The distribution of T-DNA in the genomes of transgenic *Arabidopsis* and rice

Abdelali Barakat[a,1], Patrick Gallois[a], Monique Raynal[a], Dominique Mestre-Ortega[a], Christophe Sallaud[b], Emmanuel Guiderdoni[b], Michel Delseny[a], Giorgio Bernardi[c,*]

[a] *Laboratoire de Biologie Moléculaire et Physiologie Végétale, Université de Perpignan, 52 Avenue Villeneuve, 66000 Perpignan, France*
[b] *Centre de Coopération Internationale en Recherche Agronomique pour le Developpement, Montpellier, France*
[c] *Stazione Zoologica Anton Dohrn, Laboratorio di Evoluzione Molecolare, Villa Comunale, 80121 Naples, Italy*

Received 1 March 2000

Edited by Horst Feldmann

Abstract Almost all the nuclear genes of four *Gramineae* (maize, wheat, barley, rice) and pea are located in DNA fractions covering only a 1–2% GC range and representing between 10 and 25% of the different genomes. These DNA fractions comprise large gene-rich regions (collectively called the 'gene space') separated by vast gene-empty, repeated sequences. In contrast, in *Arabidopsis thaliana*, genes are distributed in DNA fractions covering an 8% GC range and representing 85% of the genome. Here, we investigated the integration of a transferred DNA (T-DNA) in the genomes of *Arabidopsis* and rice and found different patterns of integration, which are correlated with the different gene distributions. While T-DNA integrates essentially everywhere in the *Arabidopsis* genome, integration was detected only in the gene space, namely in the gene-rich, transcriptionally active, regions of the rice genome. The implications of these results for the integration of foreign DNA are discussed.

Key words: Isochore; Plant; Gene

1. Introduction

Previous studies on plant genome organization have shown that (i) their genomes are compositionally compartmentalized [1–3]; (ii) the majority of genes from *Gramineae* (maize, barley, rice and wheat) and from a dicot, pea, are distributed in compositionally distinct DNA fractions (collectively called the gene space) having a narrow, 1–2% GC range (GC is the molar fraction of guanine+cytosine in DNA) and representing between 10% and 25% of the genomes in different plants [4–6]. Since in all cases, but especially in *Gramineae*, genes are compositionally heterogeneous [2,7], the compositional homogeneity of the gene space must be due to the narrow compositional range of intergenic sequences, which seem to be mainly formed by transposons [8]. The gene-rich regions that form the gene space are distributed over all chromosomes and are separated by vast regions of gene-empty, repeated sequences. In contrast, the *Arabidopsis thaliana* genome presents a different gene distribution in that genes are scattered over DNA fractions that cover a wide, 8%, GC range and represent about 85% of the genome [9]. This genome is also characterized by very small amounts of repeated sequences.

The different genome organizations just described may influence the pattern of transgene integration, since integration of foreign DNA seems to occur only in actively transcribing, open chromatin regions. Indeed, we noticed [4] that the gene space of the maize genome appears to correspond to the only genome compartments in which certain mobile sequences can be transposed. For instance, Mu (mutator), Ac (activator), and the majority of Cin4 elements are exclusively located in the same class of isochores as the ADH-1 gene [10], namely within the gene space. Maize transposable elements appear, therefore, to be located in gene-rich, transcriptionally active isochores. Interestingly, this is in agreement with previous observations on transcribed proviral sequences that also integrate into transcriptionally active regions of mammalian genomes (see [11] for a review), and suggests a correlation between integration and transcriptional activity.

Agrobacterium tumefaciens-mediated transformation is frequently used to transfer genes of agronomic interest to plants. Transgenes are also tools for tagging and isolating new genes and for studying their functions. Therefore, understanding transferred DNA (T-DNA) integration in the genome is very important in several respects. If the mechanisms of T-DNA transfer and integration have been relatively well studied [12,13], little is known about the distribution of T-DNA in the genome of transgenic plants and on the constraints acting on the integration mechanism.

In the case of dicots, previous studies on tomato, *Petunia hybrida* and *Crepis capillaris* have suggested that T-DNA integrates in all chromosomes [14–17]. As far as host sequence preference is concerned, the T-DNA integration pattern is, however, not known because all previous studies involved the analysis of only a few transformants. In the case of monocots, there are just no data about T-DNA integration.

In order to answer this question, we chose an approach which has been used to study gene distribution in plant and animal genomes [18], namely to localize T-DNA in compositional fractions of plant DNA, namely in DNA fractions separated according to their GC levels. In the case of *Arabidopsis*, an additional approach was also used, namely sequencing of host chromosomal regions flanking T-DNA insertion sites and finding the base composition of large DNA segments harboring these regions in data banks. Here, we report results on the

*Corresponding author. Fax: (39)-81-2455807.
E-mail: bernardi@alpha.szn.it

[1] Permanent address: Laboratoire des Sciences des Aliments, Département de Biologie, Faculté des Sciences Semlalia, BP S15, 40001 Marrakech, Morocco.

Abbreviations: BAC, bacterial artificial chromosome; BAMD, 3,6-bis-(acetatomercurimethyl)-1,4-dioxane; GC, molar fraction of guanine+cytosine in DNA; GUS, *iudA*; T-DNA, transferred DNA

PII: S0014-5793(00)01393-4

distribution of T-DNA integration sites in transgenic *Arabidopsis* and rice, two plants which have different gene distributions and different amounts of repeated sequence.

2. Materials and methods

2.1. Arabidopsis transgenic plants and construct preparations

Two series of experiments were carried out in the case of *Arabidopsis*. In the first series, transgenic plants were obtained by transformation with the pBIN19TD plasmid, which contains a gene conferring resistance to hygromycin as well as a fusion construct comprising the *uidA* (GUS) reporter gene and each one of five characterized thioredoxin promoters [19]. GC levels of the promoters corresponding to the five constructs were comprised between 29.8 and 42.4% GC, but the global GC of all constructs was quite uniform, 20–22%. All constructions were translational fusions [20]. Plant transformation was done by infiltration in planta [21]. Selection for resistant seeds was done in vitro on media containing hygromycin. The presence of transgenes in plants was checked by polymerase chain reaction (PCR) amplification using one primer in the GUS sequence and another in the thioredoxin promoter. Finally, all selected lines were investigated in their GUS expression [20]. Transgenic lines were grown for 3 weeks under standard conditions. Plants were kept in the dark for 2 days before DNA preparation to decrease the amount of starch in leaves.

In the second series of experiments, *Arabidopsis* seed lines from the Versailles INRA collection of T-DNA insertion lines were used for sequencing T-DNA junctions. These were obtained by transformation [21] using a binary vector based on Basta resistance [22]. Growing conditions were the same as for the other lines cited above.

2.2. Rice transgenic plants and construct preparations

Transgenic lines of *Oryza sativa* cultivar *japonica* (Taipei 309) were generated at the Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement (CIRAD, Montpellier, France). These plants were obtained by transformation of globular calli with the *Agrobacterium* pCAMBIA30063 plasmid, a binary plasmid containing GUS (35:GUS) and GFP (E35:SGFP) sequences and a hygromycin resistance gene (Sallaud et al., in preparation). The callus transformation procedure followed the protocol of Hiei et al. [23].

2.3. DNA fractionation and localization of inserts

Nuclear *Arabidopsis* DNA was prepared from leaves using the method of Jofuku and Goldberg [24] with minor modifications. Nuclear DNA samples of transgenic *Arabidopsis* lines with thioredoxin promoters were fractionated by preparative centrifugation [25] in Cs_2SO_4 density gradients in the presence of 3,6-bis(acetatomercurimethyl)-1,4-dioxane (BAMD). T-DNA insertion distributions were determined by Southern hybridization on gradient fractions restricted with *Hin*dIII (an enzyme not cutting the probe) and loaded on the filters proportionally to their amounts in total DNA. Fragments of the GUS gene amplified by PCR with specific primers were used as probes. Additionally, the T-DNA distribution in the *Arabidopsis* genome was analyzed using PCR amplification (see above) on each Cs_2SO_4/BAMD fraction.

Fractionation of nuclear DNA from rice was done as for *Arabidopsis* DNA. The DNA fractions corresponding to the gene space of rice were determined by hybridization with previously described [5] fractions of rice DNA corresponding to the gene space.

DNA was labelled with ^{32}P using a random priming labelling system (Amersham) and hybridizations were done as already described [26] with minor modifications.

2.4. Sequencing of T-DNA insert junctions from the Arabidopsis genome and compositional distribution of large DNA sequences harboring T-DNA

As already mentioned, insertion mutants were from the Versailles INRA collection of T-DNA insertion lines. Amplification of T-DNA flanking sequences was done by PCR walking [27]. Amplified fragments were sequenced using an ABi (Perkin-Elmer) sequencer and a dye terminator kit. In the majority of cases, this allowed us to identify bacterial artificial chromosomes (BACs) already sequenced by using the Blast program [28] on the large genomic sequences (>50 kb) available in GenBank. The GC level of BACs could be determined using Bisance software. Statistical analysis of the data was done using the Statview software.

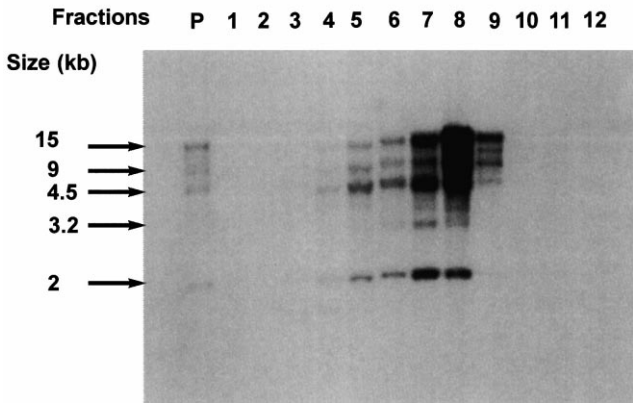

Fig. 1. Localization of T-DNA integration in the genome of transformed *Arabidopsis*. A GUS probe was hybridized on Cs_2SO_4/BAMD DNA fractions restricted with *Hin*dIII, an enzyme not cutting the GUS sequence. *Arabidopsis* DNA was pooled from 30 plants. P, pellet; 1–12, DNA fractions.

3. Results

3.1. T-DNA distribution in the Arabidopsis genome

When the GUS probe was hybridized on *Hin*dIII-restricted Cs_2SO_4/BAMD fractions from pooled DNA of 30 *Arabidopsis* transformants, hybridization signals were localized (Fig. 1) on fractions representing about 85% of the genome (as judged upon the optical density of fractions 4–9), namely an amount corresponding to the DNA fractions that contained the genes previously tested [9]. Those results were confirmed by PCR detection of T-DNA in the same fractions.

DNA sequences surrounding T-DNA integration sites in 180 transformed *Arabidopsis* lines allowed us to identify BACs corresponding to the majority of T-DNA insertions and to assess their GC levels. The compositional distribution of the BACs in which T-DNA integrations were localized showed (Fig. 2) that they covered a 32–40% GC range,

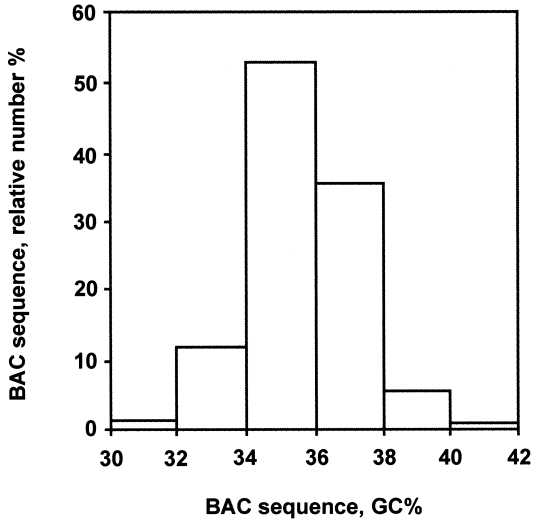

Fig. 2. Compositional distribution of BACs (from GenBank) which contained sequences flanking the T-DNA present in the genome of 180 transformed lines of *Arabidopsis*.

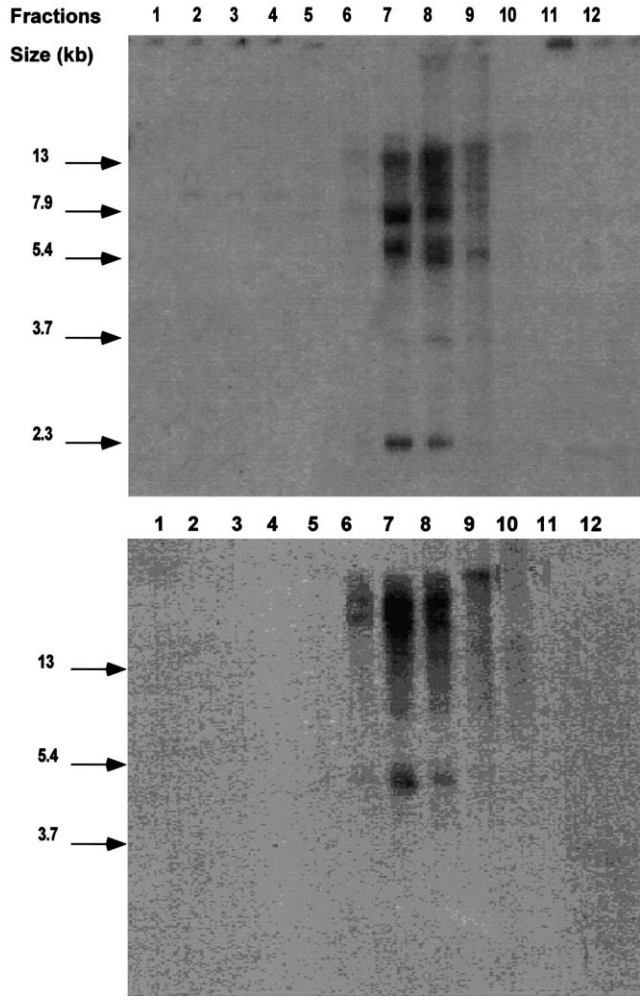

Fig. 3. Top: hybridization of GUS sequence on Cs_2SO_4/BAMD fractions from rice. 1–12, DNA fractions. Bottom: hybridization of previously described [5] gene space fractions of rice on the fractions of the top frame.

namely a range similar to that covered by the genome regions in which genes were localized and which represents about 85% of the genome [9]. Integration was maximal in DNA fragments having 34–36% GC levels, which correspond to the distribution maxima of DNA and of genes in the *Arabidopsis* genome.

3.2. T-DNA distribution in rice genome

Hybridization of the GUS sequence present in T-DNA on Cs_2SO_4/BAMD fractions of pooled DNA from transgenic rice expressing GUS gene showed that hybridization signals were centered in fractions 5–9 which correspond to the gene space, as shown by hybridization of DNA corresponding to the gene space prepared as described [5]. The same experiments on Cs_2SO_4/BAMD DNA fractions of transgenic rice not expressing the GUS gene also produced bands located in fractions corresponding to the gene space (Fig. 3).

4. Discussion

Using 210 and 30 independent transformants for *Arabidopsis* and rice, respectively, we showed that T-DNA in *Arabidopsis* and rice exhibits different integration patterns, which match the gene distribution patterns of those plants.

Global analysis of integration of five different T-DNA constructs in the *Arabidopsis* genome by PCR amplification and by hybridization experiments on Cs_2SO_4/BAMD fractions showed that T-DNA integrates in different genomes representing about 85% of the genome. This was confirmed by analysis of the compositional distribution of 180 large sequences (>50 kb) harboring independent T-DNA insertion. Indeed, insertions occurred in BACs covering an 8% GC range (32–40% GC) and representing 85% of the genome. The pattern of T-DNA integration in *Arabidopsis* appears to be due to the fact that genes occupy most of the genome and that most compartments are transcriptionally active.

In contrast, T-DNA was found to be integrated in rice genome fractions which correspond to the gene space, cover a compositional range of only 1% GC and represent only about 25% of the genome. The difference in the T-DNA integration pattern between *Arabidopsis* and rice is most probably due to their strikingly different gene (and open chromatin) distributions and to the difference in the amount of repeated sequences in their genomes.

Previous calculations predict that *Arabidopsis* genome saturation could be achieved with 70 000 insertions. Since the gene space of rice corresponds to 25% of the genome, namely to 100 Mb (which is approximately the size of the *Arabidopsis* genome), and since T-DNA preferentially inserted in the gene space, saturation of transcribed sequences by T-DNA insertion in rice may only require approximately the same number of insertions as in *Arabidopsis*. This aspect might be of even more practical importance for species with larger genomes, and is in agreement with the results [29], which showed that T-DNA insertion in *Arabidopsis* and tobacco induced transcriptional and translational gene fusion at similar frequencies in spite of very different genome sizes. These conclusions differ from previous ones resulting from studies on T-DNA insertion in genomes of tomato, *P. hybrida* and *C. capillaris* [14–16], which suggested that there was no bias of T-DNA integration. This could be explained by the fact that those studies concerned a small number of transformants and could not give a general view on the T-DNA distribution in the genome of plants studied.

A comparison of the T-DNA integration distribution in rice plants expressing and not expressing the GUS protein showed that T-DNA integrates in the gene space fraction independently of the expression of the GUS protein. Moreover, rice transformants generally contain 1–3, with an average of 1.5, T-DNA insertions (Sallaud et al., in preparation). Therefore, the 30 transformants analyzed in this study represent around 45 T-DNA insertions. Out of the 45 T-DNA insertions, in principle 30 were sufficient in the selection procedure to obtain 30 independent transgenic rice. Despite this, all T-DNA sequences detected are in the gene space. This further suggests that insertion in the gene space is mainly dictated by the open status of chromatin regions. This is in agreement with investigations [13] on tobacco transgenic plants showing that transferred sequences are located in the DNase-I sensitive domain in all transformants independently of their expression.

In conclusion, this work shows that T-DNA is inserted in almost all genome regions of *Arabidopsis*, but only in the gene space of rice, and suggests that T-DNA insertions should give access to knock-outs in most genes without requiring an ex-

ceedingly large number of insertions in the case of large genomes.

Acknowledgements: We thank George Pelletier and Nicole Bechtold (INRA, Institut National de la Recherche Agranomique, Versaille, France) for giving us *Arabidopsis* T-DNA insertion lines. We thank also our colleague Gillian Hull for reading the manuscript.

References

[1] Salinas, J., Matassi, G., Montero, L.M. and Bernardi, G. (1988) Nucleic Acids Res. 19, 5561–5567.
[2] Matassi, G., Montero, L.M., Salinas, J. and Bernardi, G. (1989) Nucleic Acids Res. 17, 5273–5290.
[3] Montero, L.M., Salinas, J., Matassi, G. and Bernardi, G. (1990) Nucleic Acids Res. 18, 1859–1867.
[4] Carels, N., Barakat, A. and Bernardi, G. (1995) Proc. Natl. Acad. Sci. USA 92, 11057–11060.
[5] Barakat, A., Carels, N. and Bernardi, G. (1997) Proc. Natl. Acad. Sci. USA 94, 6857–6861.
[6] Barakat, A., Tran Han, D., Benslimane, A., Rode, A. and Bernardi, G. (1999) FEBS Lett. 463, 139–142.
[7] Carels, N., Hatey, P., Jabbari, K. and Bernardi, G. (1998) J. Mol. Evol. 46, 45–53.
[8] San Miguel, P., Tikhonov, A., Jin, Y.K., Moutchoulskaia, N., Zakharov, D., Melake-Berkane, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z. and Bennetzen, J.L. (1996) Science 274, 765–768.
[9] Barakat, A., Matassi, G. and Bernardi, G. (1998) Proc. Natl. Acad. Sci. USA 95, 10044–10049.
[10] Capel, J., Montero, L.M., Marinez-Zapater, J.M. and Salinas, J. (1993) Nucleic Acids Res. 21, 2369–2373.
[11] Rynditch, A.V., Zoubak, S., Tsyba, L., Tryapitsina-Guley, N. and Bernardi, G. (1998) Gene 222, 1–16.
[12] Zupan, J.R. and Zambryski, P. (1995) Plant Physiol. 107, 1041–1047.
[13] Weising, K., Bohn, H. and Kahl, G. (1990) Dev. Genet. 11, 233–247.
[14] Thomas, C.M., Jones, D.A., English, J.J., Caroll, B.J., Benetzen, J.L., Harrison, K., Burbidge, A., Bishop, G.J. and Jones, J.D.G. (1994) Mol. Gen. Genet. 242, 573–585.
[15] Hobbs, S.L.A., Kpodar, P. and Delong, M.O. (1990) Plant Mol. Biol. 15, 851–864.
[16] Ambros, P.F., Matzke, A.J.M. and Matzke, M.A. (1986) EMBO J. 5, 2073–2077.
[17] Chyi, Y.S., Jorgensen,.R.A., Goldstein, D., Tanksley, S.D. and Loaiza, F.F. (1986) Mol. Gen. Genet. 204, 64–69.
[18] Bernardi, G. (2000) Gene 3–17.
[19] Rivera-Madrid, R., Mestres, D., Marinho, P., Jacquot, J.P., Decottignies, P., Miginiac-Maslow, M. and Meyer, Y. (1995) Proc. Natl. Acad. Sci. USA 92, 5620–5624.
[20] Mestre, D. and Meyer, Y. (1999) Gene 240, 307–316.
[21] Bechtold, N., Ellis, J. and Pelletier, G. (1993) C.R. Acad. Sci. Paris 316, 1194–1199.
[22] Bouchez, D., Camilleri, C. and Caboche, M. (1993) C.R. Acad. Sci. Paris 316, 1188–1193.
[23] Hiei, Y., Ohta, S., Komari, T. and Kumashiro, T. (1994) Plant J. 6, 271–282.
[24] Jofuku, K.D. and Goldberg, R.B. (1998) in: Plant Molecular Biology: A Practical Approach (Shaw, C.H., Ed.), pp. 37–66, IRL, Oxford.
[25] Cortadas, J., Macaya, G. and Bernardi, G. (1997) Eur. J. Biochem. 76, 13–19.
[26] Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.
[27] Devic, M., Albert, S., Delseny, M. and Roscoe, T.J. (1997) Plant Physiol. Biochem. 35, 331–339.
[28] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[29] Koncs, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Körber, H., Redei, G.P. and Schell, J. (1989) Proc. Natl. Acad. Sci. USA 86, 8467–8471.