Nucleic Acids Research, 1993, Vol. 21, No. 10 **2369–2373**

Non-random distribution of transposable elements in the nuclear genome of plants

J.Capel, L.M.Montero+, J.M.Martinez-Zapater and J.Salinas*
Departamento de Protección Vegetal, Centro de Investigación y Tecnología, INIA, Ctra. de la Coruña Km. 7, Madrid 28040, Spain

Received February 22, 1993; Revised and Accepted April 22, 1993

ABSTRACT

We have studied the genomic distribution of five different families of plant transposable elements by analyzing their location in DNA fractions from maize and tobacco genomes fractionated according to base composition. The results show that each family of elements is preferentially integrated in one specific fraction of its respective host genome. This demonstrates that the distribution of transposable elements in the nuclear genome of plants is not random but compartimentalized, i.e., the elements are located in specific genomic compartments characterized by having a specific G+C content and representing a small proportion of the genomes. Furthermore, these compartments seem to correspond to the genomic regions where most of the plant genes are also located, suggesting a preferential integration of transposable elements in the transcriptionally active regions of the plant genome. The implications of these results on the current applications of transposon tagging techniques are discussed.

INTRODUCTION

Transposable elements were identified in maize 40 years ago by B.McClintock, and, since then, they have been found in all the organisms where they have been sought, from bacteria to higher eukaryotes. Transposable elements have been classified in two classes according to their mode of transposition (1): (i) transposons, which transpose by a DNA–DNA mechanism, and (ii) retrotransposable elements which transpose via an RNA intermediate. In addition, these classes have been subdivided in two subclasses each according to structural differences (1). On one hand, transposons are subdivided depending on the presence of short or long terminal inverted repeats (TIR). On the other hand, retrotransposable elements are subdivided in retrotransposons, having long terminal repeats (LTR) and sharing similarities with retroviruses, and retroposons, or non-LTR elements, which bear an A-rich tail.

An important aspect of the biology of plant transposable elements concerns the distribution of integrated elements in the host genomes. This is specially interesting when considering the increasing use of transposable elements for gene cloning by transposon tagging in plants (2, 3). Currently, the accepted idea for plant transposable elements is that they are distributed randomly in the plant genomes. This idea is essentially based on restriction fragment analysis of the plant genomes which carry the elements, and on the analysis of the primary structure of the host sequences flanking them (4, 5). These two lines of evidence allow to rule out the existence of a small number of target sites of integration, but do not demonstrate that the distribution is at random. In fact, some reports have suggested that transposable elements do not integrate equally along the genomes. One of the best characterized plant transposons, the Activator (Ac) from maize, is known to transpose preferentially to linked sites in its natural host species (6, 7) and in transgenic tobacco plants carrying the elements (8). Moreover, when studying the location of 24 Ac-like elements in four maize inbred lines, it was shown that 13 elements were clustered on chromosome 4, four elements on chromosome 2 and three elements on chromosome 1, the rest of elements being scattered on different chromosomes (9). In the case of transposon Tam1 from *Antirrhinum majus*, the flanking nucleotides (40 bp) of a high number of integration sites have been characterized (10). The results revealed that Tam1 transposes preferentially to A+T rich sequences although it does not show sequence specificity of integration. This can be, however, the case of some members of the *Tourist* family, a very large family of maize transposable elements, which seem to display a target specific sequence (TAA) for integration (11). Finally, the Tnt-1 family of retrotransposons has been described to be preferentially located in A+T rich regions of the tobacco genome (12).

Recent investigations carried out in vertebrate mobile sequences have demonstrated that retroviral and retrotransposable elements are not randomly distributed in the vertebrate genomes, but preferentially located in DNA segments having a base composition which match that of the mobile sequences themselves

* To whom correspondence should be addressed

+ Present address: Instituto del Frío, CSIC, Ciudad Universitaria, Madrid 28040, Spain

(13, 14). These studies were performed by fractionating the vertebrate genomes according to base composition, using preparative centrifugation in density gradients in the presence of DNA ligands (15), and analyzing the distribution of different families of elements in the fractions so obtained. When genome fractionation was applied to plant genomes, it was shown that the nuclear genomes of plants, as those of vertebrates, are compartmentalized in that they consist in mosaics of isochores, long (>200 kb) segments of DNA fairly homogeneous in base composition. The isochores can be pooled into a small number of classes distinguished by different base composition (16). Such classes can be separated and used to study the genomic distribution of any sequence that can be probed. The study of the distribution of a number of genes in different plant genomes revealed that, similar to what is known in vertebrates, (i) plant genes are non-uniformly distributed in the genomes, and (ii) the G+C levels of the plant genes (exons, introns and individual codon positions) are linearly correlated with the G+C levels of the isochores harboring them (17).

The similarities just described between the genomes of plants and vertebrates prompted us to use the same experimental approach previously employed to study the distribution of mobile sequences in vertebrate genomes, to analyze the genomic distribution of plant transposable elements. On this way, we have characterized the distribution of five families of transposable elements, belonging to the four different subclasses described above, in the fractionated genomes of maize (Mutator, Ac and Cin4 families of elements) and tobacco (Tnt-1 family and a family of Ta1-1 homologous sequences). The results show that these plant transposable elements are non-randomly distributed in their host genomes, but preferentially located in specific genomic regions where most genes also seem to be found. The possible causes for this distribution, and the implications that these results can have in the use of transposable elements for insertional mutagenesis are discussed.

MATERIALS AND METHODS

Isolation, fractionation and G+C content analysis of nuclear DNA

Nuclear DNA (>50 kb in size) was extracted from leaves of maize (*Zea mays* L.) inbred lines RO1 (82-3010-4/2011-10) and 906-C (87-3559-6) and tobacco (*Nicotiana tabacum* L.) cv. White Burley as described elsewhere (16). Maize inbred lines 906-C and RO1 were kindly provided by B. McClintock and D.S. Robertson, respectively. Fractionation of DNA by preparative centrifugation in Cs_2SO_4 density gradients, in the presence of the DNA ligand BAMD [3,6 bis (acetato-mercuri-methyl) dioxane], was carried out as previously reported (16). Gradients were fractionated in 10 or 11 fractions, the pellet being considered as the first fraction. In the case of tobacco, we used the fractions previously characterized by Montero *et al.* (18). The G+C content of total (nuclear unfractionated) and fractionated DNAs was quantified by HPLC as already described (18).

Endonuclease digestion and hybridization

Total DNA and DNA from the genomic fractions were digested with restriction endonucleases according to the conditions given by the manufacturer, and electrophoresed on 0.8% agarose gels in TBE buffer (90mM Tris, 90mM boric acid, 1mM EDTA, pH 8.0). After electrophoresis, DNAs were stained with ethidium bromide to asses adequate loading and transferred to Hybond N^+ (Amersham, UK) nylon membranes. Hybridizations were performed at 65°C (homologous probes) or 50°C (heterologous probe) in 5×SSPE (20×SSPE is 3M NaCl, 0.2M NaH_2PO_4, 0.02M EDTA, pH7.7), 5×Denhardt's solution (1% Ficoll, 1% PVP, 1% BSA), 0.5% SDS (sodium-dodecyl-sulphate) and 0.5 mg/ml denatured herring sperm DNA. Probes (see below) were labelled by the random priming method (19). After hybridization, membranes were washed twice for 10 minutes in 2×SSPE, 0.1% SDS, and 15 minutes in 1×SSPE, 0.1% SDS at the hybridization temperature. When using homologous probes, membranes were washed additionally for 15 minutes in 0.1×SSPE, 0.1% SDS at 65°C.

Probes

The Mutator (Mu) probe was a 2.9 kb BamHI-HindIII fragment from plasmid pMJ9 (20). The Ac probe was a 4.5 kb PstI fragment from plasmid pAC2 (21). The Cin4 probe was an 800 bp EcoRI fragment from the reverse transcriptase sequence of the maize Cin4-15 retroposon (22). The Tnt-1 probe was a 330 bp BglII fragment from the reverse transcriptase sequence of the tobacco Tnt-1 retrotransposon (23). The Ta1-1 probe was a 2.3 kb SalI fragment from the sequence of the *Arabidopsis thaliana* Ta1-1 retrotransposon (24).

RESULTS

Fractionation of nuclear DNA

The fractionation of nuclear DNA from maize inbred lines RO1 and 906-C by preparative density gradient centrifugation in Cs_2SO_4/BAMD was performed as already described (17). The

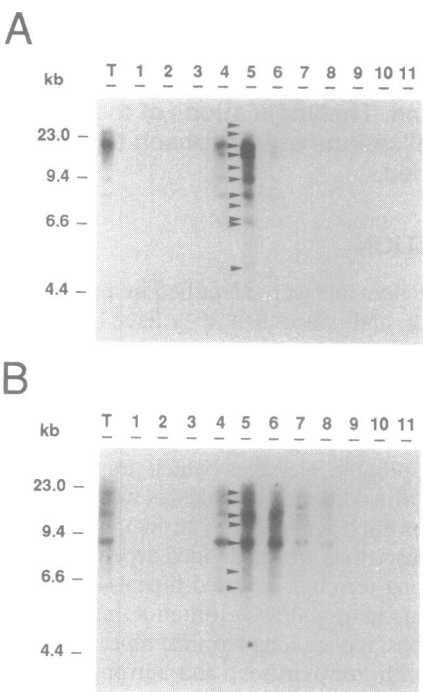

Figure 1. Location of transposons Mu (A) and Ac (B) in total unfractionated DNA (T) and in genomic fractions from maize inbred lines RO-1 and 906-C respectively. Ten micrograms of total DNA, and DNA fractions in amounts proportional to 10 μg of total DNA were processed as indicated in Materials and Methods. Arrows indicate the localization of hybridizing fragments.

relative G+C content of total unfractionated DNAs and of DNA from genomic fractions, and the relative amount that each fraction represents within the genomes are shown in Table I. In the case of tobacco, we used the genomic fractions previously characterized by Montero *et al.* (18). Features concerning these fractions are also indicated in Table I for comparison. The genome of maize is characterized by having higher G+C content and compositional heterogeneity than the genome of tobacco. In the case of maize, the range of G+C content covered by the genomic fractions is higher than 10%, whereas in tobacco it is only about 6% (Table I). These results are in close agreement with those already published (16, 18).

Location of transposable elements in fractionated DNA

In order to study the genomic distribution of plant transposable elements, DNA from the compositional fractions shown in Table I was analyzed by DNA-blot hybridization as described in Materials and Methods. To increase the probability of detecting hybridizing fragments corresponding to independent insertional events, restriction endonucleases were chosen based on the absence of target sites within the published sequences of the elements. However, it should be stressed that not all the hybridizing fragments do represent single insertions. The comigration of different restriction fragments harboring transposable elements, and/or the presence of more than one element in a single restriction fragment, could result in an underestimation of the number of insertions and would explain the different intensities observed between hybridizing bands.

Two additional considerations should be made for the correct interpretation of the DNA-blot hybridization results. First, because of the enrichment for specific DNA molecules in fractionated DNA, it is frequent to detect additional hybridizing fragments, in DNA fractions, that are not observed in total unfractionated DNA. Second, since the fractionated DNA molecules come from the unavoidable random breakage of nuclear DNA during extraction, a continuity in their distribution along the gradient is produced. Because of this continuity, each hybridizing fragment can be seen in more than one contiguous fraction (see reference 25 for a more detailed explanation). As a general rule, the fraction showing the strongest signal was always considered as the fraction containing the DNA molecules where the insertion is located.

Location of transposons in maize DNA fractions. The genomic distribution of Mu, a long TIR transposon, was analyzed in EcoRI digests of DNA fractions from the maize mutator line RO1. The hybridization results (Figure 1A) showed at least 7 hybridization bands on total unfractionated DNA. When DNA fractions were considered, eleven hybridization fragments, ranging from 5 to more than 23 kb, could be detected all of them being preferentially located in fraction 5. This result means that the Mu elements present in the genome of maize line RO1 are located in a class of isochores characterized by having an average G+C content of 44.4% and representing less than 18% of the maize genome (Table I). Similar results were obtained when studying the genomic distribution of Mu elements in the maize mutator line 82-2017-1/3012-4. In this line, the 8 hybridizing fragments detected were located in a genomic fraction characterized by having 44.2%G+C content and representing 12.4% of the maize genome (data not shown). Since the used Mu probe contains part of the Adh-1 maize gene, and this gene has been described to be included in a 20 kb EcoRI fragment (26), the 20 kb hybridizing fragment detected in fraction 5 from maize line RO1 (Figure 1A) must contain the Adh-1 gene. In fact, this has been proved to be the case by hybridizing the same membrane with a specific Adh-1 probe (data not shown). Consequently, we can conclude that, in the genome of maize, Mu elements are located in the same class of isochores than the Adh-1 gene.

To study the genomic distribution of Ac, a short TIR transposon, DNA from the compositional fractions of the maize inbred line 906-C was digested with KpnI. The hybridization pattern obtained is shown in Figure 1B. At least seven hybridizing fragments were detected in total nuclear DNA digested with KpnI. All these fragments, having from 5 to more than 23 kb, were centered on fraction 5 indicating that, in this maize line, Ac elements are inserted in a class of isochores that represent less than 15% of the genome and have an average G+C content of 44.2% (Table I). Very similar results were obtained when the genomic distribution of Ac was studied in the maize line 84766. In this line, eight hybridizing bands were detected in fractionated

Table I. G+C composition of total unfractionated DNA and of DNA from the genomic fractions obtained from Cs_2SO_4/BAMD density gradients.

Maize RO1			Maize 906-C			Tobacco		
Fraction (No)	Genome (%)	G+C (%)	Fraction (No)	Genome (%)	G+C (%)	Fraction (No)	Genome (%)	G+C (%)
1	7.2	–	1	10.9	–	1	11.9	35.20
2	5.1	39.50	2	6.8	39.60	2	1.3	35.20
3	5.5	41.95	3	7.5	41.30	3	4.0	35.80
4	10.7	43.05	4	11.6	43.80	4	30.8	36.60
5	17.5	44.40	5	14.0	44.20	5	19.2	38.00
6	14.0	45.70	6	12.4	44.60	6	19.8	38.00
7	9.8	45.85	7	9.2	45.30	7	5.8	39.20
8	6.6	46.75	8	11.4	45.70	8	3.0	39.60
9	8.9	46.90	9	8.3	46.70	9	1.3	40.20
10	8.4	49.60	10	4.7	49.40	10	2.9	41.40
11	6.3	49.80	11	3.2	47.80			
TOTAL	100.0	45.15	TOTAL	100.0	45.50	TOTAL	100.0	37.70

This table provides the relative content of Guanine+Citosine (%G+C) of total unfractionated DNA (Total) and of DNA from the genomic fractions obtained from Cs_2SO_4/BAMD density gradients. The relative amount of each DNA fraction within the genomes (%Genome) is also indicated. Values corresponding to tobacco fractions have already been shown (18). The% G+C of the maize first fractions (pellets) could not be determined. The lower G+C content of fraction 11 from maize inbred line 906-C compared to that of fraction 10 is due to contamination with A+T rich DNA from the pellet during fractionation (for more details see refs. 16, 25).

DNA, and all of them were centered in a genomic fraction having the same G+C content as the fraction where the Ac elements were located in line 906-C (data not shown).

Location of retrotransposable elements in maize and tobacco DNA fractions. The genomic distribution of Cin4, the only plant retroposon characterized so far (27), was studied in the same compositional fractions from maize inbred line 906-C that were used to analyze the distribution of Ac. The hybridization results allowed to detect 9 hybridizing fragments (Figure 2A). One of these fragments (6.8 kb) was located in fraction 4 (43.8% G+C), another fragment (8.5 kb) in fraction 6 (44.6% G+C), and the remaining seven fragments, having molecular weights between 7.6 and more than 23 kb, were centered on fraction 5 (44.2% G+C). From these results, we conclude that the majority of the Cin4 retroposons in the maize inbred line 906-C are located in the same class of isochores than the Ac transposons.

The hybridization results concerning the genomic distribution of the only plant retrotransposon known to be active, Tnt-1 (27), are displayed in Figure 2B. At least 25 hybridizing bands were detected in EcoRI digests of total DNA from tobacco cv. White Burley. On fractionated DNA, all these bands were centered on fraction 4 which represents 30% of the tobacco genome and has an average G+C content of 36.6% (Table I). A very similar distribution has been reported for this family of retrotransposons in the genome of tobacco cv. Xanthi (12).

The same tobacco fractions used to study the distribution of Tnt-1 were also used to characterize the genomic distribution of a tobacco repetitive sequence that cross-hybridized with a probe from the retrotransposon Ta1-1 from *Arabidopsis thaliana*, and probably represents the recently described tobacco Tnt-14 retrotransposon (28). When hybridized with the Ta1-1 probe, total unfractionated DNA from White Burley tobacco showed more than 20 fragments (Figure 2C). All these hybridizing fragments were centered on fraction 4 indicating that, in this tobacco cultivar, Ta1-1 homologous sequences are integrated in the same compositional fraction than Tnt-1 retrotransposons.

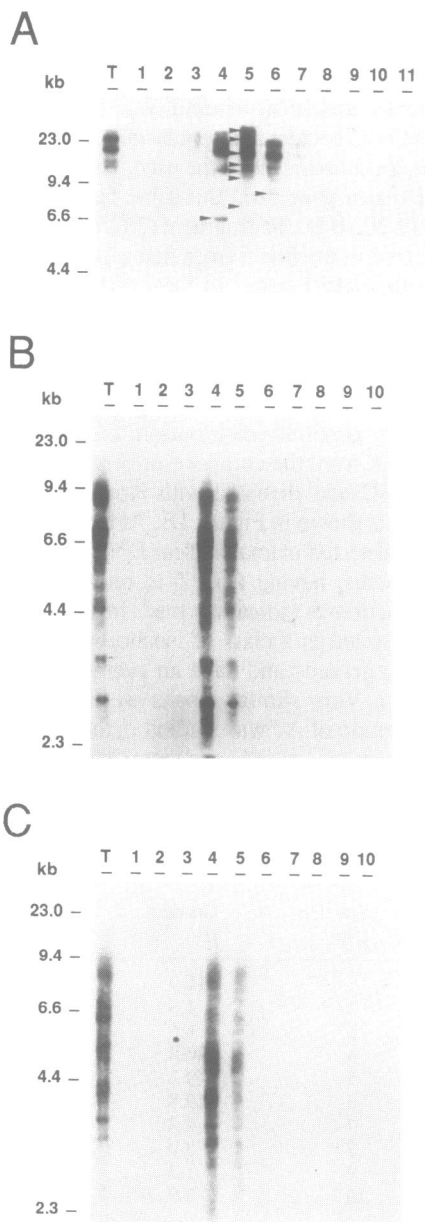

Figure 2. Location of retrotransposable elements Cin4 (A), Tnt-1 (B) and tobacco homologous sequences to Ta1-1 (C) in total unfractionated DNA (T) and in genomic fractions from maize inbred line 906-C (Cin4) and tobacco cv. White Burley (Tnt-1 and Ta1-1 homologous sequences). For other indications see legend of Figure 1.

DISCUSSION

The aim of this work was to investigate the distribution of transposable elements in the nuclear genome of plants. With this purpose, we analyzed the distribution of four different families of transposable elements, and one family of retrotransposon-homologous sequences, in the genomic fractions of two host plant species. The results obtained show that all detected elements from the Mu and Ac transposon families, and the majority of the elements from the Cin4 family of retroposons, are preferentially located in a genomic fraction characterized by having about 44% of G+C content and that represents less than one fifth of the maize nuclear genome. On the other hand, all the Tnt-1 retrotransposons and Ta1-1 homologous sequences detected are located in a genomic fraction characterized by having a G+C content of 36.6% and representing only one third of the tobacco nuclear genome. Taken together, these results indicate that the plant transposable elements analyzed, irrespective of their structures and transposition mechanisms, are not distributed at random in their host genomes but in a compartmentalized way, i.e., preferentially in specific genomic regions characterized by having a specific G+C content and representing small amounts of the total genomes. If these transposable elements would be randomly distributed in their host genomes, in all cases the probability of finding the results just described would be lower than 0.1%, this value being calculated by χ^2 analysis considering the relative amount of the total genome represented by each fraction and the number of elements found in each one of these fractions.

Since plant transposable elements do not seem to integrate in specific target sites (10, 11), the compositionally compartmentalized distribution evidenced in this work should be the result of constraints that are independent from a targeted integration. We think that such constraints could be related to a higher accessibility of the integrase complex in these regions. Consequently, integration of plant transposable elements would

occur more frequently in expanded domains of chromatin which have been shown to be related to transcriptionally active regions of the genomes or to origins of DNA replication (29). A similar hypothesis has been proposed to explain the preference that certain vertebrate retroviruses and retrotransposons have to integrate in specific restriction fragments (30, 31). Moreover, it has been indicated that retroviral sequences integrate, in their majority, in compartments of the vertebrate genomes which are active both transciptionally and recombinogenically (14).

Although the transcriptional activity of the isochores where plant transposable elements have been located is unknown, the presence of genes in these isochores can be taken as indicative of such an activity. On this way, we hybridized the membranes containing the genomic fractions employed to localize Ac, Mu and Cin-4 elements, with different maize coding sequences whose genomic distribution has already been reported (17). The results showed that most of these sequences are located in the same class of isochores as that where we have found the transposable elements (Capel et al., unpublished results). Corresponding experiments carried out with at least seven tobacco coding sequences of known genomic distribution (16, 17), also showed that all these sequences lie in the same class of isochores as that where we have located the Tnt-1 retroposons and the Ta1-1 homologous sequences (Capel et al., unpublished results). Furthermore, Matassi *et al.* (12) have also reported that Tnt-1 elements and several coding sequences are located in the same compositional fraction of the genome of tobacco cv. Xanthi. Even considering that all these data are still limited, they suggest that plant transposable elements are preferentially located in isochores containing a high density of genes and, therefore, transcriptionally active. In agreement with the hypotesis that plant transposable elements integrate in transcriptionally active regions of the host genomes, Schwarz-Sommer *et al.* (22) demonstrated that Cin4 mRNAs are transcribed from external promoters, suggesting that Cin4 retroposons integrate into transcribed regions of the maize genome. Moreover, the two genes originally used to trap the Mu and Ac maize transposons, Adh-1 and Wx respectively (20, 21), have been localized in the same class of isochores where we have found these transposable elements (Capel et al., unpublished results). Similarly, the *nia-2* gene from tobacco, which was used to trap the Tnt-1 retrotransposon (23), has also been localized in the same class of isochores (12) where we have found the Tnt-1 elements.

The results reported here have important implications to be considered when using plant transposable elements as insertional mutagens. Based on the idea that plant transposable elements are randomly distributed in the genomes, it seems reasonable to assume that the use of smaller plant genomes will increase the probability of insertion of an element into a specific gene. Our results, however, indicate that plant transposable elements are located preferentially within reduced genomic regions where, furthermore, most coding sequences are also located, thus reducing the effective genome size to the size of the gene-bearing isochores. When considering this fact, the probability of insertion of a plant transposable element within a specific gene should be less dependent from genome size than previously thought. Another consequence of the non-random distribution of transposable elements in the nuclear genome of plants is that the probability of tagging genes that are found in isochores not frequented by transposable elements is drastically reduced.

ACKNOWLEDGEMENTS

We thank J.Filipski for critical reading of the manuscript and J.L.Tadeo for technical advice. We also thank J.L.Bennetzen, M.Caboche, N.V.Federoff, Z.Schwarz-Sommer and D.F.Voytas for the gift of DNA probes. This work was supported by grant INIA 7556 from the Spanish Ministry of Agriculture.

REFERENCES

1. Finnegan D.J. (1989) Trends Genet. **5**, 103–10.
2. Balcells L., Swinborne J. and Coupland G. (1991) Trends Biotech. **9**, 31–37.
3. Gierl A. and Saedler H. (1992) Plant Mol. Biol. **19**, 39–49.
4. Döring H.P. and Starlinger P. (1986) Annu. Rev. Genet. **20**, 175–200.
5. Planckaert F. and Walbot V. (1989) Genetics **123**, 567–578.
6. Greenblatt I.M. (1984) Genetics **108**, 471–485.
7. Dooner H.K. and Belachew A. (1989) Genetics **122**, 447–457.
8. Jones J., Carland F., Lim E., Ralson E. and Dooner H.K. (1990) Plant Cell **2**, 701–707.
9. Johns M.A., Fuerstenberg S.I. and Hennelly C.A. (1990) Genet. Res. **55**, 71–80.
10. Nacken W.K.F., Piotrowiak R., Saedler H. and Sommer H. (1991) Mol. Gen. Genet. **228**, 201–208.
11. Bureau T.E. and Wessler S.R. (1992) Plant Cell **4**, 1283–1294.
12. Matassi G., Melis R., Macaya G. and Bernardi G. (1991) Nucl. Acids Res. **19**, 5561–5567.
13. Bernardi G. (1989) Annu. Rev. Genet. **23**, 637–661.
14. Rynditch A., Kadi F., Geryk J., Zoubak s., Svoboba J. and Bernardi G. (1991) Gene **186**, 165–172.
15. Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. (1985) Science **228**, 953–958.
16. Salinas J., Matassi G., Montero L.M. and Bernardi G. (1988) Nucl. Acids Res. **16**, 4269–4285.
17. Montero L.M., Salinas J., Matassi G. and Bernardi G. (1990) Nucl. Acids Res. **18**, 1859–1867.
18. Montero L.M., Filipski J., Gil P., Capel J., Martinez-Zapater J.M. and Salinas J. (1992) Nucl. Acids Res. **20**, 3207–3210.
19. Feinberg A.P. and Volgestein B. (1983) Anal. Biochem. **132**, 6–13.
20. Bennetzen J.L. (1984) J. Mol. Appl. Genet. **2**, 519–524.
21. Pohlman R.F., Federoff N.V. and Messing J. (1984) Cell. **37**, 635–643.
22. Schwarz-Sommer Z., Leclercq L., Göbel E. and Saedler H. (1987) EMBO J. **6**, 3873–3880.
23. Grandbastien M.A., Spielmann A. and Caboche M. (1989) Nature **337**, 376–380.
24. Voytas D.F. and Ausubel F.M. (1988) Nature **336**, 242 –244.
25. Salinas J., Zerial M., Filipski J. and Bernardi G. (1986) Eur. J. Biochem. **160**, 469–478.
26. Dennis E.S, Gerlach W.L., Pryor A.J., Bennetzen J.L., Inglis A., Llewellyn D., Sachs M.M., Ferl R.L. and Peacock W.J. (1984) Nucl. Acids Res. **12**, 3983–4000.
27. Grandbastien M.A. (1992) Trends Genet. **8**, 103–108.
28. Flavell A.J., Smith D.B. and Kumar A. (1992) Mol. Gen. Genet. **231**, 233–242.
29. Scherdin U., Rhodes K. and Breindl M. (1990) J. Virol. **64**, 907–912.
30. Shih C.C., Stoye J.P. and Coffin J.M. (1988) Cell **53**, 531–537.
31. Natsoulis G., Thomas W., Roghmann M.-C., Winston F. and Boeke J.D. (1989) Genetics **123**, 269–279.