Theor Appl Genet (2002) 104:860–867
DOI 10.1007/s00122-001-0818-8

Z. Price · F. Dumortier · D.W. MacDonald · S. Mayes

Characterisation of *copia*-like retrotransposons in oil palm (*Elaeis guineensis* Jacq.)

Received: 19 December 2000 / Accepted: 5 June 2001 / Published online: 8 February 2002

Abstract The work aimed to isolate and characterise *copia*-like sequences from *Elaeis guineensis*. Thirty-two different RT (reverse transcriptase) sequences were isolated from a single oil palm genome using degenerate primers. Extreme sequence heterogeneity was observed. The DNA and protein sequences were assigned to three different classes (A, B, C) on the basis of bootstrapping. We estimated the copy number of the three different classes by using a dot-blot analysis. The comparative results suggest that class-B RT sequences occur at a higher copy number in *Cocos nucifera* than in *Elaeis guineensis* and *E. oleifera*. Class-C RT sequences, which comprise the bulk of isolated sequences, occur in much higher copy number in *Elaeis guineensis* than in *Cocos nucifera*, and *E. oleifera*. Class-A sequences, which have low copy number in *Elaeis guineensis* and *Cocos nucifera*, are not present in *E. oleifera*. Our preliminary results suggest that class-C sequences represent about 5% of the *E. guineensis* genome, class B, 1% and class A, 0.1%. The methylation status of genomic domains specified by the RT probes was analysed using two pairs of restriction enzymes, each pair having the same recognition sites but different methylation sensitivities (*Msp*I, *Hpa*II and *Sau*3AI, *Mbo*I). Results from these experiments showed clearly that the genomic domains specified by the RT probes are methylated. These also suggest that a higher copy number appears to correlate to a higher degree of methylation. Our preliminary results suggest that *copia*-like retrotransposons, because of their ubiquity and diversity, have great potential as genetic markers for plant genome and biodiversity analysis in *E. guineensis*.

Communicated by J.W. Snape

Z. Price (✉) · S. Mayes · D.W. MacDonald
Cambridge University, Department of Genetics, Downing Site, Downing Street, Cambridge CB2 3EH, UK
e-mail: z.price@gen.cam.ac.uk
Fax: +44-1223-333 992

F. Dumortier
DAMI, OPRS, New Britain Palm Oil Ltd., P.O. Box 165, Kimbe, West New Britain Province, Papua New Guinea

Keywords *Elaeis guineensis* Jacq. · Retrotransposons · Ty-*copia* · Molecular markers

Introduction

The oil palm (*Elaeis guineensis* Jacq.) belongs to the subfamily *Arecoideae*, tribe *Cocoeae* and subtribe *Elaeidinae*. This subtribe includes only the genera, *Elaeis* (from the greek *elaia*, for the olive tree) and *Barcella*. The genus *Barcella* has no current commercial use. The genus *Elaeis* consists of only two species: the African oil palm, *Elaeis guineensis* Jacq., and the Latin American oil palm, *Elaeis oleifera* Cortez. They are believed to have diverged 60 M years ago and are cross-fertile. The oil palm is monoecious, alternately producing male and female inflorescences on average every 6 months, and thus outcrossing. Detailed studies of the flowers have shown, however, that each flower primordium is a potential producer of both female and male organs, though one or the other almost always remains rudimentary. The seed (kernel) and the pulp (mesocarp) of the fruit are very rich in oil, which is the main commercial product of the oil palm. The oil is extracted from the kernel (kernel oil) and from the mesocarp (palm oil). Palm oil contains an equal proportion of saturated and unsaturated fatty acids (Hartley 1988).

Breeding and selection of *E. guineensis* began in the early 1920s, and since then considerable improvements have been made both in yield and quality characters (Rosenquist 1985; Hartley 1988). Ghesquiere (1984, 1985) assessed the enzyme polymorphism within a number of breeding origins of oil palm. The long breeding cycle (over 10 years) and the extensive variation still encountered suggests that there exists considerable scope for improvement in yield (Corley 1983), disease resistance (Flood et al. 1989) and oil composition. By using marker-assisted selection this process could be speeded up. Some of the important areas of interest are shell-thickness (Mayes et al. 1997; Moretzsohn et al. 2000), *virescens* and crown disease (Breure and Soebagjo

1991). A marker for shell thickness could have a potential commercial value in breeding programmes by predicting shell type long before the plants start fruiting. Jack et al. (1995) reported the potential in oil palm for marker identification and application and Mayes et al. (1995) on the use of highly informative oil palm restriction fragment length polymorphism (RFLP) markers for genotype characterisation. Shah (1994) assessed the utility of random amplified polymorphic DNA (RAPD) markers for determining genetic variation in oil palm, and Moretzsohn (2000) produced a RAPD linkage map of the shell thickness locus in oil palm. Mayes et al. (1997) reported the construction of a RFLP map for oil palm and subsequent identification of a marker linked to shell thickness.

Single-copy genes are embedded in a complex array of tandem and dispersed types of repetitive DNA, most of which have no function. Eucaryotic retrotransposons can be divided into long terminal repeat (LTR)-containing retrotransposons and poly-A-containing retroposons based on structural characteristics (Boecke and Corces 1989; Hull 1999; Kumar and Bennetzen 1999 for reviews). Although retrotransposons have played a major role in the expansion of plant genomes (in the maize genome, retrotransposons comprise 50–80%); (San Miguel and Bennetzen 1998), it is also likely that many retrotransposon sequences have been eliminated by various mechanisms such as unequal recombination between LTRs or other types of deletions (Kumar and Bennetzen 1999). Many plant retrotransposons have mutations or deletions – they are essentially "fossil remains" of earlier transposition events. However, many other plant retrotransposons studied to date may be functionally activated by various biotic and abiotic stress factors such as cell culture (Hirochika 1993; Grandbastien 1998). Retrotransposons have been shown to be able to generate mutations through insertions near or within genes (Wessler et al. 1995; Kuncze et al. 1997). *Copia*-like retrotransposons are ubiquitous in plant genomes (Flavell 1992; Voytas et al. 1992; Kubis et al. 1998), often present in high copy number, widely dispersed on chromosomes and widely distributed within the euchromatin domains of chromosomes (Pearce et al. 1997; Linares et al. 1999; Castilho et al. 2000). Because of this ubiquity and diversity, *copia*-like retrotransposons have great potential as genetic markers for plant genome and biodiversity analysis (Lee et al. 1990; Kumar et al. 1997; Waugh et al. 1997; Ellis et al. 1998; Kalendar et al. 1999; Pearce et al. 2000).

Materials and methods

Plant material

Samples of *Elaeis guineensis* from AVROS (DM742-207), Lobe (T250/29-07 T243/25-25), Binga (T227/51-19, T227/45-20) and Dami dura origins (DM711-714) (see Rosenquist 1985), *E. oleifera* and *Cocos nucifera* were provided and dispatched by NBPOL, Papua New Guinea. The plant material was foil-packed and stored at –20 °C.

Molecular analysis

Plant DNA was isolated by using the method of Dellaporta et al. (1983). A polymerase chain reaction (PCR) analysis was carried out using degenerate primers as described in Flavell (1992) in order to amplify a 280-bp band corresponding to the reverse transcriptase domain (RT) of *copia*-like retrotransposon. The downstream priming region is well-conserved between Ty1-*copia* group members, while the upstream priming region is less conserved and absent from other retroelements. The DNA template was isolated from the palm-DM 742.207 of the AVROS origin. The annealing temperature was 45 °C. Gel pieces containing DNA fragments of the expected size of 280 bp were excised and cleaned up using the Concert rapid gel extraction kit (Gibco-BRL Gaithersburg, Md.). The purified fragments were blunt-ended with T4 DNA polymerase, phosphorylated with T4 polynucleotide kinase and cloned into [Alkaline Phosphatase (Epicentre) (treated with shrimp)] Bluescript II SK + phagemid (Stratagene) *Eco*RV cloning site. The ligated products were transformed by electroporation into *Escherichia coli* DH 10B competent cells and plated onto the LA plates with appropriate selection. The positive colonies were amplified using flanking M13 primers, gel-purified and sequenced on an ABI 377 automated sequencer. Thirty-two sequences were obtained.

Phylogenetic analysis of RT sequences

All of the putative RT DNA sequences were scanned through an NCBI database and the DNA data bank of Japan by performing BLAST and FASTA similarity searches. The multiple nucleotide and amino acid sequence alignments were conducted with the programme MULTALIN (Corpet 1988). The sequences were translated into proteins using the Protein Engine available at EBI, Hinxton. Bootstrapping is an easy but very powerful method to assess the reliabilities of phylogenetic trees (Felsenstein 1985; Marshal 1991). Bootstrap tests with 100 bootstrap replicates were performed using the PAUP 4.0b4a version for Macintosh, and the DNA and protein sequences were assigned to groups on the basis of bootstrapping. The bootstrap percentage quoted for a group is the number of trees in 100 replicates in which all members of that group appeared. If all members appear at least 95% of the time, the grouping is considered to be statistically significant.

Southern analysis

Genomic DNA was digested with a range of restriction enzymes, fractionated on 1% agarose gels and transferred to Hybond-N (Amersham) membranes. Hybridisation and washing were performed according to standard protocols (Church and Gilbert 1984). Putative RT clones and oil palm genomic DNA were used as probes. Probes were labelled with [^{32} P]-dCTP (Amersham) by random priming (Feinberg and Vogelstein 1983). The washing conditions were as follows: two washes with 2 × SSC, 0.1% SDS at room temperature, 30 min each; one wash with 0.1 SSC, 0.1% SDS at room temperature, (30 min; one wash with 0.1 SSC, 0.1% SDS at 65 °C, 30 min).

Determination of copy number

In order to determine the relative copy number of putative RT sequences, 50 ng of a subset of RT sequences belonging to different classes was separated on a 1% agarose gel, transferred by Southern blotting to Hybond-N (Amersham) membrane and probed with the original genomic DNA (200 ng) from which these products were obtained (DM 742-207).

A dot-blot analysis was also used in order to determine the copy number (Suoniemi et al. 1996; Li et al. 1999; Linares et al. 1999; Vicient et al. 1999). Samples of *E. guineensis* from AVROS (DM 742-207), Lobe (T243/25-25), Binga (T227/45-20) and Dami dura origins (DM711-714) were used in this analysis (Rosenquist

```
     1                                                                                        90
  5  XXAFLHGHLE EKIYMEQPPS FRDPGSEGKV CLLQKSLYGL KQSPRQWYKR FDSHVRSIGL FSCEFDPCVY VQSLEDGSRV FLLLYVDDMX
 32  .......... .R...DK..G .........I .......... ...S.....Q ...Y...... .R........ .......... ..........
 34  .......... .R...DK..G .........I .......... ...S.....Q ...Y...... .R........ .......... ..........
  7  ....F..... .R........ .......... .......... ...S...... ...Y...... SR........ .......... ..........
 30  ....F..... .R...K...G .......... .......... .......... ...Y.H.... SK........ .........L ..........
 11  .......... .R.......G S......... .......... ...S...... ...Y.H.... SR........ ......S..L ..........
  8  .......... .R..I....D .........I .......... .......... ...Y.H.... SRY....... .........L ..........
  9  .......... .R..I....D .........I .......... .......... ...Y.H.... SRY....... .........L ..........
 14  .......... .R.......D ....R..... .......... .......... ...Y.C.... SR........ .........L ..........
 16  .......... .R........ .....L.E.. .....F...M .......... ...YMH.... .R........ .......... ..........
  6  ....F..... .R........ ....E..... .......... ........RQ .N.Y.H.... SR.....Y.. .......... ..........
 28  .......... .R........ ....E..... .......... .....*.... ...Y...... SR........ .........L ..........
23r  ....F..... .R.......Y .......... .....L.... .........W ...Y.X.... SR........ I...K..... ..........
 21  ....F..... .R.......N ..N....... .......... .......... ...Y...... SR........ .*........ ..........
 33  .......... .W.......D .......... .......... .........W ...Y....E. .R........ .....ND..L ..........
 15  .......... .R.....L.. ....R...M. .......... ...S.....W ...Y...... SR.....Y.. .......... ..........
 17  .......... .R.......D ....R..... .....L.... ...S.....W ...Y...... SR....L... .......... ..........
 20  .......... .R........ ....R...N. .......... .*.S...... ...Y...... .R........ .......... ..........
 18  .......... .R......SD ..N.....N. ........V. ...S.....W ...Y.C.... .R*....... .......... ..........
 12  .......... .R.......D ........M. ...*...... ...S.....Q ..FY...... SRY....Y.. ........ML ..........
 24  .......Y.. .R.......D .......... .......... .*...*...W ...Y...... SR....L... .*.......L ..........
 19  ....F..... .R....K..N ...L.....I .......... .........Q ...Y.CC... .R........ .*........ ..XXXXXXX.
 10  .......E.. .Q...Q..EG .QVEEK.DH. ...K...... ........R. ..TFMTQH.Y SRSSL.S... FRK.SGN.VI Y......X..
  4  .......D.. .E..IA..K. .VEK.KKEL. .*.NI..... ..TL.C...* ..FFIV.LRF DRL.V.H.A. FC.YD...FC I.........
 22  .......NV. QE...AL-EG .VKKSK.EL. YR.D....S. ..VT.YYKQ- ...FIV.LKF DRL.V.H.AC SSGYD.E.FY I......X..
  3  .......D.. K......SEG .KVK.K.RL. .W.R...... ..AL.*Q..* ...FMVDH.Y KKILS.H... ..KFD..DFI I.S.......
 31  ....F..D.. .E...MLEE- .EIKDKKNLM .H.K.N.... .ALRW-YK.- .N.FMMEN.Y KRTS..H.MF MKKFFEDKFN I.........
  2  .......E.T KEV..I..KE .TSTDES-.. WK..R.I... ..AS.S.NMH ..KVIKMY.F IKNREES.IC -K*ANGPVV. ..V.......
 13  .......E.T .EV..I..EG .TSTDES-.. .K.*R.I... ..AS.S.NMH ..KVIKMY.F VKNGEE..I. -KWANGLVV. .FV.......
 26  .......E.T .EV..I..EG .TSIDES-.. .K..R.I... ..AS.S.NMC ..KVIKMYDF VKNGEE..I. -KWANGPVV. ..I.......
 27  .......E.T .EV..I..EG .TSTDES-.. YK..R.I... .*ASQS.NMC ..KVIKTY.F VKNGEE..I. -KWVN.PVM. ..V.......
 35  .......... .D......LG .TSSDDDH.. .K..R.I... ..AS.S.NT. .NDVIKMF.F IKN.EE...F -KKVSGSVV. ..V.......
```

Fig. 1 Multiple alignment (Corpet 1988) of 32 RT DNA sequences translated into amino acids. *Asterisk* (*) represents stop-codons, *dashes* (–) are gaps inserted to optimise the alignment. Multiple alignment was performed by using Multalin (Corpet 1988) and adjusted manually. (*x*) represents missing data

1985). Serial dilutions of total plasmid DNA (20, 15, 10, 5, 0.05, 0.01, 0.005 ng) containing RT clones belonging to the three different classes (A, B, C) and the total genomic DNA of *E.oleifera guinensis*, *E. oleifera* and *C. nucifera* DNA (0.5, 0.4, 0.3, 0.2, 0.1, 0.05 ug) were denatured and spotted onto a Hybond N (Amersham) membrane. The RT inserts were used as probes: 31 and 32 (class A), 20 and 21 (class C), 35 and 27 (class B). The hybridisation and washing conditions were the same as those used for the genomic Southern analysis. It was assumed that the dot-blot hybridisation spots with the same intensities contained the same molar amounts of the target repeat sequence. The size of the haploid *E. guineensis* genome has been estimated to be 1.7×10^9 bp (Rival et al. 1997).

Investigation of RT methylation state

The methylation status of genomic domains specified by the RT probes was analysed by using two pairs of restriction enzymes *Msp*I/*Hpa*II (CCGG) and *Sau3*AI/*Mbo*I (GATC), each with the same recognition site but differing in inhibition by methylation. *Hpa*II wil not cleave ^{m5}CCGG, C^{m5}CGG, ^{m4}CCGG, C^{m4}CGG and ^{hm5}C^{hm5}CGG, whereas *Msp*I will cut C^{m5}CGG, C^{m4}CGG, ^{m4}CCGG but not ^{m5}CCGG and ^{hm5}C^{hm5}CGG (^{m5}C stands for 5-methylcytosine, ^{m4}C for 4-methylcytosine and ^{hm5}C for 5-hydroxymethylcytosine); (McClelland et al. 1994). *Sau*3AI does not cut GAT^{m5}C, GAT^{m4}C and GAT^{hm5}C; *Mbo*I, on the other hand, cleaves GAT^{m5}C and GAT^{m4}C but does not cleave GAT^{hm5}C (McClelland et al. 1994). Samples of *E. guineensis* from AVROS (DM 742-207), Lobe (T250/29-07) and Binga (T227/51-19) origins were used in this analysis (see Rosenquist 1985).

Results

RT sequences

PCR amplification with degenerate primers (Flavell 1992) yielded a band of the expected size (approximately 280 bp). There were 32 putative RT fragments sequenced, and their sequences were used to search the NCBI database and the DNA data bank of Japan by performing BLAST and FASTA searches. All of them correspond to RT domains of *copia*-like retrotransposons from *Elaeis guineensis*: (emb|AJ271986.1|EGU271986), (emb|AJ271987.1|EGU271987), (emb|AJ271988.1|EGU271988). There were a number of other positive hits, including: *Cocos nucifera* (gi|22680|emb|X70875.1), (gi|22679|emb|X70874.1), (gi|22678|emb|X70873.1); *Oryza australiensis* (D85597|D85597); *Zea mays* (D12830|MZEPOL1); *Triticum aestivum* (D12832|WHTPOL); *Hordeum vulgare* (Z17327|HVBARE1); *Platanus occidentalis* (M94486|PTNCOPIAB); *Helianthus annuus* (M94494|HNN COPIA).

Heterogeneity among RT sequences from a single plant

Extreme diversity amongst RT genes of the Ty1-*copia* retrotransposon within a single plant and between plant species has been observed (Konieczny et al. 1991; Flavell 1992). We found a similar pattern in oil palm RT sequences isolated from one individual. Multiple nucleotide and amino acid sequence alignments with the programme MULTALIN (Corpet 1988) were performed (Fig. 1). The DNA and translated protein sequences were

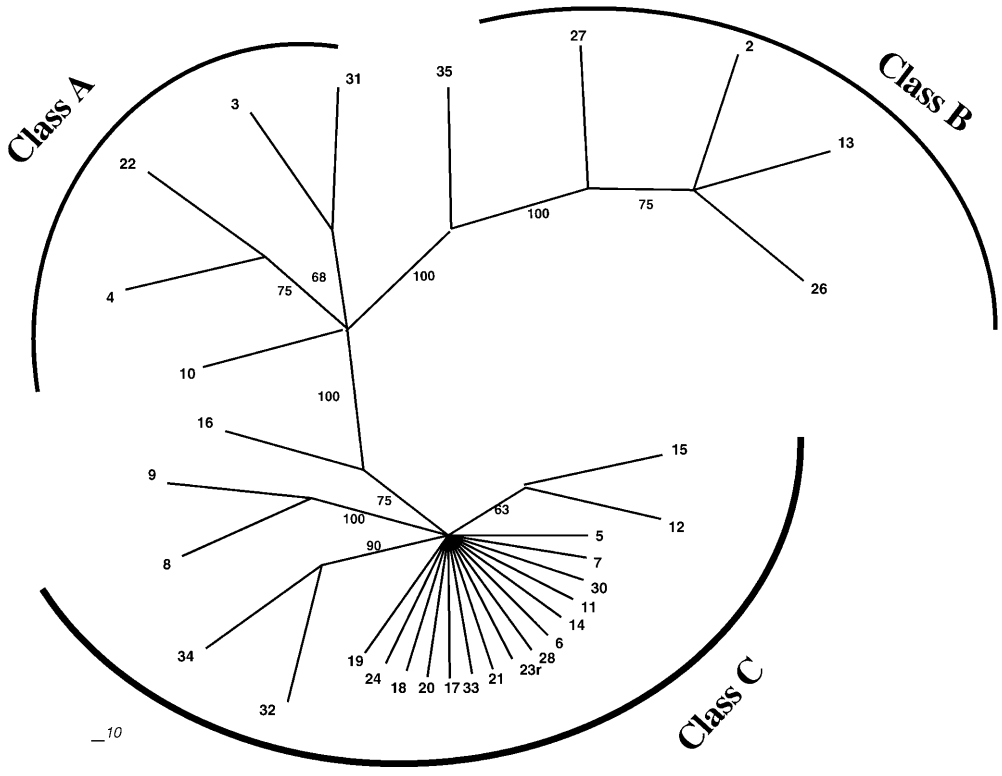

Fig. 2 Phylogram of protein sequences (PAUP, 4.0b4a version for Macintosh). The *numbers* at the *ends* of the *branches* indicate individual RT DNA sequences, translated into proteins which were obtained from a single plant (DM742-207). The *numbers on* the *branches* indicate bootstrap values for the support of the branches. The branches with bootstrap values *under 50* have been collapsed. The bootstrap percentage quoted for a group is the number of trees in 100 replicates in which all members of that group appeared. If all members appear at least 95% of the time, the grouping was considered statistically significant

assigned to groups on the basis of bootstrapping (Fig. 2). Bootstrapping RT DNA sequences and translated protein grouped the sequences into three different classes (A, B and C). Class A contains RT sequences 3, 4, 10, 22 and 31; class B, sequences 2, 13, 26, 27 and 35; class C sequences 5–9, 11, 12, 14–21, 23r, 24, 28, 30, 32–34. A phylogram of amino acid sequences revealed that classes B and C each have only one subclass: 35 and 8, 9, respectively. Although the clones 8 and 9 generate identical amino acid sequences they differ at the nucleotide level. Analysis of the unrooted phylogram of amino acid sequences showed that class B is more similar to class A than it is to class C. Translated amino acids and nucleotide sequences gave similar groupings (data not shown). We estimated the nucleotide divergence of the RT clones of the three classes (A, B, C) (Table 1). Class A shows quite a broad nucleotide divergence (0.19–0.37) and has about three times as many stop codons as the other two classes. Class B and C have quite a narrow nucleotide divergence (0.04–0.09).

Relative copy number of RT sequences

We estimated the relative copy number of different putative *copia* RT DNA sequences. The hybridisation experiment supported the segregation of putative *copia* classes on the basis of the phylogenetic analysis into three classes (A, B, C). The results suggest that class C, the one which contains the bulk of the RT sequences, has the highest relative copy number and that the other classes (A and B) have a medium to low copy number in *E. guineensis* (see Fig. 3a). Class B, which contains RT sequences 2, 13, 26, 27 and 35, was the only class which hybridizised to both *Elaeis oleifera* and *Cocos nucifera* (Fig. 3b,c) at a very high stringency (0.1 SSC, 0.1% SDS at 65 °C). RT sequence 35, despite being significantly different from RT sequences 2, 13, 26 on the basis of PAUP analysis (Fig. 2), was assigned to this group (class B) because it hybridizes to *Elaeis oleifera* and *Cocos nucifera*. The results suggest that class A and C RT-sequences occur only in *Elaeis guineensis* (Fig. 3a), while class B sequences occur in *E. guineensis*, *E. oleifera* and *Cocos nucifera*. As far as their generic relationship is concerned coconut (*C. nucifera*) belongs to the same family (*Arecaceae*) and the same tribe (*Cocoeae*) as oil palm (*E. guinnesis*). *E. oleifera* and *E. guineensis* belong to the same genera (*Elaeis*).

Dot-blot analysis

Serial dilutions were made of total plasmid DNA containing RT clones belonging to the three different classes and of total oil palm genomic DNA samples of different origin as well as of coconut and *E. oleifera*, in order to determine the copy number of the RT classes by dot-blot hybridisation (Yaňez et al. 1998; Vicient et al. 1999). The following recovered RT inserts were used as probes: 31 and 22 (class A), 20 and 21 (class C), 35 and 27 (class B) (Table 2). The hybridisations were repeated at least twice for each probe. Probes from the same class gave very similar results compared to the other classes. Under very stringent conditions we were able to detect hybridisation of *Elaeis oleifera* sequences to class C se-

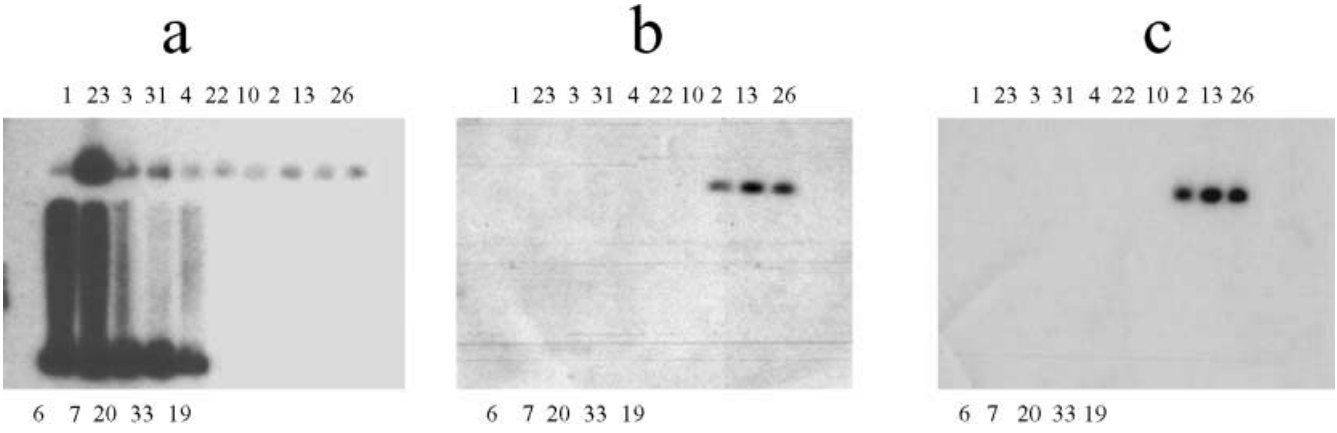

Fig. 3 Southern autoradiograph of a subset of RT-sequences (50 ng) belonging to three different classes – class A (1, 3, 31, 4, 22); class B (2, 13, 26); class C (6, 7, 20, 33, 19) – separated on a 1% agarose gel, transferred by Southern blotting to a Hybond-N (Amersham) membrane and probed with genomic DNA (200 ng) of: **a** *Elaeis guineensis* (DM742-207), **b** *E. oleifera*, **c** *Cocos nucifera*. The washing conditions were stringent: two 30-min washes with 2 × SSC, 1% SDS, one 30-min wash with 0.1 SSC, 1% SDS at room temperature and one 30-min wash with 0.1 SSC, 1% SDS at 65 °C

Table 1 Nucleotide divergence of the RT clones divided into three classes (A, B, C). Nucleotide divergence was calculated as the proportion of nucleotide sites at which the two sequences compared differ. The numbers in brackets are for clone 35, which belongs to the class B based on the PAUP analysis

Classes	Number of clones	A		B		C	
		Average	Range	Average	Range	Average	Range
A	5	0.20	0.19–0.37	0.4	0.39–0.44	0.41	0.41–0.50
B	5	–	–	0.06 (0.22)	0.05–0.09 (0.23)	0.45	0.44–0.52
C	22	–	–	–	–	0.06	0.04–0.09

Table 2 The copy number of RT DNA classes in *Elaeis guineensis*, *E. oleifera* and *Cocos nucifera*. Samples of *E. guineensis* from AVROS (DM 742-207), Lobe (T243/25-25), Binga (T227/45-20) and Dami dura origins (DM711-714) (Rosenquist, 1985) were used in this analysis

	E. guineensis (% of the genome)				*E.oleifera* (% of the genome)	*C. nucifera* (% of the genome)
	AVROS DM742-207	Lobe T243/25-25	Binga T227/45-20	DAMI dura DM711-714		
Class A	0.01	0.01	0.01	0.01	0	<0.01
Class B	0.08	0.08	0.08	0.08	0.08	0.4
Class C	0.2	0.2	0.2	0.2	0.01	0.01

quences and the hybridisation of *Cocos nucifera* sequences to those of class A and B, which was not apparent from Southern analysis of RT sequences. We did not detect hybridisation of *E. oleifera* DNA to the class A sequences (Table 2). We estimated that class A represents about 0.01% of the *E. guineensis* genome; class B, 0.08% and class C about 0.2%, based on the representation of RT sequences. On the basis of sequence data so far (RT-LTR), we would expect the complete *copia*-like elements in oil palm to be about 5–6 kb (unpublished data, Z. Price). Assuming that RT sequences are part of a complete *copia*-like retrotransposon with an estimated length of 5 kb, the *copia*-like retroelements would represent 0.1%, 1% and 5% of the *E. guineensis* genome (A, B, C, respectively), which is similar to what has been reported from other species (Yaňez et al. 1998; Vicient et al. 1999). In *C. nucifera* class B represents 0.4% and classes A and C 0.01% of the genome.

Methylation status of genomic domains specified by the RT probes

Restriction enzymes *Msp*I and *Hpa*II are diagnostic for CpG methylation. *Msp*I cuts slightly better than the methylation-sensitive enzyme (*Hpa*II) due to the ability of *Msp*I to cleave C^{m5}CGG sites. In our experiments, both enzymes generated similar hybridisation patterns when probed with RT sequences of different classes; the higher the copy number the stronger the hybridisation

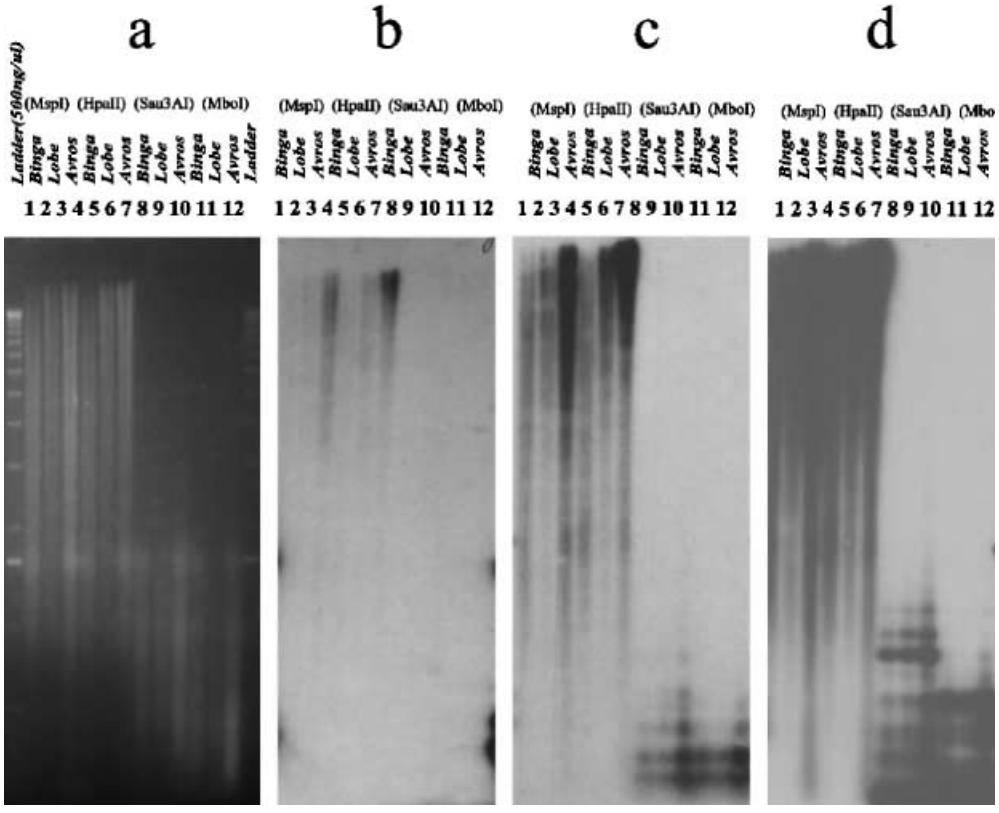

Fig. 4 Ethidium bromide-stained 1% agarose gel of: **a** genomic DNA (2.5 µg) digested with *Msp*I (*lanes* 1–3), *Hpa*II (*lanes 4–6*), *Sau*3AI (*lanes 7–9*), *Mbo*I (*lanes 10–12*); **b–d** Southern autoradiograph of genomic DNA (2.5 µg) digested with *Msp*I (*lanes 1–3*), *Hpa*II (*lanes 4–6*), *Sau*3AI (*lanes 7–9*), *Mbo*I (*lanes 10–12*) and probed with RT DNA probe 31 (class A) (**b**), RT DNA probe 35 (class B) (**c**), RT DNA probe 20 (class C) (**d**). Samples of *E. guineensis* from AVROS (DM 742-207), Lobe (T250/29-07) and Binga (T227/51-19) origins Rosenquist 1985) were used in this analysis

signal (Fig. 4). The second pair of enzymes (*Sau*3AI and *Mbo*I) with the recognition site GATC provides a test for symmetrical CpNpG methylation. The methylation-sensitive enzyme (*Sau*3AI) shows a higher than average distribution pattern under UV light/ ethidium bromide staining (Fig. 4a) than *Mbo*I, which is due to its inability to cleave GAT^{m5} C sites. We observed a difference in hybridisation patterns for the A, B and C classes that suggests that the high-copy-number class (C) appears to be more methylated than the low- and medium-copy-number classes (A, B).

Discussion

Evolutionary dynamics of Ty-*copia* retrotransposons in the palm family

One of the plant families which has been extensively studied is the grass family. Matsuoka and Tsunewaki (1999) and Gribbon et al. (1999) identified a number of Ty-*copia* retrotransposon families in *Gramineae*, and their analyses revealed several levels of organisation. The lowest organisational level that represented a distinct retrotransposon (such as *Bare-1*) was the subgroup. The subgroups were usually shared among the closely related species, and supergroups, in contrast, were spread across the entire extent of the *Gramineae*. The grass family arose approximately 60 M years (Wolfe et al. 1989). In contrast, the palm family is much older. Fossil information suggests that the subfamilies Coryphoideae, Arecoideae, Nypoideae and Calamoideae have been distinct evolutionary lines for more than 60 M years (Wilson et al. 1990). However, the base-substitution estimates, derived from the cpDNA (*rbc*L gene) restriction site data, indicate a potential 5- to 13-fold decrease in substitution rates within the palms relative to rate estimates for annual plant taxa. Wilson (1990) suggests that this is due to the fact that long-lived perennials take much longer to reproduce.

We identified three classes of *copia*-like RT sequences within the RT clones obtained from one individual plant (DM 742-207). Table 1 shows nucleotide divergence within *E. guineensis*, and it can be seen that class A shows quite a broad nucleotide divergence, which suggests that it may consist of a number of subclasses. Considering the wide range of the nucleotide divergence and the fact that the class-A sequences have about three times as many stop-codons as those of classes B and C, it seems possible that the former diversified quite early in the subfamily history leaving behind a number of subclasses with a low copy number. Although we have not detected any hybridisation to *Elaeis oleifera* (which is quite intriguing) the class is present in *Cocos nucifera* representing approximately 0.01% of the genome (Table 2). In spite of the fact that *E. guineensis* and *E. oleifera* are very closely related and are cross-fertile species belonging to the same genera, they have not been in contact since the breakup of Gondwanaland (60 M years ago). One might expect completely different evolutionary lineages of Ty-*copia* retrotransposons in

E. oleifera compared to *E. guinnesis*. The class B and C RT sequences have quite narrow nucleotide divergences within *Elaeis guineensis* and are present in high copy numbers in *Cocos nucifera* (class B) and in *E. guineensis* (class C). We estimated that there is approximately 0.5 to 0.6 nucleotide divergence between *C. nucifera* Ecorep 1, 2 and 3 ((gi|22680|emb |X70875.1), (gi|22679 |emb |X70874.1), (gi|22678|emb |X70873.1)) RT sequences and the class B RT sequences present in oil palm. Coconut Ecorep sequences are part of a *copia*-like *Eco*RI repetitive element (Rohde et al. 1992, 1995) and were generated by *Eco*RI digestion of genomic DNA. We also observed a 1.3-kb fragment after digesting coconut genomic DNA with *Eco*RI and probing with class B RT domains (unpublished data, Z. Price). Castilho et al. (2000) have recently reported finding a number of *copia*-like RT sequences but with no clear groupings. When we included those sequences into our dataset we found that most of them fell into classes A and C, class B being very under-represented by only one clone (unpublished data, Z. Price). The fact that there is so much conservation is very encouraging because any multilocus marker system based on retrotransposons would be applicable within different oil palm populations.

Methylation status of genomic domains specified by the RT probes

Methylation is the most common covalent modification of DNA found in vivo, and it plays an integral role in plant development, in regulating gene expression and in genome management (see Finnegan et al. 1998, 2000; Razin 1998 for reviews). Whereas methylation in animals is concentrated primarily in symmetric CpG dinucleotides, methylcytosine in plants and fungi can occur at C residues in any sequence context (Finnegan et al. 1998); most frequently at symmetric, CpG and CpNpG, sites (Gruenbaum et al. 1981). Assymetric C-methylation in plants was found at CpXpX (where X is any base other than G) sites in tobacco pollen (Oakeley and Jost 1996), in carrot (Zhou et al. 1998) and in transgenic petunia (Meyer and Heidmann 1994). It has been put forward by Yoder et al. (1997) that DNA methylation evolved as means of limiting the spread of transposable elements and viruses in host genomes. Using the bisulfite modification method, Goubely et al. (1999) observed that SINEs retroposons were methylated at symmetrical and assymetrical positions – mainly at CpG sites.

Results from our experiments showed that the genomic domains of *copia*-like retrotransposon as specified by RT probes are present in methylated regions. The differences in hybridisation pattern also suggest that the high-copy-number class (C) appears to be more methylated than classes A and B. Comparison of the hybridisation patterns of the two groups of restriction enzymes suggest that there is about a twofold difference in methylation at the CpG and CpNpG sites in favour of CpG methylation. This observation needs to be examined by genomic sequencing (Frommer et al. 1992; Goubely et al. 1999), which would allow us to make more concrete statements about the methylation status of *copia*-like sequences in oil palm. There is increasing evidence that reduced DNA methylation results in abnormal plant development (Finnegan et al. 1996; Richards et al. 1997; Chen et al. 1998). Interestingly, the work of Jaligot et al. (2000) showed that there was a correlation between DNA hypomethylation and the “mantled” somaclonal variation in oil palm. At present the occurrence of somaclonal variants, particularly for floral characters, restricts the application of clonal plant production in oil palm.

Acknowledgements The authors would like to acknowledge the funding provided by New Britain Palm Oil Ltd. (Papua New Guinea). Z. Price is supported by a BBSRC CASE PhD studentship.

References

Boeke JD, Corces VG (1989) Transcription and reverse transcription in retrotransposons. Annu Rev Microbiol 43:403–33

Breure CJ, Soebagjo FX (1991) Factors associated with occurrence of crown disease in oil palm (*Elaeis guineensis* Jacq) and its effect on growth and yield. Euphytica 54:55–64

Castilho A, Vershinin A, Heslop-Harrison JS (2000) Repetitive DNA and the chromosomes in the genome of oil palm (*Elaeis guineensis*). Ann Bot 85:837–844

Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenish R (1998) DNA hypomethylation leads to elevated mutation rates. Nature 395:89–93

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Corley RHV (1983) Potential productivity of tropical perennial crops. Exp Arig 19:217–237

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16:10881–10890

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

Ellis THN, Poyser IJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Ty1 copia class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

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

Felstenstein J (1985) Confidence limits on phylogenies: an aproach using the bootstrap. Evolution 39:783–791

Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc Natl Acad Sci USA 93:8449–8454

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES (1998) DNA methylation in plants. Annu Rev Plant Physiol Plant Mol Biol 49:223–47

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES (2000) DNA methylation, a key regulator of plant development and other processes. Curr Opin Genet Dev 10:217–223

Flavell AJ (1992) Ty1-*copia* group retrotransposons and the evolution of retroelements in the eukaryotes. Genetica 86:203–214

Flood J, Cooper RM, Lees PE (1989) An investigation into the pathogenity of four isolates of *Fusarium oxysporum* from South America Africa and Malaysia to clonal oil palm. J Phytopathol 124:80–88

Frommer M, Mcdonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA 89: 1827–1831

Ghesquiere M (1984) Enzyme polymorphism in oil palm (*Elaeis guineensis* Jacq).1. Genetic control of 9 enzyme-systems. Oleagineaux 39:561–574

Ghesquiere M (1985) Enzyme polymorphism in oil palm (*Elaeis guineensis* Jacq). 2. Variability and genetic structure of 7 origins of oil palm. Oleagineaux 40:529–537

Goubely C, Arnaud P, Tatout C, Heslop-Harrison JS, Deragon JM (1999) S1 SINE retroposons are methylated at symmetrical and non-symmerical positions in *Brassica napus*: identification of a preferred target site for assymetrical methylation. Plant Mol Biol 39:243–255

Grandbastien MA (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci 3:181–87

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

Gruenbaum Y, Naveh-Many T, Cedar H, Razin I (1981) Sequence specificity of methylation in higher plant DNA. Natur 292: 860–862

Hartley CWS (1988) The oil palm, 3rd edn. Longmans, London

Hirochika H (1993) Activation of tobacco retrotransposons during tissue-culture. EMBO J 12:2521–28

Hull R (1999) Classification of reverse transcribing elements: a discussion document. Arch Virol 144/1:209–214

Jack PL, Dimitrijevic TAF, Mayes S (1995) Assessment of nuclear mitochondrial and chloroplast RFLP markers in oil palm (*Elaeis guineensis* Jacq.). Theor Appl Genet 90:643–649

Jaligot E, Rival A, Beulé T, Dussert S, Verdeil Jl (2000) Somaclonal variation in oil palm (*Elaeis guineensis* Jacq.): the DNA methylation hypothesis. Plant Cell Rep 19:684–690

Kalendar R, Grog T, Regina M, Suoniemi A, Schulman A (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98:704–711

Konieczny A, Voytas DF, Cummings MP, Ausubel FM (1991) A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics 127:801–9

Kubis S, Schmidt T, Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plant genomes. Anal Bot 82:45–48

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

Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS (1997) The Ty1-*copia* group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. Genetica 100:205–217

Kuncze R, Saedler H, Lonnig WE (1997) Plant transposable elements. Adv Bot Res 27:331–340

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

Li ZY, Chen SY, Zheng XW, Zhu LH (1999) Identification and chromosomal localization of a transcriptionally active retrotransposon of *Ty*-3 gypsy type in rice. Genome 43:404–408

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

Marshall CR (1991) Statistical tests and bootstrapping: assessing the reliability of phylogenies based on distance data. Mol Biol Evol 8:836–891

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

Mayes S, James CM, Horner SF, Jack PL, Corley RHV (1995) The application of restriction fragment polymorphism to genetic fingerprinting in oil palm (*Elaeis guineensis* Jacq.). Mol Breed 2:175–180

Mayes S, Jack PL, Marshall DF, Corley HRV (1997) Construction of a RFLP map genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). Genome 40:116–122

Meyer P, Heidmann I (1994) Epigenetic variants of a transgenic petunia line show hypermethylation in transgene DNA – an indication for specific recognition of foreign DNA in transgenic plants. Mol Gen Genet 243:390–399

McClelland M, Nelson M, Rashke E (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Res 22:3640–3659

Moretzsohn MC, Nunes CDM, Ferriera ME, Grattapaglia D (2000) RAPD linkage mapping of the shell thickness locus in oil palm (*Elaeis guineensis* Jacq.). Theor Appl Genet 100:63–70

Oakeley EJ, Jost JP (1996) Non-symmetrical cytosine methylation in tobacco pollen DNA. Plant Mol Biol 31:927–930

Pearce SR, Harrison G, Heslop-Harrison JS (1997) Characterization and genomic organization of Ty1-copia retrotransposon in rye (*Secale cereale*). Genome 40:617–625

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

Razin A (1998) CpG methylation chromatin structure and gene silencing – a three way connection. EMBO J 17:4905–4908

Richards EJ (1997) DNA methylation and plant development. Trends Genet 13:319–323

Rival A, Beule T, Barre P, Hamon S, Duval Y, Noirot M (1997) Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq) tissue cultures and seed derived plants. Plant Cell Rep 16:884–887

Rohde W, Salamini F, Ashburner R, Randles JW (1992) An *Eco*RI repetitive sequence family of the coconut palm *Cocos nucifera* L. shows sequence homology to *copia*-like elements. J Genet Breed 46:391–394

Rohde W, Kullaya J, Rodriguez J, Ritter E (1995) Genome analysis of *Cocos nucifera* L. by PCR amplification of spacer sequences separating a subset of *copia*-like *Eco*RI repetitive elements. J Genet Breed 49:179–186

Rosenquist EA (1985) The genetic base of oil palm breeding populations. Proc Palm Oil Res Inst Malaysia 10:10–27

San Miguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 81:37–44

Shah FH, Rashid O, Simons AJ, Dunsdon A (1994) The utility of RAPD markers for determination of genetic variation in oil palm (*Elaeis guineensis* Jacq.). Theor Appl Genet 89:713–718

Suoniemi A, Anamthawat-Jónsson K, Arna T, Shulman AH (1996) Retrotransposon *BARE*-1 is a major, dispersed component of the barley (*Hordeum vulgare* L) genome. Plant Mol Biol. 30:1321–1329

Vicient CM, Suoniemi A, Anamthvat-Jónsson K, Tanskanen J, Beharav A, Nevo E, Shulman AH (2000) Retrotransposon BARE-1 and its role in genome evolution of Hordeum. Plant Cell 11:1760–1784

Voytas DF, Cummings P, Konieczny A, Ausubel FM, Rodermel SR (1992) *Copia*-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89:7124–7128

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BT, Powell W (1997) Genetic distribution of BARE 1 retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet 253:681–694

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITES important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–21

Wilson MA, Gaut B, Clegg MT (1990) Chloroplast DNA evolves slowly in the palm family (Arecaceae). Mol Biol Evol 7:303–314

Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH (1989) Date of the monocot-dicot divergence estimated from the chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:6201–6205

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 13:335–40

Yañez M, Verdugo E, Rodriguez M, Prat S, Ruiz-Lara S (1998) Highly heterogeneous families of Ty1/*copia* retrotransposons in the *Lycopersicon* genome.Gene 222:223–228

Zhou Y, Magill CW, Magill JM, Newton RJ (1998) An apparent case of nonsymetrical and sustained strand specific hemimethylation in the *Dc8* gene of carrot. Genome 41:23–33