Cytogenet Genome Res 110:91–107 (2005)
DOI: 10.1159/000084941

Cytogenetic and Genome Research

LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model

C. Vitte,[a] O. Panaud[b]

[a] Laboratoire Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay;
[b] Laboratoire Genome et Developpement des Plantes, Perpignan (France)

Manuscript received 10 December 2003; accepted in revised form for publication by J.-N. Volff 14 April 2004.

Abstract. Long Terminal Repeat (LTR) retrotransposons are ubiquitous components of plant genomes. Because of their copy-and-paste mode of transposition, these elements tend to increase their copy number while they are active. In addition, it is now well established that the differences in genome size observed in the plant kingdom are accompanied by variations in LTR retrotransposon content, suggesting that LTR retrotransposons might be important players in the evolution of plant genome size, along with polyploidy. The recent availability of large genomic sequences for many crop species has made it possible to examine in detail how LTR retrotransposons actually drive genomic changes in plants. In the present paper, we provide a review of the recent publications that have contributed to the knowledge of plant LTR retrotransposons, as structural components of the genomes, as well as from an evolutionary genomic perspective. These studies have shown that plant genomes undergo genome size increases through bursts of retrotransposition, while there is a counteracting process that tends to eliminate the transposed copies from the genomes. This process involves recombination mechanisms that occur either between the LTRs of the elements, leading to the formation of solo-LTRs, or between direct repeats anywhere in the sequence of the element, leading to internal deletions. All these studies have led to the emergence of a new model for plant genome evolution that takes into account both genome size increases (through retrotransposition) and decreases (through solo-LTR and deletion formation). In the conclusion, we discuss this new model and present the future prospects in the study of plant genome evolution in relation to the activity of transposable elements.

Copyright © 2005 S. Karger AG, Basel

LTR retrotransposons are ubiquitous in the plant kingdom (Kumar and Bennetzen, 1999) and are actually the main constituents of large plant genomes. In addition, because they transpose via an mRNA intermediate, LTR retrotransposons indeed tend to increase their copy number while they are active. Retrotransposition is therefore now considered to be a major force in plant genome evolution. As a consequence, over the past few years, the study of LTR retrotransposons has become a main focus in plant structural and evolutionary genomics.

In this review, we provide an update of the current knowledge of plant LTR retrotransposons, with a particular emphasis on their impact on plant genome size. Other forces, such as polyploidy, are well-known to contribute to genome size variations. We however have chosen to focus on the impact of LTR retrotransposons on genome size. Other forces will therefore not be discussed in this paper.

Supported by the French Ministry of Research and Education.

Request reprints from O. Panaud
Laboratoire Genome et Developpement des Plantes
52 avenue Paul Alduy, FR–66860 Perpignan cedex (France)
telephone: +33468661773; fax: +33468668499; e-mail: panaud@univ-perp.fr

KARGER
Fax + 41 61 306 12 34
E-mail karger@karger.ch
www.karger.com

© 2005 S. Karger AG, Basel
1424–8581/05/1104–0091$22.00/0

Accessible online at:
www.karger.com/cgr

Current knowledge on plant retrotransposons and their genomic and evolutionary dynamics

Overview of LTR retrotransposon characterization

The first plant LTR retrotransposons described were mostly characterized as byproducts of defective mutant characterization (Harberd et al., 1987; Grandbastien et al., 1989; Jin and Bennetzen, 1989; Camirand and Brisson, 1990) or by chance while looking for polymorphisms (Voytas and Ausubel, 1988). Later attempts to isolate genes of interest, through the production of genetic markers (Bhattacharyya et al., 1997; Erdmann et al., 2002) and the construction of small insert genomic libraries (Lee et al., 1990; Manninen and Schulman, 1993; Hu et al., 1995; Li et al., 2000; van Leeuwen et al., 2003) also led to the fortuitous characterization of LTR retrotransposons. Some families have also been discovered while analysing repeated, heterochromatic or centromeric sequences (Pelissier et al., 1995; Nakajima et al., 1996; Ananiev et al., 1998; Presting et al., 1998; Linares et al., 1999, 2001; Francki, 2001; Kentner et al., 2003), as well as retroviruses (Thomson et al., 1998). More recently, the availability of large insert libraries, along with the development of genome sequencing projects, has yielded numerous large plant genomic sequences and therefore facilitated the discovery of new LTR retrotransposons. This has been particularly the case for plant species of agronomic interest such as barley (Panstruga et al., 1998; Wei et al., 1999; Shirasu et al., 2000; Rostoks et al., 2002), maize (SanMiguel et al., 1996; Fu and Dooner, 2002), rice (Llaca et al., 1998, GenBank accession AF111709; Tarchini et al., 2000), wheat (Wicker et al., 2001; SanMiguel et al., 2002) and tomato (Mao et al., 2001). In addition, the completion of *Arabidopsis thaliana* and rice *(Oryza sativa)* genome projects has enabled the full characterization of LTR retrotransposons for which only partial sequences were cloned (Panaud et al., 2002). It has also made possible systematic computer-based mining of transposable elements (TEs), leading to the characterization of nearly complete repertoires of LTR retrotransposon families in *A. thaliana* (Kapitonov and Jurka, 1999; Marin and Llorèns, 2000; Terol et al., 2001; Wright and Voytas, 2001) and rice (McCarthy et al., 2002).

The availability of these numerous plant LTR retrotransposon sequences has allowed the discovery of several common features, such as the presence of primer binding sites (PBS) and conserved coding domains such as reverse transcriptase *(RT)* and integrase *(Int)*. Degenerate polymerase chain reaction (PCR) amplifications of these conserved motives has been applied on many plant species (Flavell et al., 1992a, b; Hirochika et al., 1992; VanderWiel et al., 1993; Pearce et al., 1996, 1997; Wang et al., 1997; Gribbon et al., 1999; Nakatsuka et al., 2002; Stergiou et al., 2002) and the partial sequences produced have been exploited for chromosomal localisation and phylogenetic studies of LTR retrotransposons, without any prior knowledge of the complete corresponding element. For some of these partially known elements, a full characterization has been subsequently achieved through the screening of genomic libraries using the partial regions as probes (Konieczny et al., 1991; Hirochika et al., 1992; Costa et al., 1999; Kumekawa et al., 1999; Balint-Kurti et al., 2000; Lall et al., 2002). In addition, the discovery of mutations induced by LTR retrotransposons suggested that active elements could be cloned from cDNA libraries. This was the case for *Tos17* of rice (Hirochika et al., 1996) and *Tto1* of tomato (Hirochika, 1993).

So far, over 150 LTR retrotransposon families have been found and fully characterized in more than 20 species of angiosperms. They are presented in Table 1.

Plant genome organization and LTR retrotransposon localization

Reassociation kinetics studies showed that non-transcribing repeat (NTR)-DNA is an integral part of most plant genomes (Flavell et al., 1974). Such NTR-DNA is unevenly distributed in plant genomes, as shown by reports based on density gradient centrifugation (Barakat et al., 1998), cytogenetic studies (Curtis and Lukaszewski, 1991; Gill et al., 1991) and comparison of genetic and physical maps (Faris et al., 2000; Kunzel et al., 2000; Sandhu et al., 2001). They are mainly located around centromeres, but are also present as large blocks separating gene-rich regions (Sandhu and Gill, 2002). They are mainly composed of retrotransposons and pseudogenes (derived from multigene families for instance) and are highly heterochromatic (Sandhu and Gill, 2002).

The uneven distribution of genes along genomes has been confirmed by the recent analyses of large genomic sequences (mainly from Bacterial Artificial Chromosome [BAC], Yeast Artificial Chromosome [YAC] and Transformation-competent Artificial Chromosome [TAC] clones) that have been performed for several species such as barley (Panstruga et al., 1998; Feuillet and Keller, 1999; Wei et al., 1999; Shirasu et al., 2000; Dubcovsky et al., 2001; Rostoks et al., 2002; Brunner et al., 2003), maize (SanMiguel et al., 1996; Tikhonov et al., 1999; Fu et al., 2001; Fu and Dooner, 2002), rice (Chen and Bennetzen, 1996; Tarchini et al., 2000; Dubcovsky et al., 2001), sorghum (Tikhonov et al., 1999), wheat (Rahman et al., 1997; Feuillet and Keller, 1999; Keller and Feuillet, 2000; Wicker et al., 2001, 2003; SanMiguel et al., 2002), lotus (Sato et al., 2001; Nakamura et al., 2002), peach (Georgi et al., 2003), soybean (Foster-Hartnett et al., 2002) and tomato (Ku et al., 2000; Mao et al., 2001): for genomes not exceeding 500 Mb in size, the observed gene density is close to that predicted under a random gene distribution model and approaches the observed gene density described for *A. thaliana* (~1 gene/4–5 kb). For species with larger genomes, however, gene densities appear to be much higher than predicted by the random distribution model. For example, most of the reported gene densities for barley range from 1 gene/12 kb to 1 gene/20 kb, depending on the genomic regions studied, which is much higher than the expected value of ~1 gene/200 kb (see Table 2 for details). These observations suggest that genes are not evenly distributed in the genomes and that there is a bias towards gene-rich regions among the genomic sequences available. Most of the early studies were indeed performed on regions of agronomic interest, thus containing genes. More recent analyses on randomly-chosen genomic regions confirmed that gene-poor sequences (~1 gene/100 kb, Rostoks et al., 2002) are also found in barley. Hence, large genomes seem to be organized into gene-dense regions interrupted by heterochromatic gene-poor blocks containing mostly retrotransposons.

Table 1. Description of plant LTR retrotransposons. This table highlights how each retrotransposon family was discovered, its copy number and the corresponding method of estimation. Characteristic structural parameters such as total length (in kbp), LTR length (in bp) and classification into *gypsy*-like and *copia*-like classes are also presented. The name of the species corresponds to the species where the retrotransposon was first discovered. When the retrotransposon was subsequently found in other species, this information is not presented.

RE	Type	Total size	LTR size	Method of identification[a]	Copy no.	Method of estimation	Reference[b]
Monocots							
Poaceae							
Barley *(Hordeum vulgare)*							
BARE-1	Gypsy	12100	1829	Seq. analysis of a genom. lib. clone	16600 +/– 600	Slot blot hyb.	Manninen and Schulman, 1993 Vicient et al., 1999
BAGY-1	Gypsy	14400	4202–4208	BAC seq. of the *Mlo* region	n.a.	n.a.	Panstruga et al., 1998
BAGY-2	Gypsy	>8000	1523	YAC seq. of the *Rar1* locus	n.a.	n.a.	Shirasu et al., 2000
Cereba	Gypsy	9800	154	Genom. lib. scr. using centromeric prob (*int*)	~1500	Gen. lib. scr. using *int* probe	Presting et al., 1998
Horgy-1	Gypsy	12100	n.a.	BAC seq. of the *Mla* region	n.a.	n.a.	Wei et al., 2002
Horpia-1	Copia	>3600	n.a.	BAC seq. of the *Mla* region	n.a.	n.a.	Wei et al., 2002
Horpia-2	Copia	5100	n.a.	BAC seq. of the *Mla* region	n.a.	n.a.	Wei et al., 2002
Horpia-3	Copia	2000	n.a.	BAC seq. of the *Mla* region	n.a.	n.a.	Wei et al., 2002
Inga	Copia	11800	1571	BAC seq. of four different chrom.	n.a.	n.a.	Rostoks et al., 2002
Nikita	solo	n.a.	2930	YAC seq. of the *Rar1* locus	n.a.	n.a.	Shirasu et al., 2000
Sabrina	n.a.	6800	1620	YAC seq. of the *Rar1* locus	n.a.	n.a.	Shirasu et al., 2000
Sukkula	solo	n.a.	4960	YAC seq. of the *Rar1* locus	n.a.	n.a.	Shirasu et al., 2000
Zea diploperennis							
Grande-1	Gypsy	13800	630	Gen. lib. scr. for glutelin genes and following clone charact.	1300–1700	Slot blot hyb.	Martinez-Izquierdo et al., 1997
Maize *(Zea mays)*							
B5	n.a.	6100	299	Charact. as an insertion in the *Wx* gene	n.a.	n.a.	Varagona et al., 1992 Vignols et al., 1995
Bs1	Copia	3200	302	Charact. of an *Adh1*-null allele	1–5	Southern blot hyb.	Johns et al., 1985, 1989 Jin and Bennetzen, 1989
CentA	Gypsy	4600	1304	Analysis of centromeric regions	n.a.	n.a.	Ananiev et al., 1998
Cin1	solo	n.a.		Charact. as an insertion in *Wx* gene	n.a.	n.a.	Shepherd et al., 1984
Cinful	Gypsy	8500	600	YAC seq. of the *Adh1*-F locus	n.a.	n.a.	SanMiguel et al., 1996, 1998
Fourf	Copia	7000	1100	YAC seq. of the *Adh1*-F locus	Hundreds	Reverse southern blot hyb.	SanMiguel et al., 1996
G	n.a.	5000	n.a.	Charact. as an insertion in *Wx* gene	n.a.	n.a.	Varagona et al., 1992
Grande-zm	Gypsy	10500	600	YAC seq. of the *Adh1*-F locus	n.a.	n.a.	SanMiguel et al., 1996
Hopscotch	Copia	4800	231	Charact. as an insertion in *Wx* gene	2–6	n.a.	White et al., 1994
Huck	Gypsy	11500	1500	YAC seq. of the *Adh1*-F locus	n.a.	n.a.	SanMiguel et al., 1996
Ji	Copia	8500	1300	YAC seq. of the *Adh1*-F locus	~50000	YAC hyb.	SanMiguel et al., 1996
Kake	n.a.	6500	200	YAC seq. of the *Adh1*-F locus	Hundreds	Reverse southern blot hyb.	SanMiguel et al., 1996
Magellan	Gypsy	>5700	341	Charact. as an insertion in the *Wx* gene	4–8	Southern blot hyb.	Purugganan and Wessler, 1994 Chavanne et al., 1998
Milt	n.a.	4500	700	YAC seq. of the *Adh1*-F locus	n.a.	n.a.	SanMiguel et al., 1996
Opie	Copia	8000	1300	YAC seq. of the *Adh1*-F locus	>30000	YAC hyb.	SanMiguel et al., 1996
PREM-1	n.a.	8300	3502–3524	Charact. of tissue-specific genes	10000–40000	Slot blot hyb. ; lib. Scr.	Turcich et al., 1994 Fu et al., 2001
PREM-2	Copia	9400	1307	Genom. lib. scr. using *PREM-1* LTR as probe	30000	Gen. lib. scr.	Turcich et al., 1996
Reina	Gypsy	5500	300	YAC seq. of the *Adh1*-F locus	~10	Southern blot hyb.	SanMiguel et al., 1996
Stonor	Copia	6000	560	Charact. as an insertion in *Wx* gene	30–40	Southern blot hyb.	Varagona et al., 1992 Marrillonet and Wessler, 1998
Tekay	Gypsy	12100	3441	n.a.	n.a.	n.a.	San Miguel and Bennetzen, 1998 (unpubl.)[1]
Victim	Copia	5500	100	YAC seq. of the *Adh1*-F locus	Hundreds	Reverse Southern blot hyb.	SanMiguel et al., 1996
Xilon1	n.a.	11700	2703–2707	BAC seq. of the *Bz-McC* locus	n.a.	n.a.	Fu and Dooner, 2002
Zdel	Gypsy	n.a.	n.a.	Charact. as an insertion in *Grande-1*	~100	Southern blot hyb.	Vicient and Martinez-Izquierdo, 1997
Zeon-1	Gypsy	7300	648	Charact. of the rearranged 27 kDa γ zein locus	>1250	Southern blot hyb.	Hu et al., 1995
Oat *(Avena strigosa)*							
As17	Copia	n.a.	646	Charact. of repeated DNA fragments	13000	Slot blot hyb.	Linares et al., 1999
TAS-1	Copia	>2600	n.a. (p.c.)	Charact. of repeated DNA fragments	10000	n.a.	Linares et al., 2001
Oryza australiensis							
RIRE1	Copia	8300	1523	Analysis of the interspersed seq. pOa4	7500–64000	Slot blot hyb.	Nakajima et al., 1996 Noma et al., 1997
Rice *(Oryza sativa)*							
Copia-like A	Copia	7700	275	BAC seq. of the *Adh1-2* region	n.a.	n.a.	Tarchini et al., 2000
Copia-like B	Copia	3000	n.a.	BAC seq. of the *Adh1-2* region	n.a.	n.a.	Tarchini et al., 2000
Copia-like C	Copia	5000	1093	BAC seq. of the *Adh1-2* region	n.a.	n.a.	Tarchini et al., 2000
Copia-like D	Copia	5700	503	BAC seq. of the *Adh1-2* region	n.a.	n.a.	Tarchini et al., 2000
Dagul	Gypsy	13400	3622	RDA cloning and *in silico* charact.	n.a.	n.a.	Panaud et al., 2002
Gypsy-like A	Gypsy	8900	3367	BAC seq. of the *Adh1-2* region	n.a.	n.a.	Tarchini et al., 2000
Gypsy-like B	Gypsy	7000	702	BAC seq. of the *Adh1-2* region	n.a.	n.a.	Tarchini et al., 2000
Gypsy-like C	Gypsy	11500	3072	BAC seq. of the *Adh1-2* region	n.a.	n.a.	Tarchini et al., 2000
Hopi/Osr27	Gypsy	12800	1102	RDA cloning and *in silico* charact.	900	Based on no. of hits on 259Mb	Panaud et al., 2002 McCarthy et al., 2002

Table 1 (continued)

RE	Type	Total size	LTR size	Method of identification[a]	Copy no.	Method of estimation	Reference[b]
Houba/Tos5/ Osr13	Copia	6400	967	Genome lib. scr., RDA cloning and *in silico* charact.	650	Based on no. of hits on 259Mb	Hirochika et al., 1992 Panaud et al., 2002 McCarthy et al., 2002
Osr1	Copia	6400	965	LTR_STRUC genome scanning	250	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr2	Copia	4900	267	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr3	Copia	5200	146	LTR_STRUC genome scanning	50–100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr4	Copia	5700	350	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr5	Copia	6100	477	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr6	Copia	5200	440	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr7	Copia	8900	1608	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr8	Copia	9200	1220	LTR_STRUC genome scanning	1100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr9	Copia	n.a.	n.a.	LTR_STRUC genome scanning	50–100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr10	Copia	n.a.	n.a.	LTR_STRUC genome scanning	400	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr11	Copia	n.a.	n.a.	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr12	Copia	4700	221	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr14	Copia	8400	319	LTR_STRUC genome scanning	350	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr15	Copia	5100	262	LTR_STRUC genome scanning	250	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr16	Copia	6900	300	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr17	Copia	6000	501	LTR_STRUC genome scanning	50–100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr18	Copia	n.a.	n.a.	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr19	Copia	4700	205	LTR_STRUC genome scanning	50–100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr20	Copia	5500	286	LTR_STRUC genome scanning	50–100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr22	Copia	4600	191	LTR_STRUC genome scanning	50–100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr23	Copia	4800	209	LTR_STRUC genome scanning	50–100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr24	Copia	4900	221	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr25/ Dasheng	n.a.	6800	417	LTR_STRUC genome scanning	800–1300	Gen. lib. scr., computer search	McCarthy et al., 2002 Jiang et al., 2002a and 2002b
Osr28	Gypsy	18000	2195	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr29	Gypsy	9000	656	LTR_STRUC genome scanning	550	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr30	Gypsy	13000	1507	LTR_STRUC genome scanning	1500	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr31	Gypsy	7400	787	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr32	Gypsy	n.a.	n.a.	LTR_STRUC genome scanning	50–100	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr34	Gypsy	12800	3292	LTR_STRUC genome scanning	450	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr35	Gypsy	5700	423	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr36	Gypsy	5200	319	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr37	n.a.	4400	794	LTR_STRUC genome scanning	600	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr38	Gypsy	5500	332	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr39	Gypsy	5200	368	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr40	Gypsy	11400	564	LTR_STRUC genome scanning	600	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr41	Gypsy	15700	518	LTR_STRUC genome scanning	300	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr42	Gypsy	5600	358	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Osr43	n.a.	1800	291	LTR_STRUC genome scanning	<50	Based on no. of hits on 259Mb	McCarthy et al., 2002
Osr44	n.a.	1200	148	LTR_STRUC genome scanning	<50	Based on no. of hits on 259 Mb	McCarthy et al., 2002
Retrosat1/RI RE2/Osr26	Gypsy	10900	440	52,6 kb genom. seq., RDA cloning and in silico charact.	500	Based on no. of hits on 259 Mb	Llaca et al., 1998 (unpubl.)[2] Ohtsubo et al., 1999 McCarthy et al., 2002
Retrosat2	Gypsy	12800	3293–3295	BAC seq. analysis	n.a.	n.a.	Llaca et al., 1999 (unpubl.)[3]
RIRE3	Gypsy	12000	3154	Scr. of genom. lib. using chance-defined *RT* primers	25	Searches on 30% of the genome seq.	Kumekawa et al., 1999 Panaud et al., 2002 Vitte and Panaud, 2003
RIRE7	Gypsy	7600	858	Charact. of an insertion within a *RIRE3* copy	410–1700	Slot blot hyb.	Kumekawa et al., 2001
RIRE8/Osr33	Gypsy	11800	2948	Scr. of genom. lib. using chance-defined *RT* primers	550	Based on no. of hits on 259 Mb	Kumekawa et al., 1999 Panaud et al., 2002 McCarthy et al., 2002
RIRE9	Gypsy	n.a. (p.c.)	n.a. (p.c.)	Disease resistance gene isolation	~1600	Dot blot hyb.	Li et al., 2000
Spip	n.a.	10800	3315	*In silico* charact.	39	Searches on 50% of the genome seq.	Vitte and Panaud (submitted)[4]
Squiq	n.a.	8600	3410	*In silico* charact.	17	Searches on 50% of the genome seq.	Vitte and Panaud (submitted)[5]
Tos1	Copia	n.a.	143–145	Genome lib. scr. using primer complementary to PBS	~30	Genom. lib. scr.	Hirochika et al., 1992
Tos17/Osr21	Copia	4100	138	Tissue cultured followed by RT-PCR method	1–4	Southern blot hyb.	Hirochika et al., 1996 McCarthy et al., 2002
Tos18-Tos20	n.a.	n.a. (p.c.)	n.a. (p.c.)	Tissue culture followed by RT-PCR method	n.a.	n.a.	Hirochika et al., 1996
Tos2	Copia	n.a.	220–245	Genome lib. scr. using primer complementary to PBS	~30	Genom. lib. scr.	Hirochika et al., 1992
Tos3	Copia	5200	115	Genome lib. scr. using primer complementary to PBS	n.a.	n.a.	Hirochika et al., 1992
Tos4	Copia	n.a. (p.c.)	n.a. (p.c.)	Genome lib. scr. using *RT* probes	n.a.	n.a.	Hirochika et al., 1992
Tos6-Tos16	n.a.	n.a. (p.c.)	n.a. (p.c.)	RT PCR amplification and cloning followed by cross-hyb.	n.a.	n.a.	Hirochika et al., 1996
Rye *(Secale cereale)*							
Bilby	Copia	n.a. (p.c.)	n.a. (p.c.)	Analysis of centromeric repetitive elements	n.a.	n.a.	Francki, 2001
Sorghum *(Sorghum bicolor)*							
Levithan	n.a.	15200	4560	n.a.	n.a.	n.a.	Liu and Bennetzen, 2000 (unpubl.)[7]
Retrosor1	Gypsy	13500	701	Analysis of retroelements in a gene-dense region	n.a.	n.a.	Llaca et al., 1999 (unpubl.)[6]

Table 1 (continued)

RE	Type	Total size	LTR size	Method of identification[a]	Copy no.	Method of estimation	Reference[b]
Wheat *(Triticum aestivum)*							
Angela	Copia	8500	1720	211 kb BAC seq. analysis	n.a.	n.a.	Wicker et al., 2001
Barbara	Copia	9800	n.a.	211 kb BAC seq. analysis	n.a.	n.a.	Wicker et al., 2001
Claudia	Copia	>3000	n.a. (p.c.)	211 kb BAC seq. analysis	n.a.	n.a.	Wicker et al., 2001
Daniela	Gypsy	13300	936	211 kb BAC seq. analysis	n.a.	n.a.	Wicker et al., 2001
Erika-1	Gypsy	14100	4200	211 kb BAC seq. analysis	n.a.	n.a.	Wicker et al., 2001
Fatima	Gypsy	9100	481	211 kb BAC seq. analysis	n.a.	n.a.	Wicker et al., 2001
Greti	Gypsy	4300	n.a. (p.c.)	211 kb BAC seq. analysis	n.a.	n.a.	Wicker et al., 2001
Heidi	Gypsy	>6000	n.a. (p.c.)	211 kb BAC seq. analysis	n.a.	n.a.	Wicker et al., 2001
Latidu	Gypsy	5000	464	215 kb BAC seq. analysis	n.a.	n.a.	SanMiguel et al., 2002
Nusif	Gypsy	>7800	463	215 kb BAC seq. analysis	n.a.	n.a.	SanMiguel et al., 2002
Wham	Gypsy	10000	1415–1418	215 kb BAC seq. analysis	n.a.	n.a.	SanMiguel et al., 2002
Wis2	Copia	8600	1755	Charact. of an insert in the *Glu 1A-2* gene	200	Southern blot hyb.	Harberd et al., 1987 Moore et al., 1991 Lucas et al., 1992
Other Monocots							
Banana *(Musa acuminata)*							
Monkey	Gypsy	n.a. (p.c.)	n.a. (p.c.)	Screen of l genom. lib. using heterologous probes	600–1500	Dot blot hyb.	Balint-Kurti et al., 2000
Iris *(Iris sp.)*							
IRRE1	Gypsy	11000	2800–3000	Screen of a genom. lib. of repeated seq.s	100000	Dot blot hyb. and lib. scr.	Kentner et al., 2003
Lily *(Lilium henryi)*							
Del1-46	Gypsy	9300	2406–2415	Genome lib. scr. using dispersed repeated seq. probes	>13000	DNA reannealing; South. blot. hyb.	Sentry and Smyth, 1989 Smyth et al., 1989
Pineapple *(Ananas comosus)*							
dea-1	Gypsy	n.a. (p.c.)	n.a. (p.c.)	*RT, RNaseH* and *Int* cloning using single primer-PCR	n.a.	n.a.	Thomson et al., 1998
Dicots							
Brassicaceae							
Arabidopsis thaliana							
AtC1	Copia	4900	355	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC2	Copia	4500	216	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC3	Copia	4600	274	Computer mining on 460 kb from chrom. III	2	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC4	Copia	5000	379	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC5	Copia	4700	296	Computer mining on 460 kb from chrom. III	2	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC6	Copia	5100	318	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC7	Copia	5600	732	Computer mining on 460 kb from chrom. III	2	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC8	Copia	4800	128	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC9	Copia	4800	174	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC10	Copia	5000	440	Computer mining on 460 kb from chrom. III	6	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC11	Copia	5700	713	Computer mining on 460 kb from chrom. III	2	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC12	Copia	4800	130	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC13	Copia	5000	158	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC14	Copia	5700	734	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC15	Copia	5200	407	Computer mining on 460 kb from chrom. III	2	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC16	Copia	4600	244	Computer mining on 460 kb from chrom. III	2	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC17	Copia	5100	396	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
AtC18	Copia	4900	130	Computer mining on 460 kb from chrom. III	1	Searches on 92.8% of the genome seq.	Terol et al., 2001
Athila	Gypsy	10500	1539, 1552	Genom. lib. scr. using heterochromatic fragments as probe	Up to 30	Genom. lib. scr. using internal probe	Pelissier et al., 1995, 1996 Wright and Voytas, 1998
Athila 2	Gypsy	11300	1744, 1752	Computer mining on unannotated genome seq.	n.a.	n.a.	Wright and Voytas, 1998
Athila 3	Gypsy	8100	>1200	Computer mining on unannotated genome seq.	n.a.	n.a.	Wright and Voytas, 1998
Athila1-2	Gypsy	7600	1386, 1419	Genome lib. scr.	n.a.	n.a.	Wright and Voytas, 1998
AtRE1	Copia	4800	167	Analysis of the 100cM map unit of chrom. 1	1–2	Southern blot hyb.	Kuwahara et al., 2000
AtRE2	Copia	n.a.	166	Analysis of the 100cM map unit of chrom. 1	1–2	Southern blot hyb.	Kuwahara et al., 2000
Evelknievel	Copia	4700	~200	Comparison of *CMT1* gene organization in distinct ecotypes	1 to few	Southern blot hyb.	Henikoff and Comai, 1998
Gimli	Gypsy	5200	341	Computer mining on the non-redundant NCBI databases	n.a.	n.a.	Marin and Lloréns, 2000
Gloin	Gypsy	5400	359	Computer mining on the non-redundant NCBI databases	n.a.	n.a.	Marin and Lloréns, 2000
Legolas	Gypsy	7700	1347	Computer mining on the non-redundant NCBI databases	n.a.	n.a.	Marin and Lloréns, 2000
Meta-1	Copia	4800	162	Computer mining on chrom. II seq.	n.a.	n.a.	Kapitonov and Jurka, 1999
Ta1	Copia	5200	514	RFLP polymorphism in ecotype Kas-1	1–3	Southern blot hyb.	Voytas and Ausubel, 1988
Ta2	Copia	3000	520	Genome lib. scr. using *Ta1 RT* probes	1	Southern blot hyb.	Konieczny et al., 1991
Ta3	Copia	5000	485–499	Genome lib. scr. using *Ta1 RT* probes	1	Southern blot hyb.	Konieczny et al., 1991
Tat1	Gypsy?	5000	~430	Charact. of a genom. clone containing the *SAM1* locus	2–10	Southern blot hyb.	Peleman et al., 1991 Wright and Voytas, 1998

Table 1 (continued)

RE	Type	Total size	LTR size	Method of identification[a]	Copy no.	Method of estimation	Reference[b]
Tat4	Gypsy	11900	453, 452	Charact. of a genom. clone	n.a.	n.a.	Wright and Voytas, 1998
Tft1	Gypsy	8400	1327	Computer mining on the non-redundant NCBI databases	n.a.	n.a.	Marin and Lloréns, 2000
Tft2	Gypsy	>6600	n.a.	Computer mining on the non-redundant NCBI databases	n.a.	n.a.	Marin and Lloréns, 2000
Tma1	Gypsy	7800	1164, 1158	Computer mining on unannotated genome seq.	n.a.	n.a.	Wright and Voytas, 1998
Tma2	Gypsy	8400	1161, 1488	Computer mining on unannotated genome seq.	n.a.	n.a.	Wright and Voytas, 1998
Tma3	Gypsy	7800	1155, 1054	Computer mining on unannotated genome seq.	n.a.	n.a.	Wright and Voytas, 1998
Tma4	Gypsy	4500	>1200	Computer mining on unannotated genome seq.	n.a.	n.a.	Wright and Voytas, 1998
Fabaceae							
Pea *(Pisum sativum)*							
Cyclops	Gypsy	12300	1504–1594	Charact. of intron 3 of cytosolic NAD-specific GAPDH gene	1000	Genom. lib. scr.	Chavanne et al., 1998
Pdr1	Copia	3900	156	Charact. of a genom. clone containing two legumin genes	200	Southern blot hyb., lib. scr. and S-SAP	Lee et al., 1990 Ellis et al., 1998 Vershinin and Ellis, 1999
Haricot *(Phaseolus vulgaris)*							
Tpv2	Copia	4800	296–297	Genom. lib. scr. using RT-PCR fragment as probe	~40	Southern blot hyb.	Garber et al., 1999
Tpv3g	Gypsy	>2300	n.a. (p.c.)	Charact. of a RAPD marker closely linked to *P* gene	~100	Southern blot hyb. and lib. scr.	Erdmann et al., 2002
Pigeonpea *(Cajanus cajan)*							
Panzee	Copia	5000	380	Genom. lib. scr. using *RT* probe	High	Southern blot hyb.	Lall et al., 2002
Soybean *(Glycine max)*							
diaspora	Gypsy	>4100	n.a. (p.c.)	n.a.	n.a.	n.a.	Laten, 1999 (unpubl.)[8] Vicient et al., 2001
SIRE-1	Copia	11000	1420	Genom. lib. scr. using a single primer amplified probe	>200	Genom. lib. scr.	Laten and Morris, 1993 Laten et al., 1998
Tgmr	Copia	5000	249–260	charact. of a RAPD marker closely linked to rpsl-k gene	Low	Southern blot hyb.	Bhattacharyya et al., 1997
Solanaceae							
Potato *(Solanum tuberosum)*							
Tst1	Copia	5060	285	Charact. of an insertion in a starch phosphorylase gene	n.a.	n.a.	Camirand and Brisson, 1990
Tobacco *(Nicotiana tabacccum)*							
Tnt1	Copia	5300	610	Charact. of spontaneous NR-deficient (NR-) mutant lines	>100	Southern blot hyb.	Grandbastien et al., 1989
Tto1	Copia	5500	574	PCR amplification from cDNA prepared from protoplasts	~30	Slot blot hyb.	Hirochika, 1993
Tomato *(Lycopersicon esculentum)*							
ACO1RPT	Copia	11300	422–424	Inv-PCR to get the *LEACO-1* upstream regulatory seq.	High	Southern blot hyb.	Blume et al., 1997
Lere1	Copia	5500	276	Bac seq. of the *Jointless* locus	n.a.	n.a.	Mao et al., 2001
Retrolyc1	Copia	>3600	596–610	Screen of genom. lib. using *Tnt1-94* probes	30–40	Southern blot hyb.	Costa et al., 1999 Araujo et al., 2001
Other Dicots							
Melon *(Cucumis melo)*							
CURE	Gypsy	10000	2600	Charact. of a genom. clone containing genes *Hlh* and *Drzf*	n.a.	n.a.	van Leeuwen et al., 2003
Grapevine *(Vitis vinifera)*							
Tvv1	Copia	5000	150–156	Chrom. walking after conserved *RT* domain amplification	>28	PCR amplification of 28 UTRs	Pelsy and Merdinoglu, 2002
Vine-1	Copia	2400	287	Charact. as an insertion in *Adhr* gene	Moderate	Southern blot hyb.	Verriès et al., 2000

[a] n.a.: not available; p.c.: partial characterization; seq.: sequence; scr.: screening; chrom.: chromosome(s); genom.: genomic; lib.: library; charact.: characterization; hyb.: hybridization; unpubl.: unpublished. solo: solo-LTR.

[b] [1]GenBank accession no. AF050455; [2]GenBank accession no. AF111709; [3]GenBank accession no. AF111709; [4]GenBank accession no. AY355292; [5]GenBank accession no. AY355293; [6]GenBank accession no. AF098806; [7]GenBank accession no. U07816; [8]GenBank accession no. AF095730.

Table 2. Expected versus observed gene density at different loci in plant genomes

Genome	1C[a]	Type of region	Observed (gene/kb)	Expected[b] (gene/kb)	Reference
Arabidopsis thaliana	125	Global sequence	1 g/4-5 kb	1 g/5 kb	AGI, 2000
Peach *(Prunus persica)*	270	48.5 kb in *evg* region	1 g/7 kb	1 g/11 kb	Georgi et al., 2003
Lotus *(Lotus japonicus)*	466	5.4 Mb from different regions	1 g/8 kb	1 g/18 kb	Sato et al., 2001
Lotus *(Lotus japonicus)*	466	6.5 Mb from different regions	1 g/9 kb	1 g/18 kb	Nakamura et al., 2002
Medicago truncatula	470	na	1 g/6-10 kb	1 g/18 kb	Young et al., 2003
Rice *(Oryza sativa)*	490	28 kb in *Sh2-A1* region	1 g/8 kb	1 g/19 kb	Chen and Bennetzen, 1996
Rice *(Oryza sativa)*	490	340 kb in *Adh1-Adh2* region	1 g/10.3 kb	1 g/19 kb	Tarchini et al., 2000
Rice *(Oryza sativa)*	490	340 kb from chromosome 2	1 g/6.1 kb	1 g/19 kb	Mayer et al., 2001
Rice *(Oryza sativa)*	490	50 kb from chromosome 3	1 g/10 kb	1 g/19 kb	Dubcovsky et al., 2001
Tomato *(Lycopersicon esculentum)*	950	105 kb in *Ovate* region	1 g/6.2 kb	1 g/37 kb	Ku et al., 2000
Tomato *(Lycopersicon esculentum)*	950	119 kb in *Jointless* region	1 g/8 kb	1 g/37 kb	Mao et al., 2001
Sorghum *(Sorghum bicolor)*	1103	78 kb in *Adh1* region	1 g/5 kb	1 g/43 kb	Tikhonov et al., 1999
Soybean *(Glycine max)*	1115	200 kb BAC-end and subclone	1 g/14 kb	1 g/44 kb	Foster-Hartnett et al., 2002
Maize *(Zea mays)*	2670	78 kb around 22-kDa alpha zein genes[c]	1 g/6 kb	1 g/105 kb	Llaca and Messing, 1998
Maize *(Zea mays)*	2670	225 kb in *Adh1* region	1 g/25 kb	1 g/105 kb	Tikhonov et al., 1999
Maize *(Zea mays)*	2670	60 kb in *bz* region gene	1 g/3.2 kb	1 g/105 kb	Fu et al., 2001
Barley *(Hordeum vulgare)*	5000	60 kb in *Mlo* region	1 g/20 kb	1 g/196 kb	Panstruga et al., 1998
Barley *(Hordeum vulgare)*	5000	16 kb in *HvLrk* region	1 g/15 kb	1 g/196 kb	Feuillet and Keller, 1999
Barley *(Hordeum vulgare)*	5000	261 kb *Mla* region [c]	1 g/18 kb	1 g/196 kb	Wei et al., 1999
Barley *(Hordeum vulgare)*	5000	66 kb in *Rar1* region	1 g/22 kb	1 g/196 kb	Shirasu et al., 2000
Barley *(Hordeum vulgare)*	5000	102 kb from chromosome 5H	1/20 kb	1 g/196 kb	Dubcovsky et al., 2001
Barley *(Hordeum vulgare)*	5000	417 kb from different regions	1 g/12 kb to 1 g/103 kb	1 g/196 kb	Rostoks et al., 2002
Barley *(Hordeum vulgare)*	5000	112 kb in *Rph7* region [c]	1 g/20 kb region aver.	1 g/196 kb	Brunner et al., 2003
Triticum taushii	n.a.	16 kb in *SBE-1* region [c]	1 g/5 kb	1 g/196 kb	Rahman et al., 1997
Triticum taushii	n.a.	100 kb in *Cre3* region	1 g/15 kb	1 g/196 kb	Keller and Feuillet, 2000
Triticum taushii	n.a.	75 kb in *Lrk10* region	1 g/15 kb	1 g/196 kb	Keller and Feuillet, 2000
Triticum monococcum	6100	150 kb in *Lrk10* region	>1 g/25 kb	1 g/239 kb	Keller and Feuillet, 2000
Wheat *(Triticum aestivum)*	16980	16 kb in *Lrk10* region	1 g/5 kb	1 g/222 kb	Feuillet and Keller, 1999
Wheat *(Triticum aestivum)*	16980	211 kb from chromosome 1Am	1 g/42 kb region aver.	1 g/222 kb	Wicker et al., 2001
Wheat *(Triticum aestivum)*	16980	215 kb from chromosome 5Am	1 g/43 kb region aver.	1 g/222 kb	SanMiguel et al., 2002

[a] 1C genome sizes have been extracted from the angiosperm C-value database (http://www.rbgkew.org.uk/cval/homepage.html), except for *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000). They are presented in mega base-pair (Mbp = 10^6 bp). *Triticum taushii* (not available).
[b] Expected gene densities were calculated by dividing the gene content that has been estimated for *Arabidopsis thaliana* (*i.e.,* 25000 genes) by the 1C genome size value estimated for each species. For hexaploid wheat *Triticum aestivum*, the number of genes used for the calculation was 75000, to compensate for ploidy level.
[c] Corresponds to a region containing duplicated genes.

Moreover, the analysis of the large genomic sequences obtained from regions of agronomic interest has provided insights into the structural organization of gene-dense regions. It has revealed that these regions can be separated into gene-rich and inter-gene sub-regions. The gene-density of the gene-rich sub-regions is comparable to that of *A. thaliana* (from a few kilobases to ~20 kb, see Table 2 for details). The inter-gene sub-regions are composed of repetitive DNA, with the predominance of LTR retrotransposons, Miniature Inverted Terminal Elements (MITEs) and transposons. However, due to their large size, LTR retrotransposons are generally the largest component of such regions. In the gene-rich regions of large genomes such as maize, barley and wheat, LTR retrotransposons are often inserted one within another, thus forming stretches of nested retrotransposons (SanMiguel et al., 1996, 1998; Wei et al., 1999; Fu et al., 2001; Wicker et al., 2001; 2003; Fu and Dooner, 2002; Rostoks et al., 2002). In a plant with a smaller genome like rice, however, LTR retrotransposons seem to be more dispersed in such regions (Tarchini et al., 2000), suggesting that differences in terms of retrotransposon content may have led to genome size differences not only in the large heterochromatic blocks, but also within the gene-rich regions.

Taken together, these observations suggest that large plant genomes are composed of large blocks of heterochromatic DNA, comprised mainly of retrotransposons interspersing gene-rich regions that are composed of stretches of repetitive sequences interrupted by gene-rich sub-regions with a gene-density comparable to that of *Arabidopsis thaliana*.

LTR retrotransposons are of ancient origin and highly dynamic

The conservation of coding domain sequences such as *RT* and *Int* within *copia*-like and within *gypsy*-like LTR retrotransposons has allowed primer design for degenerate PCR (Flavell et al., 1992a, b; Hirochika et al., 1992; VanderWiel et al., 1993; Pearce et al., 1996, 1997). This technique allowed the cloning of these regions in numerous plant species and helped to unravel the evolutionary dynamics of LTR retrotransposon families within different plant lineages.

In the Poaceae family, the phylogenetic study of a large sample of *copia*-like *RT* sequences suggested an ancient origin of this type of retrotransposon (Gribbon et al., 1999; Matsuoka and Tsunewaki, 1999), as some retrotransposon supergroups (defined as sharing at least 60% sequence identity in the *RT* domain) were shared by all the Poaceae species analysed. In

addition, there appeared to be retrotransposons that are specific to plant families, tribes, and even species. This suggests that *copia*-like retrotransposons existed early in the angiosperm history and diverged into heterogeneous subgroups before the modern plant orders arose.

Similarly, the analysis of *RT* sequences from numerous plant species from both monocots and dicots has unravelled that at least four lineages of *gypsy*-like LTR retrotransposons coexisted in the last common ancestor of monocots and dicots, some 200 Mya (Marin and Llorèns, 2000), demonstrating that *gypsy*-like elements are also of ancient origin.

When the complete sequence of an LTR retrotransposon is available, a more detailed analysis of its transpositional history is possible. SanMiguel et al. (1998) have proposed a method to date the insertion of LTR retrotransposons in the maize genome, that they named "paleontology of retroelements". The dating is based on the estimate of the nucleotide divergence of the two LTR sequences of each retrotransposon, given the synonymous substitution rate of the *adh1* and *adh2* genes in the Poaceae family (6.5×10^{-9}, Gaut et al., 1996). As a consequence of the LTR retrotransposon replication cycle, the two LTRs of a newly inserted copy are indeed identical in sequence. Over time, mutations accumulate and lead to the divergence of the two LTRs, whose extent is proportional to the time elapsed since the insertion. Using an estimate of the divergence rate of these particular sequences, it is thus possible to translate the divergence of the two LTRs of a given element into an estimated insertion date.

The authors have shown that the retrotransposition of several families of LTR retrotransposons has been very active in maize over the past few million years (SanMiguel et al., 1998). Such a method has been consequently used to perform paleontological studies in other species such as *A. thaliana* (Wright and Voytas, 1998, 2001; Kapitonov and Jurka, 1999; Marin and Llorèns, 2000), barley (Wei et al., 1999), rice (Jiang et al., 2002a, b; McCarthy et al., 2002; Vitte and Panaud, 2003), tomato (Mao et al., 2001) and wheat (SanMiguel et al., 2002). They have revealed that most of the retrotransposon copies found within plant genomes have inserted within the last few million years, thus confirming that retrotransposons have been active in the recent history of flowering plants.

In addition, several LTR retrotransposons show polymorphic insertion patterns within domesticated variety pools of barley (Waugh et al., 1997; Vicient et al., 2001), pea (Ellis et al., 1998; Gribbon et al., 1999; Pearce et al., 2000) and of diverse Solanaceae species such as tobacco, tomato and pepper (Grandbastien, pers. comm.). This therefore suggests that these elements have amplified very recently, i.e. after the domestication of these species that occurred during the late Neolithic, some 10,000 years ago, even though care has to be taken, as recent results on pea have shown that the polymorphism observed within species did not reflect recent amplification but rather recombination or introgression (Vershinin et al., 2003). Paradoxically, for most of the LTR retrotransposon families that have been described in the literature, most of the copies retrieved from genomic sequences have been shown to be defective (due to stop-codons, insertions, deletions, frameshifts or rearrangements) and thus inactive.

Taken together, these observations suggest that a given LTR retrotransposon family can propagate through only a few functional copies, whereas most of the other copies are inactive. This hypothesis has been supported by the analysis of the phenetic relationships within three rice *gypsy*-like retrotransposon families (Vitte and Panaud, 2003). The authors have demonstrated that in all cases only a few master copies have indeed been at the origin of the copies present in the rice genome.

The next step towards understanding the evolutionary dynamics of LTR retrotransposon families in plant genomes is (i) to get an overview of the dynamics from an extended number of retrotransposon families within a genome and (ii) to compare the relative dynamics of a given retrotransposon family in several plant species. This will be possible in the near future, as complete genome projects will provide access to large databases of LTR retrotransposon sequences for several plant species.

LTR retrotransposons and plant genome size

LTR retrotransposons increase plant genome size and create genome size differences

LTR retrotransposons transpose via an mRNA intermediate and thus potentially increase their copy number in their host genome during their replication cycle. In the following section, we will examine how the activity of LTR retrotransposons could contribute significantly to genome size increases in plants.

The increase-only model

Over the last few years, extensive studies in the Poaceae family have provided good insight into the impact that LTR retrotransposons have on plant genome size. This family indeed offers a particular feature that makes it a good model for the analysis of genome evolution: during 60–70 My of evolution (i.e., since the origin of the family, Crepet and Feldman, 1991; Clark et al., 1995), the family has diversified into species the genomes of which vary greatly in size (from 0.5 pg/2C for *Oropetium thomaeum* to 27.6 pg/2C for *Lygeum spartum*), whereas gene content and gene order are conserved among them (Ahn and Tanksley, 1993). Such variations cannot merely be explained by differences in ploidy level or large duplications (there is, for instance, more than an 11-fold difference between the genome sizes of barley and rice, two diploid grasses).

In this family, the percentage of gene-containing regions is negatively correlated with genome size (they have been estimated to be 7, 12, 17 and 24% for wheat, barley, maize and rice, respectively) and the NTR-DNA portion has been shown to be proportional to genome size (Flavell et al., 1974). Hence, differences in genome size could be the result of size variations in the large heterochromatic gene-poor blocks, probably through retrotransposon amplifications. However, these observations did not rule out the possible impact of gene-dense intergenic regions on genome size evolution.

Large genomic sequences at orthologous positions are now available for several cereal species, thus allowing micro-colinearity analyses within the family and unravelling the evolu-

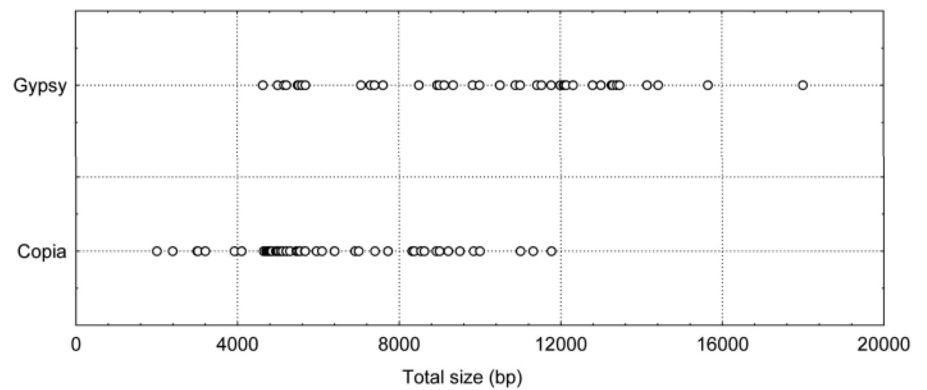

Fig. 1. Comparison of total length of *copia*-like and *gypsy*-like LTR retrotransposons. The total size of the retrotransposons is shown in bp. Each dot corresponds to one retrotransposon family. The original data is shown in Table 1.

tionary dynamics of the gene-dense regions. Results of these studies showed that, whereas gene content and gene order are still conserved to some extent at the micro-level in the Poaceae family, there is no correspondence between the TEs that comprise most of the intergenic regions (SanMiguel et al., 1996; Chen et al., 1997, 1998; Bennetzen et al., 1998). This observation was also made between two varieties of maize (Fu and Dooner, 2002), suggesting that the genomic differentiations induced by retrotransposons can be very rapid. In addition, retrotransposon content and genome size were positively correlated, which strongly suggests that LTR retrotransposons are the main factors contributing to variations in plant genome size. These observations led some authors to propose an increase-only model for the evolution of genome size in the Poaceae family (Bennetzen and Kellogg, 1997). This model posits that, within a genome, the retrotransposons undergo large amplification events, thus increasing genome size. Differences in genomes sizes would reflect different histories of retrotransposon amplification in distinct lineages.

Structural parameters determine the impact of a retrotransposon family on genome size

The total length of LTR retrotransposons observed so far in plants ranges from ~2,000 bp to ~18,000 bp. The *copia*-like elements range from ~2,000 bp to ~11,800 bp, whereas the *gypsy*-like elements range from ~4,650 bp to ~18,000 bp (see Fig. 1 for details of the distribution). This difference is statistically highly significant (Mann-Whitney test, $P < 0.000001$). In addition, the copy number of the retrotransposon families described so far in plants range from a few to several thousand copies (see Table 1 for details), depending on the family.

Consequently, it is now clear that all the LTR retrotransposon families of a given genome do not impact upon its evolution to the same extent. One has thus to take into account both the element size and its copy number as two key parameters for the overall contribution of retrotransposons to plant genome evolution. In addition, comparison of both the retrotransposon family diversity and the copy number per family will provide some insights on the relative differences that have led to genome size differentiations.

New insights from computer-based studies: non-autonomous retroelements are important factors in genome size increases

Over the past twenty years, several non-autonomous but active retrotransposons have been described, the best characterized being the maize elements *Bs1* (Johns et al., 1985, 1989; Jin and Bennetzen, 1989) and *Zeon* (Hu et al., 1995). *Bs1* has been characterized as a new insertion in the maize *Adh1* locus. It contains regions that are similar to the classical protease, RT, RNAseH and endonuclease regions of retroviral *pol* genes. However, these regions are split into sections, and therefore not functional. Similarly, the *Zeon* retrotransposon has been discovered as a new insertion within the 27-kDa γ *zein* locus whereas it lacks the open reading frame encoding the *RT* gene and is therefore not autonomous. Both *Bs1* and *Zeon* retrotransposons might thus have been trans-activated by their corresponding autonomous elements during their replication cycle. The precise evolutionary dynamics of these elements remain however still unclear.

Computer-assisted data mining of TEs has proven to be the most efficient method of analysing the evolutionary dynamics of LTR retrotransposon within a genome and thus their impact on genome evolution. Although many mined elements are typical of known groups (*copia* and *gypsy* types), new types of LTR-like retrotransposons have been described. These new elements, unlike the *Bs1* and *Zeon* elements, which show degenerate or partial *pol* regions, completely lack the gag/pol polyprotein that is necessary to undergo a retrotransposition cycle.

Terminal Repeats In Miniature (TRIM) elements possess the classical structure of LTR retrotransposons but are distinguished by their small overall size (> 540 bp), their small terminal repeats (> 140 bp on average) and the absence of the coding regions that are typical of LTR retrotransposons (Witte et al., 2001). TRIM elements have been described in several distinct plant lineages such as Poaceae, Brassicaceae, Fabaceae and Solanaceae. They are thus common to monocots and dicots and seem to have an ancient origin. However, copies from some TRIM families show high sequence identity (up to 98%) and the finding of related empty sites (Le et al., 2000) in *A. thaliana* suggests that at least some copies have inserted recently, most probably by trans-activation through autonomous elements that are yet to be discovered (Witte et al., 2001).

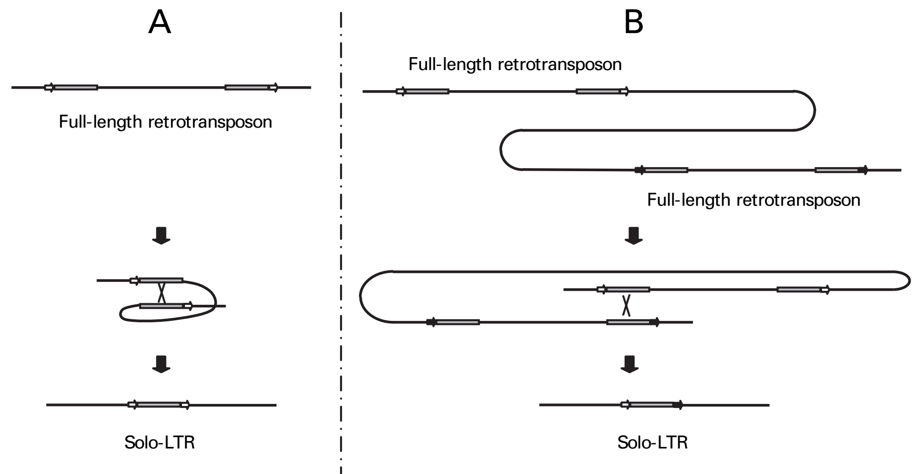

Fig. 2. Intra- and inter-element homologous unequal recombination: (**A**) Homologous recombination between two LTRs from the same retrotransposon copy: formation of a solo-LTR. In this case, one complete LTR (the solo-LTR) is remaining in the genome. The duplicated target site, represented by two white arrows, is conserved after the recombination. (**B**) Homologous recombination between two LTRs from two different retrotransposon copies. Such a recombination leads to the removal of a large fragment of DNA, comprising the internal region and one LTR of each copy involved in the recombination. In this case, the duplicated target site is not conserved (the final solo-LTR is flanked by a black and a white arrow). Figure adapted from Devos et al., 2002.

More recently, several other non-autonomous LTR elements have been found in the rice genome (*Dasheng*, Jiang et al., 2002a, b; *Spip* and *Squiq*, Vitte and Panaud, submitted results), whose LTRs share high sequence identity with known rice retrotransposons but which contain an internal region that is unrelated to the internal region of the known corresponding LTR retrotransposons. Specifically, these new elements completely lack the gag/pol polyprotein that is necessary to undergo a retrotransposition cycle. The conservation of structural features known to be important for the replication cycle such as LTR, PBS and Poly-Purine Tract (PPT) between the two protagonists of the autonomous/non-autonomous couples (*RIRE2/Dasheng*, Jiang et al., 2002a, b, *RIRE3/Spip* and *RIRE8/Squiq*, Vitte and Panaud, submitted results) suggests that these non-autonomous LTR retrotransposons are trans-activated by autonomous partners. Like TRIMs, they have amplified recently: for *Dasheng*, the sequence identity between the two LTRs of full-length elements (see SanMiguel et al., 1998 for detail of the dating method) ranges between 92.7 and 100%, with the majority of the elements (83%) showing >99.5% LTR sequence identity. Using a substitution rate of 6.5×10^{-9} (Gaut et al., 1996) most of the copies are thus estimated to have inserted within the last 500,000 years (Jiang et al., 2002a). For *Spip*, 35 of 37 copies show >99% LTR sequence identity. Using the substitution rate cited above, most of *Spip* copies are thus less than 800,000 years old (Vitte and Panaud, submitted results). *Squiq* shows the same trend, with 10 of 11 full-length copies (~90%) showing >99% LTR identity (Vitte and Panaud, submitted results).

This new type of LTR element has amplified recently in the rice genome, thus contributing to a recent genomic increase. In addition, the discovery of three non-autonomous families with classical LTR retrotransposon partners suggests that non-autonomous elements appeared several times in the evolution of rice. Comparing the amplification dynamics of such elements between distinct species will augment the insights given by coding LTR retrotransposon studies and further reveal the impact of LTR retrotransposons on plant genome evolution.

Genome size increases due to retrotransposon amplifications are counterbalanced by genomic contractions: towards an increase/decrease model of plant genome size evolution

One main argument, albeit indirect, in favour of the increase-only model of genome evolution was the lack of observed mechanisms for decreasing genome size. However, Southern blot analyses using probes from different regions (*RT*, *Int*, and LTR) of the *BARE-1* element showed that the ratio of LTR to internal region varied greatly among species both within the genus *Hordeum* (Vicient et al., 1999) and within wild barley (*H. spontaneum*, Kalendar et al., 2000) and was in both cases negatively correlated with the contribution of the *BARE-1* element to genome size (Vicient et al., 1999; Kalendar et al., 2000). These results suggest that, at least in barley and related species, a decreasing force exists, which tends to remove LTRs. In the following section, we present two decreasing processes, LTR-LTR recombination and deletion formation, and show how they can lead to a significant decrease in genome size. We then propose a molecular mechanism by which they could occur.

Elimination of LTR retrotransposons: solo-LTRs vs. deletions

Solo-LTRs are LTRs that are flanked by direct repeats (target site duplication) and are not connected to any internal region (Fig. 2). They are hypothesized to originate from recombination between the two LTRs of full-length elements, leading to the elimination of the internal region along with one LTR. Although such solo-LTRs have often been observed (Shepherd et al., 1984; Sentry and Smyth, 1989; SanMiguel et al., 1996; Noma et al., 1997; Chen et al., 1998; Han et al., 2000), only a few copies were generally found compared to the numerous full-length corresponding copies observed in the host genomes. This mechanism of elimination had therefore been considered to be too weak to counteract the genomic inflations caused by massive bursts of LTR retrotransposon amplification as had been described in maize (Bennetzen and Kellogg, 1997; SanMiguel et al., 1998). In barley, however, results from Shirasu et al.

(2000) suggest that both intra- and inter-element recombination events have occurred between LTR sequences around the *Rar1* locus. The authors thus proposed solo-LTR formation as a possible mechanism that could counteract LTR retrotransposon amplification. Indeed, whereas the formation of solo-LTRs by intra-element recombination could only reduce genome size to a small degree (one LTR from the complete element is still remaining), inter-element recombination is a mechanism that could eliminate complete copies and counteract massive amplifications of retrotransposons (Fig. 2). Considering an island-like distribution of genes in the genome, such large deletions should be transparent to selection and thus heritable.

In rice, the occurrence of solo-LTRs was analysed for three *gypsy*-like LTR retrotransposon families (Vitte and Panaud, 2003). The results showed that solo-LTRs are abundant in rice genome and that the relative abundance of solo-LTRs to complete copies varies from one retrotransposon family to the other (~1:1.77, ~1:10 and ~1:0.4 solo-LTRs:complete-copies ratios, for the *hopi*, *Retrosat1* and *RIRE3* families, respectively). In addition, whereas inter-element recombination has been described in barley (Shirasu et al., 2000), analysis of the duplicated target sites of these three LTR retrotransposon families in rice showed that most solo-LTR copies have originated from intra-element recombination events (Vitte and Panaud, 2003). The authors have also investigated the timing of solo-LTR formation relatively to full-length copies. The clustering of each solo-LTR with a group of complete copies suggested that solo-LTR formation is concomitant with retrotransposon amplification.

Solo-terminal repeats have also been described for both TRIMs and LTR non-autonomous elements (Witte et al., 2001; Jiang et al., 2002a, b). Most of these solo-LTRs also show conserved target site duplication and are thus considered to have resulted from intra-copy recombination.

Overall, these data show that, although solo-LTR formation seems to occur in several plant species and in both autonomous and non-autonomous elements, differences in LTR recombination do not seem to be the main mechanism that could explain large differences in genome size. In *A. thaliana,* the analysis of 291 LTR retrotransposons belonging to 12 families revealed 87 intact elements, 101 solo-LTRs, five elements showing traces of intra-element unequal recombination and 98 elements harbouring traces of illegitimate recombination leading to deletions (Devos et al., 2002). Thus, in this small plant genome, two mechanisms appear to have contributed to reducing the genome size: solo-LTR formation through unequal homologous recombination, and formation of deletions through illegitimate recombination. Deletions have also been shown to be a frequent event in maize TEs (Masson et al., 1987; Marillonnet and Wessler, 1998) and have been described for wheat LTR retroelements (Wicker et al., 2001, 2003), suggesting that formation of deletions by a mechanism independent of homologous recombination is key for DNA elimination in flowering plants. Differences in the efficiency of such a mechanism among plant genomes could thus explain (at least in part) differences in final genome size.

Such a bias towards deletions had been previously proposed as a force contributing to differences in genome size in *Drosophila:* Petrov and Hartl (1998) estimated that "Dead on arrival" copies from long interspersed nuclear elements (LINEs) lost 50% of their DNA in 14 Myrs through spontaneous deletion. In addition, results from previous studies (Petrov et al., 1996) demonstrated that mutations in *Drosophila* are biased towards deletions and that genome size difference between *Drosophila* and Laupala cricket could be explained by a difference in deletion rate (~40 times lower in the case of the 11-fold larger genome of Laupala, Petrov et al., 2000). The gradual removal of sequences through illegitimate recombination could thus be a counterbalancing force against genome expansions. Such illegitimate recombination could occur due to errors in DNA replication, double-strand break repair or other unknown mechanism.

Double-strand break repair as a molecular mechanism for deletion and solo-LTR formation

Recently, it was proposed that double-strand break repair, through homologous or illegitimate recombination, might influence genome size and organization (Kirik et al., 2000). Double-strand breaks were induced in both *A. thaliana* and tobacco (two dicotyledonous plants species differing >20-fold in genome size) by the rare-cutting I-*Sce*I restriction endonuclease and the size of the subsequent deletions were compared. The two types of mechanisms that have been postulated for eukaryotes (Nicolas et al., 1995), i.e. junctions without homologies (simple ligation) and single-strand annealing (associated with small patches of homologous nucleotides), were observed for the two plant species, and no significant difference of ratio between the two was seen (1.5 times more junctions with small homology patches than without in tobacco, vs. 1.35 in *Arabidopsis*).

However, significant differences in deletion size were observed. First, the average deletion size was larger in tobacco than in *Arabidopsis* (1,341 bp compared to 920 bp). Second, insertions of filler sequences were associated with 40% of linked-ends cases in tobacco, whereas such insertions were not observed in *Arabidopsis*. Interestingly, such insertions of filler sequences have also been reported to be associated with deletions in maize, another plant with a large genome (Wessler et al., 1990). Taken together, these observations suggest that differences in double-strand break repair mechanism might exist between species and that they may contribute to differences in genome size. In addition, the observation of small duplicated patches of nucleotides in deletions of both *Arabidopsis* (Devos et al., 2002) and *Drosophila* (Petrov and Hartl, 1998; Petrov et al., 2000) suggest that they could result from single strand annealing repair. Hence, differences in genome size could result from, at least in part, differences in the efficiency of this mechanism. The numerous deletions observed however imply that double-strand breaks might be frequent, though extremely damaging. One possibility is that such double-strand breaks could be induced by the excision of class II TEs or by abortive events of class I and class II TEs insertion. The occurrence of double-strand breaks would thus be linked to the global activity of TEs within a genome.

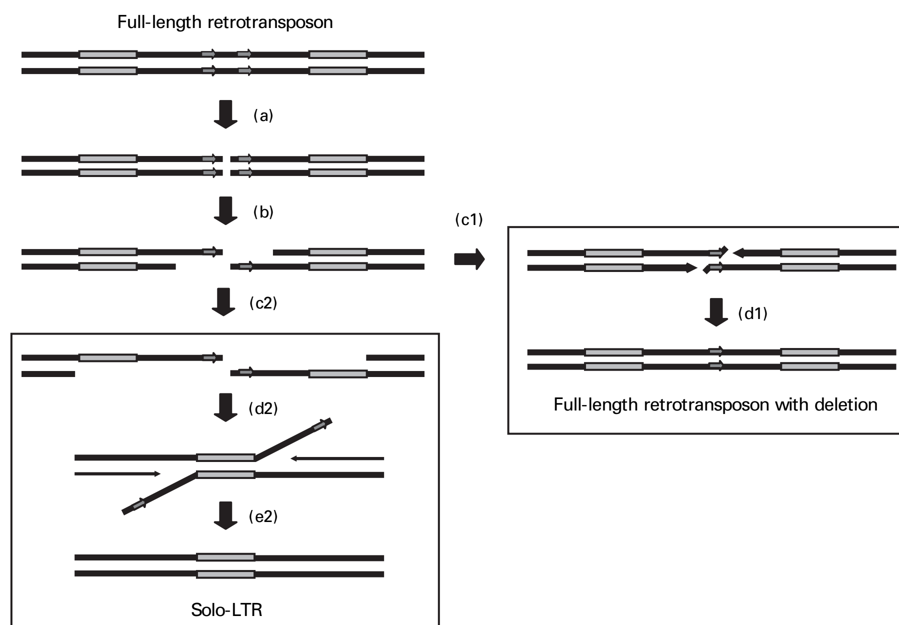

Fig. 3. Model of deletion and solo-LTR formation through double-strand break repair: Genomic DNA is represented by a black line. Full-length LTR retrotransposon is represented in grey, with two long boxes showing the LTRs. Small arrows correspond to small patches of identical nucleotide that have the same orientation. (a) Double strand break; (b) degradation of the single strands following the double strand break; (c1) annealing of the homologous small sequences single strands and elongation of the genomic DNA single strands; (d1) following of elongation and final ligation, leading to the formation of a deletion; (c2) the degradation of the single strands continues up to the LTR boundaries; (d2) annealing of the single strands LTR homologous sequences and elongation of the genomic DNA single strands; (e2) following of elongation and final ligation, leading to the formation of a solo-LTR.

The relative rates of solo-LTR formation versus deletion remain unknown. Some authors recently devised a system to study the occurrence of double-strand break repair through homologous recombination or non-homologous end joining (illegitimate recombination) in tobacco (Siebert and Puchta, 2002). Their results show that when genomic double-strand breaks are induced in close proximity to homologous sequences, these sequences can be used for repair through recombination in up to one third of the cases. A parallel can be made with LTR retrotransposons, with LTR sequences corresponding to the homologous sequences. Such double-strand breaks using LTRs would induce solo-LTR formation through recombination, whereas the remaining repair events would induce deletions through illegitimate recombination. Hence, under this model, the ratio between solo-LTR and deletions for a given double strand would be 1:3 in tobacco.

However, in this particular system, the double-strand break was induced directly next to the homologous sequences. This is not the case for retrotransposons, where the two LTRs of a complete copy can be more than 12 kb apart. In yeast, it has been demonstrated that the addition of a 4.4-kb interval between two repeats decreases the efficiency of homologous recombination by 3 fold as compared with directly adjacent repeats (Fishman-Lobell et al., 1992). It is thus possible that the deletions:solo-LTR ratio is even more biased toward deletions for retrotransposons, suggesting that solo-LTR formation is not a major force compared to deletion formation in plant genomes. Siebert and Puchta (2002) suggested that if homologous sequences are present in close proximity to the break, they will be used for the rejoining, leading to unequal homologous recombination. If no such sequences are available, short patches of identical nucleotides (which are numerous in plant genomes and in retrotransposons in particular) will be used instead, leading to deletion through illegitimate recombination (see Fig. 3 for mechanistic details). Hence, differences between retrotransposon families in terms of internal sequence length could induce differences in solo-LTR occurrence and the number of solo-LTRs in a given genome would also depend on the structural features of the retrotransposon families from which they are derived.

Conclusion

The current model of plant genome evolution

The large amount of data that has been accumulated for plant LTR retrotransposons during the past decades has revealed that plant genomes have both undergone genome amplifications (through retrotransposition) and contractions (through either homologous or illegitimate recombination). It is thus now clear that both these forces have an impact on plant genome size, leading to the emergence of an increase-decrease model for plant genome evolution (Fig. 4). This model posits that plant genomes have undergone (and still undergo) genomic expansions through the amplification of both autonomous and non-autonomous elements. As a result, a significant portion of the genome consists of DNA sequences that are non-genic. This portion, which is not under direct selection, is then eliminated. The rapid turnover of TE sequences through this process leads to a fast differentiation of plant genomes, which is evidenced by the lack of correspondence between the inter-genic regions of orthologous loci in various taxa (e.g., the Poaceae).

The next step towards defining a model for plant genome evolution is to determine precisely how these two antagonist forces actually drive overall changes in plant genomes, especially those changes related to genome size. So far, it is not yet clear whether differences in genome size are due to the relative extent of the amplification force, of the contraction force, or

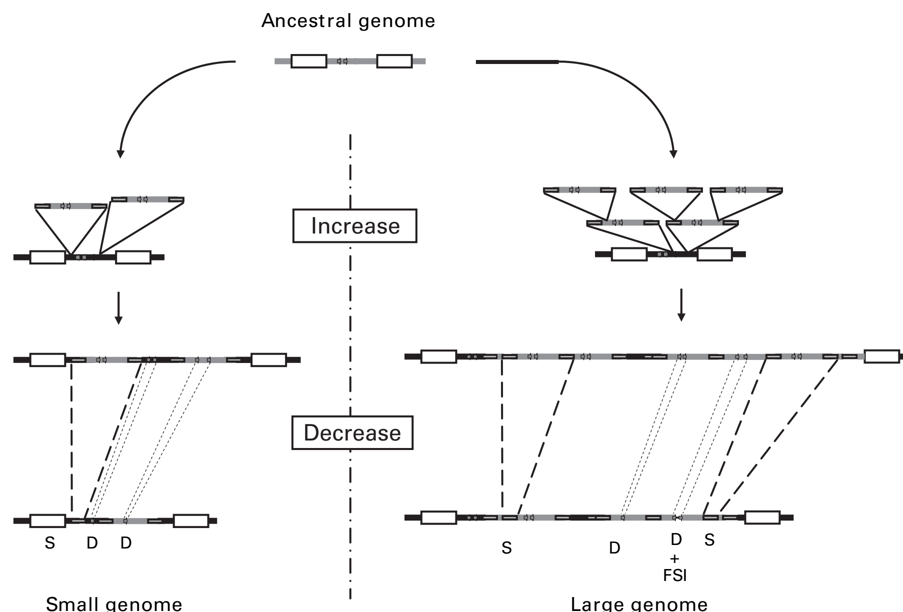

Fig. 4. Current increase-decrease model of plant genome evolution: White rectangles correspond to genes, genomic DNA is represented by a black line. Small patches of homologous nucleotides are shown by two small arrows. LTR retrotransposons are represented by a line boarded by two rectangles that figure the two LTRs. They contain arbitrarily one set of short patches. In the ancestral genome, genes are separated by a short genomic sequence, with short patches of homologous nucleotides (they might be more numerous, but only one is shown for the clarity of the scheme). Following retrotransposon amplification, a large amount of DNA is added in the genome, leading to a genome increase. On the left, only a small amount of copies have inserted, leading to a dispersed organization of the retrotransposons and a small genome increase. On the right, massive amplification of retrotransposons leads to a nested organization of the copies and a large genome increase. The portion brought by retrotransposon amplification is not under direct selection and starts to be eliminated, both through solo-LTR (S) and deletion (D) occurrence. On both left and right, solo-LTR and deletions occur. However, on the right part, deletions are smaller than in the left part, and accompanied by filler sequences insertion (FSI). Hence, the differences between a small and a large genome could be due to a difference of LTR retrotransposon amplification strength or to a difference in the contraction efficiency, or to both.

both. When comparing rice with Poaceae species having a larger genome (e.g., maize, barley or wheat), it appears that the species with large genomes have undergone retrotransposition bursts of larger extent.

The impact of the decreasing force on genome size is however less clear, as no direct evidence has yet been shown. Indirect evidence from the genomic paleontology of LTR retrotransposons from maize (SanMiguel et al., 1998), rice (Witte et al., 2001; Jiang et al., 2002a, b; McCarthy et al., 2002; Vitte and Panaud, 2003), *A. thaliana* (Wright and Voytas, 1998, 2001; Kapitonov and Jurka, 1999; Marin and Llorens, 2000), tomato (Mao et al., 2001) and wheat (SanMiguel et al., 2002) however suggests that this force may have had an impact on genome size: the data available clearly show that the vast majority of LTR retrotransposons found in the genomes of these species have inserted within the last few million years.

This can be interpreted in two alternative ways: either the genomes of these species were devoid of LTR retrotransposons only a few million years ago, or the elements that were present in their ancestral genomes have been eliminated during the same time period. Because LTR retrotransposons are found in all living organisms, it is almost impossible to believe that this type of TE, mostly inherited vertically through the reproduction of their host, could have colonized all the genomes in the plant kingdom in less than five million years. Thus, the hypothesis that old elements may indeed have been lost seems more plausible. Albeit no decrease has directly been estimated, these observations suggest that the elimination of LTR retrotransposons has resulted in significant decreases in genome size in some plant lineages.

However, detailed information about the extent and timing of the elimination process is still to be estimated in order to further complete our understanding of its impact on plant genome size. If this process is continuous and of limited extent, then the differences in genome size should essentially reflect the differences in the retrotranspositional activity. On the other hand, if TE elimination is efficient (that is, if it removes substantial numbers of DNA segments in a relatively short time period), then it should be regarded as an important factor of genome size decrease. In animals, significant differences in DNA loss have been shown between two insect species with different genome size. There is therefore a need to compare the relative efficiency of the decreasing force in several plants with different genome sizes. This will provide direct evidence on the impact of DNA loss in plants.

Future prospects

Origin of massive LTR retrotransposon bursts

As mentioned above, studies of the molecular paleontology of LTR retrotransposons suggest that, in plants, genomic expansions have occurred through concomitant bursts of several retrotransposon families in the recent past. Comparing the data obtained from rice, maize and *A. thaliana*, for which numerous retrotransposon insertions have been dated, there is a striking coincidence of multiple rapid amplifications around 1–1.5 million years ago. This raises the question of a possible cause, or common origin, of such concomitant events. It is now well documented that retrotransposition is induced by biotic and abiotic stresses (Wessler, 1996; Grandbastien, 1998, see also Melayah et al., 2001 for more recent data). The most recent episodes of global cooling took place during the Pleistocene (1.8 million – 10,000 years ago). Moreover, global demethylation, including demethylation of part of a retrotransposon sequence, was recently shown to be induced by cold stress in maize seeds (Steward et al., 2002). The alternating cycles of global cooling and warming during this period could certainly be regarded as an abiotic stress to which the ancestors of today's land plants were subjected. Whether such repeated stresses could be at the origin of massive retrotransposition events remains purely speculative, however it provides an attractive working hypothesis that could be tested further when additional data on genomic paleontology become available for a broader sample of plant taxa.

Extent and timing of the contraction force

The extent and timing of the elimination process remain largely unclear, simply because the size and structure of the ancestral genomes are unknown. However, comparative genomic studies of closely related species (that have diverged within the last few million years) should allow us to analyse orthologous LTR retrotransposons and therefore estimate the rate of the decreasing force in several lineages.

Moreover, computer-based large-scale analyses of numerous families of LTR retrotransposons will soon provide tools for studying the decay of the LTR retrotransposons that have transposed recently and may provide some insight on the rate and extent of the deletion process in rice (Vitte, unpublished results). As soon as such analyses will be handled on several plant species with distinct genome sizes, correlation estimates between genome size and contraction rate will be performed and provide direct evidence of the impact of the decreasing force on plant genome size.

Acknowledgements

We thank A. Sarr and A. Frary for their useful comments on the manuscript.

References

Ahn S, Tanksley SD: Comparative linkage maps of the rice and the maize genomes. Proc Natl Acad Sci USA 92:7980–7984 (1993).

Ananiev EV, Phillips RL, Rines HW: Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci USA 95:13073–13078 (1998).

Arabidopsis Genome Initiative (AGI): Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815 (2000).

Araujo PG, Casacuberta JM, Costa AP, Hashimoto RY, Grandbastien MA, Van Sluys MA: *Retrolyc1* subfamilies defined by different U3 LTR regulatory regions in the *Lycopersicon* genus. Mol Genet Genomics 266:35–41 (2001).

Balint-Kurti PJ, Clendennen SK, Dolezelova M, Valarik M, Dolezel J, Beetham PR, May GD: Identification and chromosomal localization of the *monkey* retrotransposon in *Musa* sp. Mol Gen Genet 263:908–915 (2000).

Barakat A, Matassi G, Bernardi G: Distribution of genes in the genome of *Arabidopsis thaliana* and its implications for the genome organization of plants. Proc Natl Acad Sci USA 95:10044–10049 (1998).

Bennetzen JL, Kellogg EA: Do plants have a one-way ticket to genomic obesity? Plant Cell 9:1509–1514 (1997).

Bennetzen JL, SanMiguel P, Chen M, Tikhonov A, Francki M, Avramova Z: Grass genomes. Proc Natl Acad Sci USA 95:1975–1978 (1998).

Bhattacharyya MK, Gonzales RA, Kraft M, Buzzell RI: A *copia*-like retrotransposon *Tgmr* closely linked to the *Rps1-k* allele that confers race-specific resistance of soybean to *Phytophthora sojae*. Plant Mol Biol 34:255–264 (1997).

Blume B, Barry CS, Hamilton AJ, Bouzayen M, Grierson D: Identification of transposon-like elements in non-coding regions of tomato ACC oxidase genes. Mol Gen Genet 254:297–303 (1997).

Brunner S, Keller B, Feuillet C: A large rearrangement involving genes and low-copy DNA interrupts the microcollinearity between rice and barley at the *Rph7* locus. Genetics 164:673–683 (2003).

Camirand A, Brisson N: The complete nucleotide sequence of the *Tst1* retrotransposon of potato. Nucleic Acids Res 18:4929 (1990).

Chavanne F, Zhang DX, Liaud MF, Cerff R: Structure and evolution of *Cyclops*: a novel giant retrotransposon of the *Ty3/Gypsy* family highly amplified in pea and other legume species. Plant Mol Biol 37: 363–375 (1998).

Chen M, Bennetzen JL: Sequence composition and organization in the *Sh2/A1*-homologous region of rice. Plant Mol Biol 32:999–1001 (1996).

Chen M, SanMiguel P, de Oliveira AC, Woo SS, Zhang H, Wing RA, and Bennetzen JL: Microcolinearity in *sh2*-homologous regions of the maize, rice and sorghum genomes. Proc Natl Acad Sci USA 94: 3431–3435 (1997).

Chen M, SanMiguel P, Bennetzen JL: Sequence organization and conservation in *sh2/a1*-homologous regions of sorghum and rice. Genetics 148:435–443 (1998).

Clark LG, Zhang W, Wendel JF: A phylogeny of the grass family (*Poaceae*) based on *ndhF* sequence data. Syst Bot 20: 436–460 (1995).

Costa AP, Scortecci KC, Hashimoto RY, Araujo PG, Grandbastien MA, Van Sluys MA: *Retrolycl-1*, a member of the *tntl* retrotransposon super-family in the *Lycopersicon peruvianum* genome. Genetica 107:65–72 (1999).

Crepet WL, Feldman GD: The earliest remains of grasses in the fossil record. J Botany 78:1010–1014 (1991).

Curtis CA, Lukaszewski AJ: Genetic linkage between C-bands and storage protein genes in chromosome 1B of tetraploid wheat. Theor Appl Genet 81:245–252 (1991).

Devos K, Brown JK, Bennetzen JL: Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079 (2002).

Dubcovsky J, Ramakrishna W, SanMiguel PJ, Busso CS, Yan L, Shiloff BA, Bennetzen JL: Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes. Plant Physiol 125: 1342–1353 (2001).

Ellis TH, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ: Polymorphism of insertion sites of *Ty1-copia* class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19 (1998).

Erdmann PM, Lee RK, Bassett MJ, McClean PE: A molecular marker tightly linked to *P*, a gene required for flower and seedcoat color in common bean (*Phaseolus vulgaris* L.), contains the *Ty3-gypsy* retrotransposon *Tpv3g*. Genome 45:728–736 (2002).

Faris JD, Haen KM, Gill BS: Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics 154:823–835 (2000).

Feuillet C, Keller B: High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA 96:8265–8270 (1999).

Fishman-Lobell J, Rudin N, Haber JE: Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. Mol Cell Biol 12:1292–1303 (1992).

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A: *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20:3639–3644 (1992a).

Flavell AJ, Smith DB, Kumar A: Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. Mol Gen Genet 231:233–242 (1992b).

Flavell RB, Bennett MD, Smith JB, Smith DB: Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 12:257–269 (1974).

Foster-Hartnett D, Mudge J, Larsen D, Danesh D, Yan H, Denny R, Penuela S, Young ND: Comparative genomic analysis of sequences sampled from a small region on soybean *(Glycine max)* molecular linkage group G. Genome 45:634–645 (2002).

Francki MG: Identification of *Bilby*, a diverged centromeric *Ty1-copia* retrotransposon family from cereal rye (*Secale cereale* L.). Genome 44: 266–274 (2001).

Fu H, Dooner HK: Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99:9573–9578 (2002).

Fu H, Park W, Yan X, Zheng Z, Shen B, Dooner HK: The highly recombinogenic *bz* locus lies in an unusually gene-rich region of the maize genome. Proc Natl Acad Sci USA 98:8903–8908 (2001).

Garber K, Bilic I, Pusch O, Tohme J, Bachmair A, Schweizer D, Jantsch V: The *Tpv2* family of retrotransposons of *Phaseolus vulgaris*: structure integration characteristics and use for genotype classification. Plant Mol Biol 39:797–807 (1999).

Gaut BS, Morton BR, McCaig BC, Clegg MT: Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA 93:10274–10279 (1996).

Georgi LL, Wang Y, Reighard GL, Mao L, Wing RA, Abbott AG: Comparison of peach and *Arabidopsis* genomic sequences: fragmentary conservation of gene neighborhoods. Genome 46:268–276 (2003).

Gill BS, Friebe B, Endo TR: Standard karyotype and nomenclature system for description of chromosome bands and structural aberration in wheat *(Triticum aestivum)*. Genome 34:830–839 (1991).

Grandbastien MA: Activation of plant retrotransposons under stress conditions. Trends Plant Sci 3:181–187 (1998).

Grandbastien MA, Spielmann A, Caboche M: *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337:376–380 (1989).

Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A, Flavell AJ: Phylogeny and transpositional activity of *Ty1-copia* group retrotransposons in cereal genomes. Mol Gen Genet 261:883–891 (1999).

Han CG, Frank MJ, Ohtsubo H, Ohtsubo E: New transposable elements identified as insertions in rice transposon *Tnr1*. Genes Genet Syst 75:69–77 (2000).

Harberd NP, Flavell RB, Thompson RD: Identification of a transposon-like element in a *Glu-1* allele of wheat. Mol Gen Genet 209:326–332 (1987).

Henikoff S, Comai L: A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. Genetics 149:307–318 (1998).

Hirochika H: Activation of tobacco retrotransposons during tissue culture. EMBO J 12:2521–2528 (1993).

Hirochika H, Fukuchi A, Kikuchi F: Retrotransposon families in rice. Mol Gen Genet 233:209–216 (1992).

Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M: Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93:7783–7788 (1996).

Hu W, Das OP, Messing J: Zeon-1, a member of a new maize retrotransposon family. Mol Gen Genet 248:471–480 (1995).

Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler SR: *Dasheng*: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. Genetics 161:1293–1305 (2002a).

Jiang N, Jordan IK, Wessler SR: *Dasheng* and *RIRE2*. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. Plant Physiol 130:1697–16975 (2002b).

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

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

Johns MA, Babcock MS, Fuerstenberg SM, Fuerstenberg SI, Freeling M: *Bs1*: an unusually compact retrotransposon in maize. Plant Mol Biol 12: 633–642 (1989).

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH: Genome evolution of wild barley *(Hordeum spontaneum)* by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97:6603–6607 (2000).

Kapitonov VV, Jurka J: Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107:27–37 (1999).

Keller B, Feuillet C: Colinearity and gene density in grass genomes. Trends Plant Sci 5:246–251 (2000).

Kentner EK, Arnold ML, Wessler SR: Characterization of high-copy-number retrotransposons from the large genomes of the Louisiana iris species and their use as molecular markers. Genetics 164:685–697 (2003)

Kirik A, Salomon S, Puchta H: Species-specific double-strand break repair and genome evolution in plants. EMBO J 19:5562–5566 (2000).

Konieczny A, Voytas DF, Cummings MP, Ausubel FM: A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics 127:801–809 (1991).

Ku HM, Vision T, Liu J, Tanksley SD: Comparing sequenced segments of the tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. Proc Natl Acad Sci USA 97:9121–9126 (2000).

Kumar A, Bennetzen JL: Plant retrotransposons. Annu Rev Genet 33:479–532 (1999).

Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E: Identification and characterization of novel retrotransposons of the *gypsy* type in rice. Mol Gen Genet 260:593–602 (1999).

Kumekawa N, Ohmido N, Fukui K, Ohtsubo E, Ohtsubo H: A new *gypsy*-type retrotransposon *RIRE7*: preferential insertion into the tandem repeat sequence *TrsD* in pericentromeric heterochromatin regions of rice chromosomes. Mol Genet Genomics 265:480–488 (2001).

Kunzel G, Korzun L, Meister A: Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154:397–412 (2000).

Kuwahara A, Kato A, Komeda Y: Isolation and characterization of *copia*-type retrotransposons in *Arabidopsis thaliana.* Gene 244:127–136 (2000).

Lall IP, Maneesha, Upadhyaya KC: *Panzee*, a *copia*-like retrotransposon from the grain legume pigeonpea (*Cajanus cajan* L.). Mol Genet Genomics 267:271–280 (2002).

Laten HM, Morris RO: SI retrotransposon-1, a long interspersed repetitive DNA element from soybean with weak sequence similarity to retrotransposons: initial characterization and partial sequence. Gene 134:153–159 (1993).

Laten HM, Majumdar A, Gaucher EA: SI retrotransposon-1 a *copia/Ty1*-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc Natl Acad Sci USA 95:6897–6902 (1998).

Le QH, Wright S, Yu Z, Bureau T: Transposon diversity in *Arabidopsis thaliana.* Proc Natl Acad Sci USA 97:7376–7381 (2000).

Lee D, Ellis THN, Turner L, Hellens RP, Cleary WG: A *copia*-like element in *Pisum* demonstrates the uses of disperse sequences in genetic analysis. Plant Mol Biol 15:707–722 (1990).

Li ZY, Chen SY, Zheng XW, Zhu LH: Identification and chromosomal localization of a transcriptionally active retrotransposon of *Ty3-gypsy* type in rice. Genome 43:404–408 (2000).

Linares C, Serna A, Fominaya A: Chromosomal organization of a sequence related to LTR-like elements of *Ty1-copia* retrotransposons in *Avena* species. Genome 42:706–713 (1999).

Linares C, Loarce Y, Serna A, Fominaya A: Isolation and characterization of two novel retrotransposons of the *Ty1-copia* group in oat genomes. Chromosoma 110:115–123 (2001).

Llaca V, Messing J: Amplicons of maize *zein* genes are conserved within genic but expanded and constricted in intergenic regions. Plant J 15:211–220 (1998).

Lucas H, Moore G, Murphy G, Flavell RB: A family of retrotransposons and associated genomic variation in wheat. Mol Biol Evol 9:716–728 (1992).

Manninen I, Schulman AH: *BA* retrotransposon-1, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). Plant Mol Biol 22:829–846 (1993).

Mao L, Begum D, Goff SA, Wing RA: Sequence and analysis of the tomato *JOINTLESS* locus. Plant Physiol 126:1331–1340 (2001).

Marillonnet S, Wessler SR: Extreme structural heterogeneity among the members of a maize retrotransposon family. Genetics 150:1245–1256 (1998).

Marin I, Llorèns C: *Ty3/Gypsy* retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. Mol Biol Evol 17:1040–1049 (2000).

Martinez-Izquierdo JA, Garcia-Martinez J, Vicient CM: What makes *Grande1* retrotransposon different? Genetica 100:15–28 (1997).

Masson P, Surosky R, Kingsbury JA, Fedoroff NV: Genetic and molecular analysis of the *Spm*-dependent *a-m2* alleles of the maize *a* locus. Genetics 117:117–137 (1987).

Matsuoka Y, Tsunewaki K: Evolutionary dynamics of *Ty1-copia* group retrotransposons in grass shown by reverse transcriptase domain analysis. Mol Biol Evol 16:208–217 (1999).

Mayer K, Murphy G, Tarchini R, Wambutt R, Volckaert G, Pohl T, Dusterhoft A, Stiekema W, Entian KD, Terryn N, et al: Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. Genome Res 11:1167–1174 (2001).

McCarthy EM, Liu J, Lizhi G, McDonald JF: Long terminal repeat retrotransposons of *Oryza sativa*. Genome Biol 3:10 research0053 (2002).

Melayah D, Bonnivard E, Chalhoub B, Audeon C, Grandbastien MA: The mobility of the tobacco *Tnt1* retrotransposon correlates with its transcriptional activation by fungal factors. Plant J 28:159–168 (2001).

Moore G, Lucas H, Batty N, Flavell R: A family of retrotransposons and associated genomic variation in wheat. Genomics 10:461–468 (1991).

Nakajima R, Noma K, Ohtsubo H, Ohtsubo E: Identification and characterization of two tandem repeat sequences (*TrsB* and *TrsC*) and a retrotransposon (*RIRE1*) as genome-general sequences in rice. Genes Genet Syst 71:373–382 (1996).

Nakamura Y, Kaneko T, Asamizu E, Kato T, Sato S, Tabata S: Structural analysis of a *Lotus japonicus* genome. II. Sequence features and mapping of sixty-five TAC clones which cover the 6.5 Mb regions of the genome. DNA Res 9:63–70 (2002).

Nakatsuka A, Iwami N, Matsumoto S, Itamura H, Yamagishi M: *Ty1-copia* group retrotransposons in persimmon (*Diospyros kaki* Thunb.). Genes Genet Syst 77:131–136 (2002).

Nicolas AL, Munz PL, Young CS: A modified single-strand annealing model best explains the joining of DNA double-strand breaks mammalian cells and cell extracts. Nucleic Acids Res 23:1036–1043 (1995).

Noma K, Nakajima R, Ohtsubo H, Ohtsubo E: *RIRE1*, a retrotransposon from wild rice *Oryza australiensis.* Genes Genet Syst 72:131–140 (1997).

Ohtsubo H, Kumekawa N, Ohtsubo E: *RIRE2*, a novel gypsy-type retrotransposon from rice. Genes Genet Syst 74:83–91 (1999).

Panaud O, Vitte C, Hivert J, Muszlak S, Talag J, Brar D, Sarr A: Characterization of transposable elements in the genome of rice (*Oryza sativa* L.) using Representational Difference Analysis. Mol Gen Genomics 268:113–121 (2002).

Panstruga R, Buschges R, Piffanelli P, Schulze-Lefert P: A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Res 26:1056–1062 (1998).

Pearce SR, Harrison G, Li D, Heslop-Harrison J, Kumar A, Flavell AJ: The *Ty1-copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250:305–315 (1996).

Pearce SR, Harrison G, Heslop-Harrison PJ, Flavell AJ, Kumar A: Characterization and genomic organization of *Ty1-copia* group retrotransposons in rye (*Secale cereale*). Genome 40:617–625 (1997).

Pearce SR, Knox M, Ellis TH, Flavell AJ, Kumar A: Pea *Ty1-copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Mol Gen Genet 263:898–907 (2000).

Peleman J, Cottyn B, Van Camp W, Van Montagu M, Inze D: Transient occurrence of extrachromosomal DNA of an *Arabidopsis thaliana* transposon-like element, *Tat1*. Proc Natl Acad Sci USA 88:3618–3622 (1991).

Pelissier T, Tutois S, Deragon JM, Tourmente S, Genestier S, Picard G: *Athila*, a new retroelement from *Arabidopsis thaliana*. Plant Mol Biol 29:441–452 (1995).

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

Pelsy F, Merdinoglu D: Complete sequence of *Tvv1*, a family of *Ty1 copia*-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking. Theor Appl Genet 105:614–621 (2002).

Petrov DA, Hartl DL: High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol Biol Evol 15:293–302 (1998).

Petrov DA, Lozovskaya ER, Hartl DL: High intrinsic rate of DNA loss in *Drosophila.* Nature 384:346–349 (1996).

Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL: Evidence for DNA loss as a determinant of genome size. Science 287:1060–1062 (2000).

Presting GG, Malysheva L, Fuchs J, Schubert I: A *Ty3/gypsy* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J 16:721–728 (1998).

Purugganan MD, Wessler SR: Molecular evolution of *magellan*, a maize *Ty3/gypsy*-like retrotransposon. Proc Natl Acad Sci USA 91:11674–11678 (1994).

Rahman S, Abrahams S, Abbott D, Mukai Y, Samuel M, Morell M, Appels R: A complex arrangement of genes at a starch branching enzyme I locus in the D-genome donor of wheat. Genome 40:465–474 (1997).

Rostoks N, Park YJ, Ramakrishna W, Ma J, Druka A, Shiloff BA, SanMiguel PJ, Jiang Z, Brueggeman R, Sandhu D, et al: Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Funct Integr Genomics 2:51–59 (2002).

Sandhu D, Gill KS: Gene-containing regions of wheat and the other grass genomes. Plant Physiol 128: 803–811 (2002).

Sandhu D, Champoux JA, Bondareva SN, Gill KS: Identification and physical localization of useful genes and markers to a major gene-rich region on wheat group 1S chromosomes. Genetics 157:1735–1747 (2001).

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, et al: Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768 (1996).

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL: The paleontology of intergene retrotransposons of maize. Nature Genet 20:43–45 (1998).

SanMiguel P, Ramakrishna W, Bennetzen JL, Busso CS, Dubcovsky J: Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). Funct Integr Genomics 2:70–80 (2002).

Sato S, Kaneko T, Nakamura Y, Asamizu E, Kato T, Tabata S: Structural analysis of a *Lotus japonicus* genome. I. Sequence features and mapping of fifty-six TAC clones which cover the 5.4 Mb regions of the genome. DNA Res 8:311–318 (2001).

Sentry JW, Smyth DR: An element with long terminal repeats and its variant arrangements in the genome of *Lilium henryi.* Mol Gen Genet 215:349–354 (1989).

Shepherd NS, Schwarz-Sommer Z, Blumberg J, Gupta M, Wienand U, Saedler H: Similarity of the *Cin1* repetitive family of *Zea mays* to eukaryotic transposable elements. Nature 307:185–187 (1984).

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P: A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915 (2000).

Siebert R, Puchta H: Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome. Plant Cell 14:1121–1131 (2002).

Smyth DR, Kalitsis P, Joseph JL, Sentry JW: Plant retrotransposon from *Lilium henryi* is related to *Ty3* of yeast and the *gypsy* group of *Drosophila*. Proc Natl Acad Sci USA 86:5015–5019 (1989).

Stergiou G, Katsiotis A, Hagidimitriou M, Loukas M: Genomic and chromosomal organization of *Ty1-copia*-like sequences in *Olea europaea* and evolutionary relationships of *Olea* retroelements. Theor Appl Genet 104:926–933 (2002).

Steward N, Ito M, Yamaguchi Y, Koizumi N, Sano H: Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. J Biol Chem 277:37741–37746 (2002).

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A: The complete sequence of 340 kb of DNA around the rice *ADH1-ADH2* regions reveals interrupted colinearity with maize chromosome 4. Plant Cell 12:381–391 (2000).

Terol J, Castillo MC, Bargues M, Perez-Alonso M, de Frutos R: Structural and evolutionary analysis of the *copia*-like elements in the *Arabidopsis thaliana* genome. Mol Biol Evol 18:882–892 (2001).

Thomson KG, Thomas JE, Dietzgen RG: Retrotransposon-like sequences integrated into the genome of pineapple, *Ananas comosus.* Plant Mol Biol 38:461–465 (1998).

Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein ND, Bennetzen JL, Avramova Z: Colinearity and its exceptions in orthologous *ADH* regions of maize and sorghum. Proc Natl Acad Sci USA 96:7409–7414 (1999).

Turcich MP, Mascarenhas JP: Pretrotransposon M-1, a putative maize retroelement has LTR (long terminal repeat) sequences that are preferentially transcribed in pollen. Sexual Plant Reprod 7:2–11 (1994).

Turcich MP, Bokhari-Riza A, Hamilton DA, He C, Messier W, Stewart CB, Mascarenhas JP: Pretrotransposon M-2, a *copia*-type retroelement in maize is expressed preferentially in early microspores. Sexual Plant Reprod 9:65–74 (1996).

van Leeuwen H, Monfort A, Zhang HB, Puigdomenech P: Identification and characterisation of a melon genomic region containing a resistance gene cluster from a constructed BAC library. Microcolinearity between *Cucumis melo* and *Arabidopsis thaliana*. Plant Mol Biol 51:703–718 (2003).

VanderWiel PL, Voytas DF, Wendel JF: Copia-like retrotransposable element evolution in diploid and polyploid cotton *(Gossypium L.)*. J Mol Evol 36:429–447 (1993).

Varagona MJ, Purugganan M, Wessler SR: Alternative splicing induced by insertion of retrotransposons into the maize *waxy* gene. Plant Cell 4:811–820 (1992).

Verriès C, Bes C, This P, Tesniere C: Cloning and characterization of Vine-1, an LTR-retrotransposon-like element in *Vitis vinifera* L and other *Vitis* species. Genome 43:366–376 (2000).

Vershinin AV, Ellis TH: Heterogeneity of the internal structure of PDR1, a family of Ty1/copia-like retrotransposons in pea. Mol Gen Genet 262:703–713 (1999).

Vershinin AV, Allnutt TR, Know MK, Amborse MJ, Ellis THN: Transposable elements reveal the impact of introgression rather than transposition in *Pisum* diversity, evolution and domestication. Mol Biol Evol 20:2067–2075 (2003).

Vicient CM, Martinez-Izquierdo JA: Discovery of a *Zdel* transposable element in *Zea* species as a consequence of a retrotransposon insertion. Gene 184:257–261 (1997).

Vicient CM, Suoniemi A, Anamthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH: Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum.* Plant Cell 11: 1769–1784 (1999).

Vicient CM, Kalendar R, Schulman AH: Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res 11:2041–2049 (2001).

Vignols F, Rigau J, Torres M, Capellades M, Puigdomenech P: The brown *midrib3* (*bm3*) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. Plant Cell 7:407–416 (1995).

Vitte C, Panaud O: Formation of Solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. Mol Biol Evol 20:528–540 (2003).

Voytas DF, Ausubel FM: A *copia*-like transposable element family in *Arabidopsis thaliana.* Nature 336: 242–244 (1988).

Wang S, Zhang Q, Maughan PJ, Saghai Maroof MA: *Copia*-like retrotransposons in rice: sequence heterogeneity, species distribution and chromosomal locations. Plant Mol Biol 33:1051–1058 (1997).

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

Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing RA, Leister D, Schulze-Lefert P, Wise RP: The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. Genetics 153:1929–1948 (1999).

Wei F, Wing RA, Wise RP: Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 14:1903–1917 (2002).

Wessler SR: Turned on by stress. Plant retrotransposons. Curr Biol 6:959–961 (1996).

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

White SE, Habera LF, Wessler SR: Retrotransposons in the flanking regions of normal plant genes: a role for *copia*-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91:11792–11796 (1994).

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B: Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26:307–316 (2001).

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B: Rapid genome divergence at orthologous low molecular weight *Glutenin* loci of the A and Am genomes of wheat. Plant Cell 15:1186–1197 (2003).

Witte CP, Le QH, Bureau T, Kumar A: Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 98:13778–13783 (2001).

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

Wright DA, Voytas DF: *Athila4* of *Arabidopsis* and *Calypso* of soybean define a lineage of endogeneous retroviruses. Genome Res 12: 122–131 (2001).

Young ND, Mudge J, Ellis TH: Legume genomes: more than peas in a pod. Curr Opin Plant Biol 6:199–204 (2003).