest an alternative explanation, in addition to the well-documented examples of deletion events involving homeologous loci[19–21].

The involvement of the genic insertions in the modulation of the expression of the ancestral intact genes through transcriptional silencing mechanisms[22,23] could be important for the heterosis phenomenon that does not have yet a satisfactory molecular explanation[24,25]. Owing to the presence of genic fragments rather than entire genes, in the short term it is likely that gene-containing helitron elements are selectively neutral from the point of view of their protein-coding potential. Over evolutionary time, they nevertheless offer opportunities for the appearance of new functions, owing to the formation of clusters of gene fragments that are correctly spliced and transcribed and that, in rare cases, could produce functional fusion proteins. A similar, though less frequent, phenomenon was recently described for Mutator-like elements in maize and rice[26] and was previously observed in *A. thaliana*[27,28].

The evidence provided here indicates that helitron elements cause genic sequence insertions in maize by an unknown molecular mechanism and can create multiple partial copies of genes (pseudogenes). The two helitron insertions present in the bronze1 region were recently described and the isolation of a copy nearly identical to one of the two was reported[29], further supporting the idea that nonautonomous helitrons containing gene fragments behave as transposable elements and are present in multiple copies in the genome. Our conservative estimate of 10,000 nonshared gene fragments in the maize genome, a large fraction of which are likely to have been mobilized by helitrons, reflects only those that are polymorphic between the two inbred lines we studied. Extrapolating the total number of insertions and of gene fragments involved to the existing maize population is impossible. Helitron gene duplication and exon shuffling activity may make the definition of the gene complement of maize very difficult, even though most differences documented to date are pseudogenes by our current

Table 3 Summary of RT-PCR amplification results for nonshared genic fragments

	Seedling				Root				Leaf			
Gene cluster	B73	Mo17	B73/Mo17	Mo17/B73	B73	Mo17	B73/Mo17	Mo17/B73	B73	Mo17	B73/Mo17	Mo17/B73
GHIJKLM (B73-9002)												
H9002	+	+	+	+	+	+	+	+	+	+	+	+
J9002	–	–	–	–	–	–	–	–	–	–	–	–
K9002	–	–	–	–	+	+	+	+	–	–	–	–
L9002	+	+	+	+	+	+	+	+	+	+	+	+
M9002	–	+	+	+	–	+	+	+	–	+	+	+
K9002 + L9002	–	–	–	–	–	–	–	–	–	–	–	–
L9002 + M9002	–	+	+	+	–	+	+	+	–	+	+	+
NOPQ (B73-9002)												
O9002	+	+	+	+	+	+	+	+	+	+	+	+
P9002	+	+	+	+	+	+	+	+	+	+	+	+
Q9002	+	+	+	+	+	+	+	+	+	+	+	+
N9002 + O9002	–	+	+	+	–	+	+	+	–	+	+	+
O9002 + Q9002	–	+	+	+	–	+	+	+	–	+	+	+
P9002 + Q9002	–	+	+	+	–	+	+	+	–	+	+	+
RST (B73-9002)												
S9002	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
T9002	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
S9002 + T9002	–	–	–	–	+	+	+	+	–	–	–	–
HI (B73-9008)												
H9008	+	+	+	+	+	+	+	+	+	+	+	+
I9008	+	+	+	+	+	+	+	+	+	+	+	+
H9008 + I9008	+	–	+	+	+	–	+	+	+	–	+	+
RS (Mo17-9009)												
R9009	+	+	+	+	+	+	+	+	+	+	+	+
S9009	+	+	+	+	+	+	+	+	+	+	+	+
R9009 + S9009	+	+	+	+	+	+	+	+	+	+	+	+

+, positive amplification; –, negative amplification; ND, not determined. Primer pairs and their locations are listed in **Supplementary Table 2** online. Primer sets across neighboring clustered genes are underlined.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

npg

definition. The maize genome is in flux, and its genetic diversity is profoundly affected by the presence of polymorphic insertions due to DNA transposons that have modulated the gene content of a species to an unanticipated extent. It remains to be seen whether helitron-like elements have had a similar role in shaping genome structure and diversity in other plant species and whether other, yet undetected mobile element classes with similar effects can be found.

METHODS

Physical map alignment and analysis of shared and nonshared genes. We identified allelic contigs from the B73 (Arizona Genomics Institute) and Mo17 (ref. 7) maps on the basis of shared overgo probes[8] as described[6]. The BAC clones used in the overgo probe hybridization experiment provided a tenfold genome coverage for B73 relative to an eightfold coverage for Mo17. We considered only overgo probes that were localized to ten contigs or fewer to avoid probes identifying repetitive sequences. We defined allelic contigs as those sharing at least five probes. We identified 695 such pairs of contigs, covering ~63% of the Mo17 physical contig map. We identified shared genes as those probes that were present in both contigs, even if they hybridized to a single BAC clone in each contig, and nonshared genes as those probes that did not hybridize to one of the two contigs and hybridized to at least two BAC clones in the other contig. This is a rather conservative criterion that we adopted to avoid the effect of false positive hybridizations. When comparing allelic BAC contigs, we counted nonshared probes only in fully overlapping parts that were defined as those contained in the outermost shared probes.

Sequence analysis and gene annotation. The search for autonomous helitron elements in maize used protein sequences encoded by helitron elements identified in *A. thaliana* and rice[1] as a query in tBLASTN analyses against the nr and HTS sections of GenBank. We examined alignments to detect conserved residues in the HEL protein-coding gene. We identified 5′ and 3′ ends by comparison of the two identified elements. We searched for cDNA sequences with homology to the rice HEL protein-coding gene in the dbEST section of GenBank using tBLASTN and compared them with the identified putative autonomous elements using BLASTN to find those that showed >90% identity over at least 200 bases. The search for new nonautonomous helitron elements in maize used the 20 most 5′ most and the 37 most 3′ bases of all elements (including the conserved target site on both sides; **Table 1**) interrupted by a stretch of 20 Ns in a BLASTN search of the nr section of GenBank. We considered a hit to be positive when both ends matched the same genomic sequence in the correct orientation and at a distance between 1 and 25 kb. We annotated the exon-intron structure of nonshared genes on the basis of BLASTX alignments to known proteins from GenBank or BLASTN alignment to public and private maize expressed sequence tags. We used FGENSH splicing site predictions in some cases. We treated discontinuous alignments with the same entry in the database as consecutive exons of the same gene. For the nonautonomous elements (**Supplementary Table 4** online), we used only BLASTX alignments to known proteins from GenBank to identify protein-coding regions. We considered only matches with $E < 10^{-10}$ to be significant. We estimated the time of insertion of LTR retrotransposons and the divergence between allelic sequences as described[6].

Plant materials. We germinated seeds of maize inbred lines B73 and Mo17 and of their reciprocal hybrids in well arrays (4 cm in diameter) containing agriperlite in growth chambers (26 °C constant temperature; photoperiod 14 h light–10 h dark; relative humidity 90%). We collected seedling, juvenile root and leaf material from single individuals for RNA extraction. To sample juvenile tissues, we collected the whole whorl above the first leaf collar, with

the exclusion of the second leaf, as seedling samples. We collected radicles, lateral and seminal roots after the appearance of the tip of the fourth leaf as root samples. We transferred additional plants to 20–cm-diameter pots containing common soil and grew them until the appearance of the sixth leaf tip to collect the fifth leaf as leaf samples.

DNA extraction. We extracted genomic maize DNA using the DNeasy Plant Maxi Kit (Qiagen). DNA from oat-maize addition lines[16] was provided by E. Ananiev (Pioneer Hi-Bred Intl.).

PCR primer design. We designed PCR primers in exon sequences using Primer3 software (Whitehead Institute for Biomedical Research). We used default parameters, except for the following ones: primer size 21 bp (minimum 20 bp, maximum 22 bp), GC% 58 (minimum 53%, maximum 80%) and GC Clamp 1. We combined primers to amplify either segments of a single gene or segments of adjacent genes. All PCR primers used are described in **Supplementary Table 5** online.

mRNA purification and reverse transcription. For each maize genotype, we synthesized cDNA separately from two samples of blended tissue material obtained from five individuals. We extracted total RNA using TRIzol Reagent (Invitrogen) in accordance with the manufacturer's protocol. We purified poly(A^+) RNA from 500–1,000 μg of total RNA using the Qiagen Oligotex Midi Kit. We treated a sample of 2 μg of poly(A^+) RNA with DNase I (Amplification Grade, Sigma-Aldrich). We carried out a control PCR reaction, using the histone 2A–specific primers H2A_fwd and H2A_rev on a 2-μl aliquot of DNase-treated poly(A^+) RNA, to test for any genomic DNA contamination. We reverse-transcribed the remaining DNase-treated poly(A^+) RNA to cDNA using the Promega Reverse Transcription System with random primers on the first sample and oligodT primer on the second sample. We increased the total reaction volume to 40 μl and prolonged the first-strand synthesis step to 60 min. Reverse transcriptase was heat-deactivated for 5 min at 95 °C. We verified the cDNA synthesis reaction quality using the supplied positive control system.

Genomic and RT-PCR. We carried out both genomic PCR and RT-PCR reactions using the HotStarTaq Master Mix Kit (Qiagen) in a final reaction volume of 15 μl (7.5 μl of HotStarTaq Mix (2×); 1 μl of forward primer (10 μM); 1 μl of reverse primer (10 μM); 2 μl of template; and water to equalize volume). We used 20 ng of genomic DNA for genomic PCR and 2 μl of 1:5 v/v dilution of cDNA in sterile ddH_2O as template in genomic and RT-PCR reactions, respectively. We included genomic DNA from oat and the maize donor line as negative and positive controls in the oat-maize addition lines PCR analysis. We carried out thermal cycling in a GeneAmp PCR System 9700 thermo-cycler (Applied Biosystems) under the following conditions: one step of 10 min at 95 °C; followed by 35 cycles of 45 s at 95 °C, 45 s at 60 °C and 90 s at 72 °C; and one final extension of 7 min at 72 °C. We adjusted annealing temperature and elongation time on the basis of the primer and amplification size conditions. We visualized PCR results on 1.5% ethidium bromide 1× TBE-agarose gels and sequenced them as described[6].

URLs. BAC contig maps of line B73 are available from the Arizona Genomics Institute (http://www.genome.arizona.edu/fpc/maize/). Helitron elements identified in *A. thaliana* are available at http://www.girinst.org/~vladimir/RC/Data1.htm. We obtained FGENSH splicing site predictions from http://www.softberry.com/berry.phtml. Primer3 is available at http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank F. Salamini and S. Tingey for critical reading of the manuscript and Y. Zhang for writing the script used for the overgo probe comparison in the two physical maps. M.M. is supported by a DuPont Young Professor Grant. This research is partly supported by an Italian Ministry of University and Research, PRIN projects grant to M.M.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Received 1 April; accepted 27 June 2005
Published online at http://www.nature.com/naturegenetics/

1. Kapitonov, V.V. & Jurka, J. Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **98**, 8714–8719 (2001).
2. Poulter, R.T., Goodwin, T.J. & Butler, M.I. Vertebrate helentrons and other novel *Helitrons*. *Gene* **313**, 201–212 (2003).
3. Lal, S.K., Giroux, M.J., Brendel, V., Vallejos, C.E. & Hannah, L.C. The maize genome contains a *helitron* insertion. *Plant Cell* **15**, 381–391 (2003).
4. Fu, H. & Dooner, H.K. Intraspecific violation of genetic colinearity and its implications in maize. *Proc. Natl. Acad. Sci. USA* **99**, 9573–9578 (2002).
5. Song, R. & Messing, J. Gene expression of a gene family in maize based on noncollinear haplotypes. *Proc. Natl. Acad. Sci. USA* **100**, 9055–9060 (2003).
6. Brunner, S., Fengler, K., Morgante, M., Tingey, S. & Rafalski, A. Evolution of DNA sequence nonhomologies among maize inbreds. *Plant Cell* **17**, 343–360 (2005).
7. Meyers, B.C., Scalabrin, S. & Morgante, M. Mapping and sequencing complex genomes: let's get physical! *Nat. Rev. Genet.* **5**, 578–588 (2004).
8. Gardiner, J. *et al.* Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. *Plant Physiol.* **134**, 1317–1326 (2004).
9. Bennetzen, J.L., Coleman, C., Liu, R., Ma, J. & Ramakrishna, W. Consistent over-estimation of gene number in complex plant genomes. *Curr. Opin. Plant Biol.* **7**, 732–736 (2004).
10. Palmer, L.E. *et al.* Maize genome sequencing by methylation filtration. *Science* **302**, 2115–2117 (2003).
11. Messing, J. *et al.* Sequence composition and genome organization of maize. *Proc. Natl Acad. Sci. USA* **101**, 14349–14354 (2004).
12. Ramakrishna, W., Emberton, J., Ogden, M., SanMiguel, P. & Bennetzen, J.L. Structural analysis of the maize rp1 complex reveals numerous sites and unexpected mechanisms of local rearrangement. *Plant Cell* **14**, 3213–3223 (2002).
13. Craig, N.L., Craigie, R., Gellert, M. & Lambowitz, A.M. *Mobile DNA II* (American Society of Microbiology Press, Washington, DC, 2002).
14. Gupta, S., Gallavotti, A., Stryker, G.A., Schmidt, R.J. & Lal, S.K. A novel class of *Helitron*-related transposable elements in maize contains portions of multiple pseudogenes. *Plant Mol. Biol.* **57**, 115–127 (2005).
15. Feschotte, C. & Wessler, S.R. Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* **98**, 8923–8924 (2001).
16. Kynast, R.G. *et al.* A complete set of maize individual chromosome additions to the oat genome. *Plant Physiol.* **125**, 1216–1227 (2001).
17. Okagaki, R.J. *et al.* Mapping maize sequences to chromosomes using oat-maize chromosome addition materials. *Plant Physiol.* **125**, 1228–1235 (2001).
18. Song, R., Llaca, V. & Messing, J. Mosaic organization of orthologous sequences in grass genomes. *Genome Res.* **12**, 1549–1555 (2002).
19. Lai, J. *et al.* Gene loss and movement in the maize genome. *Genome Res.* **14**, 1924–1931 (2004).
20. Swigonova, Z., Bennetzen, J.L. & Messing, J. Structure and evolution of the r/b chromosomal regions in rice, maize, and sorghum. *Genetics* **169**, 891–906 (2005).
21. Ilic, K., SanMiguel, P.J. & Bennetzen, J.L. A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. *Proc. Natl. Acad. Sci. USA* **100**, 12265–12270 (2003).
22. Hamilton, A.J. & Baulcombe, D.C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
23. van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N. & Stuitje, A.R. Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* **2**, 291–299 (1990).
24. Duvick, D.N. Biotechnology in the 1930s: the development of hybrid maize. *Nat. Rev. Genet.* **2**, 69–74 (2001).
25. Birchler, J.A., Auger, D.L. & Riddle, N.C. In search of the molecular basis of heterosis. *Plant Cell* **15**, 2236–2239 (2003).
26. Jiang, N., Bao, Z., Zhang, X., Eddy, S.R. & Wessler, S.R. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **431**, 569–573 (2004).
27. Yu, Z., Wright, S.I. & Bureau, T.E. Mutator-like elements in *Arabidopsis thaliana*. Structure, diversity and evolution. *Genetics* **156**, 2019–2031 (2000).
28. Le, Q.H., Wright, S., Yu, Z. & Bureau, T. Transposon diversity in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **97**, 7376–7381 (2000).
29. Lai, J., Li, Y., Messing, J. & Dooner, H.K. Gene movement by *Helitron* transposons contributes to the haplotype variability of maize. *Proc. Natl. Acad. Sci. USA* **102**, 9068–9073 (2005).
30. Meyers, B.C., Tingey, S.V. & Morgante, M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* **11**, 1660–1676 (2001).

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics