Volume 14 Number 20 1986 Nucleic Acids Research

A binary vector for transferring genomic libraries to plants

C.Simoens[1], Th.Alliotte[1], R.Mendel[2], A.Müller[2], J.Schiemann[2], M.Van Lijsebettens[1], J.Schell[1,3], M.Van Montagu[1] and D.Inzé[1]

[1]Laboratorium voor Genetica, Rijksuniversiteit Gent, B-9000 Gent, Belgium, [2]Zentralinstitut für Genetik und Kulturpflanzenforschung, Akademie der Wissenschaften der DDR, DDR-4325 Gatersleben, GDR and [3]Max-Planck-Institut der Züchtungsforschung, D-5000 Köln 30, FRG

Received 12 August 1986; Accepted 22 September 1986

SUMMARY

The transformation of mutant plants with a complete recombinant library derived from wild-type DNA followed by assay of transformed plants for complementation of the mutant phenotype is a promising method for the isolation of plant genes. The small genome of *Arabidopsis thaliana* is a good candidate for attempting this so-called shotgun transformation. We present the properties of an *A. thaliana* genomic library cloned in a binary vector, pC22. This vector, designed to introduce genomic libraries into plants, contains the *ori*V of the Ri plasmid pRiHRl by which it replicates perfectly stably in *Agrobacterium*. Upon transfer of the library from *E. coli* to *A. tumefaciens* large differences in transfer efficiencies of individual recombinant clones were observed. There is a direct relation between transfer efficiency and stability of the recombinant clones both in *E. coli* and *A. tumefaciens*. The stability is independent of the insert size, but seems to be related to the nature of the insert DNA. The feasibility of shotgun transformation and problems of statistical sampling are discussed.

INTRODUCTION

Agrobacterium tumefaciens is a bacterial plant pathogen causing crown gall tumours on most dicotyledonous plants. The tumorous growth is the result of the introduction of a well-defined region (T-region) of a tumour-inducing (Ti) plasmid into the plant nuclear genome. This transferred DNA (T-DNA) encodes genes which alter the normal plant hormone metabolism by introducing new auxin and cytokinin biosynthetic pathways. Apart from the T-region, another portion of the Ti plasmid, called the virulence (*vir*) region, is essential for tumour induction. This *vir* region encodes *trans*-acting functions which are likely to be involved in transfer of the T-region to plant cells (for a recent review, see [1]).

The T-region is flanked by 25-bp direct repeats, called borders [2-4]. The right border of the nopaline T-DNA is essential for transfer and/or integration of the T-region in the plant genome [5,6]. The left 25-bp repeat delimits the DNA to be transferred at the other end. Genetic studies

© IRL Press Limited, Oxford, England.

have shown that no T-DNA genes are required for transfer and integration [7-9]. Thus, any sequence situated between the 25-bp repeats can be transferred into plant cells in the presence of the vir functions. This property has been used to construct vectors specially designed to introduce foreign genes into plants.

A first generation of vectors was based on Ti plasmids in which the T-region is substituted for pBR322 sequences. A pBR derivative carrying the gene to be introduced into plants can be transferred into Agrobacterium where homology recombination between the two plasmids can result in a cointegrate which has the desired gene between the two borders [9,10].

A second generation of vectors does not rely on homology recombination. This is an important advantage for the proposed shotgun cloning of plant genes since the recombination step lowers the transfer efficiency of plasmids from E. coli to A. tumefaciens. This system of plant cell transformation requires two autonomously replicating plasmids in Agrobacterium. One plasmid contains the trans-acting vir functions, while the other one harbours the sequence to be introduced into plant cells between the T-DNA borders. The latter plasmid has been referred to as a "binary vector" [11]. Binary vectors replicate both in E. coli and A. tumefaciens and can easily be mobilized from E. coli to A. tumefaciens [12-15]. In this paper, we present the construction of a binary vector, called pC22, designed to introduce genomic libraries into plant cells.

We also describe the construction and properties of a partial Sau3A genomic library of Arabidopsis thaliana into pC22. A. thaliana has been proposed as a model for plant molecular research [16]. Since it has a particularly small haploid genome (about 7×10^7 bp), only a limited number of clones are needed to represent the complete genome. The A. thaliana genomic library into pC22 could be used in the future for gene isolation by complementation of well-characterized plant mutants. The statistical dispersion of the library during transfer from E. coli to Agrobacterium and subsequently to plants will be discussed.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phage clones

The bacterial strains, plasmids, and phage clones used in this work are listed in Table 1.

Media and antibiotics

The A. tumefaciens strains were grown at 28°C in YEB medium [25]; E.

Table 1. Bacterial strains, plasmids, and phage clones

	Antibiotic resistance	Relevant characteristics	Origin
Bacterial strains			
E. coli			
HB101	Str	*pro leu gal hsdR hsdM recA*	[17]
K803		r_K^- m_K^- *supE supF58 lacY gal met trp*	[18]
A. tumefaciens			
C58C1Rif^R	Rif	Rif^R derivative of C58, cured for pTiC58	[19]
Plasmids			
pC10	Sm Sp Cb	Binary vector with Sa origin of replication	this work
pHSG262	Km	Small cosmid vector	[20]
pLJbB11	Km	pHSG262 containing the 8.1-kb *Bam*HI-11 fragment of pRiHR1	[21]
pLGV23neo	Cb	pBR322 derivative containing a chimeric Pnos-Km^R gene	[22]
pGV2275	Km	pTiB6S3 derivative from which the total T-region, including the borders, is substituted by a *Bam*HI-*Eco*RI fragment carrying the Tn*5* Km^R gene	M. Dubois
pMP90	Gm	pTiC58 derivative, carrying a Gm^R marker, in which the total T-region, including the borders, is deleted	[23]
pGV2260	Cb	pTiB6S3 derivative from which the total T-region, including the borders, is substituted by pBR322 sequences	[10]
pGV831	Sm Sp Cb	pBR325 derivative carrying a Pnos-Km^R gene between T-DNA borders	[10]
pRK2013	Km	ColE1 replicon containing the RK2 transfer genes	[24]
Phage clones			
λbAT002		A λ clone carrying the 10-kb repeat unit of the large ribosomal DNAs (18S and 26S) of *A. thaliana*	[16]

Abbreviations : Cb, carbenicillin; Gm, gentamycin; Km, kanamycin; Rif, rifampicin; Sp, spectinomycin; Str, Sm, streptomycin.

coli strains were grown in LB medium at 37°C [26]. When necessary, YEB and LB media were solidified with 1.5% Difco agar. Concentrations of antibiotics were as described [27]. Plant tissue cultures were maintained on Linsmaier and Skoog (LS) medium [28]. Axenic tissue was obtained by adding cefotaxime, Claforan® (Hoechst), to a final concentration of 500 µg per ml culture medium. Kanamycin (Sigma) concentrations are expressed in mg/l kanamycin acid sulphate.

Mobilisation from E. coli to A. tumefaciens

Recombinant clones were mobilized from E. coli to A. tumefaciens in a triparental mating using HB101(pRK2013) as helper strain [29,30]. The E. coli strains were grown at 37°C in LB medium supplemented with antibiotics for which the resistance is encoded by the resident plasmids. The Agrobacterium acceptor strains were grown in YEB medium at 28°C for 24 hours. Equal amounts (1:1:1) of the three cultures (previously washed) were mixed and 100 µl was plated on solid LB medium. The conjugation mix was incubated at 33°C and dilutions were plated on YEB medium supplemented with 100 µg/ml rifampicin and 100 µg/ml spectinomycin. The transfer efficiencies were expressed as the number of transconjugants versus the total number of acceptor colonies. The total Arabidopsis genomic library was transferred as above, except that the library in E. coli was only grown for 8 hours at 37°C.

DNA preparations

For large and small scale plasmid DNA preparations, the procedure of [31] was used. Total plant DNA was prepared according to [32]. Chloroplast DNA of Arabidopsis thaliana (variety Columbia) was a generous gift of Dr. Redei (Columbia, U.S.A).

DNA manipulations

All DNA manipulations were conducted using standard procedures [18].

Construction of a cosmid library of A. thaliana

This was essentially done as described [18]. The cosmid vector pC22 (Figure 1) was cut with BamHI and dephosphorylated with calf intestinal alkaline phosphatase. Total Arabidopsis thaliana DNA was partially digested with Sau3A. An enriched fraction of high molecular weight fragments between 20 and 30 kb, obtained from a sucrose gradient, was ligated to a tenfold molar excess of dephosphorylated vector at a final concentration of 225 µg/ml. In vitro packaging into bacteriophage λ particles was carried out according to the recommendation of the manufacturer (Amersham). The packaging mixes were subsequently transduced to E. coli strain HB101 or K803. Recombinants were selected on LB medium with 100 µg/ml carbenicillin.

Plasmid stability test

Plasmid stability in both E. coli and A. tumefaciens was tested as follows. An overnight culture, grown in selective conditions (with 100 µg/ml of spectinomycin), was 1000-fold diluted in a non-selective LB medium. Subsequently, the overnight grown bacteria were plated on non-selective

agarmedia, and 100 colonies were tested for the presence of the spectinomycin and carbenicillin resistance markers of the resident plasmid. The overnight grown culture in non-selective medium was further diluted 1000-fold and tested as above, for two additional cycles of overnight growth.

Cocultivation

Cocultivation was as described [33]. Transformed calli were selected for kanamycin resistance as follows. After addition of cefotaxime, the protoplasts were allowed to grow for a further 3 weeks. The resulting small calli were then spread on an agar plate without kanamycin. After 2 weeks the growing calli were transferred individually onto an agar plate with 50 mg/l kanamycin. This method is used to calculate the transformation frequency as the ratio of the number of resistant calli to the total number of calli tested.

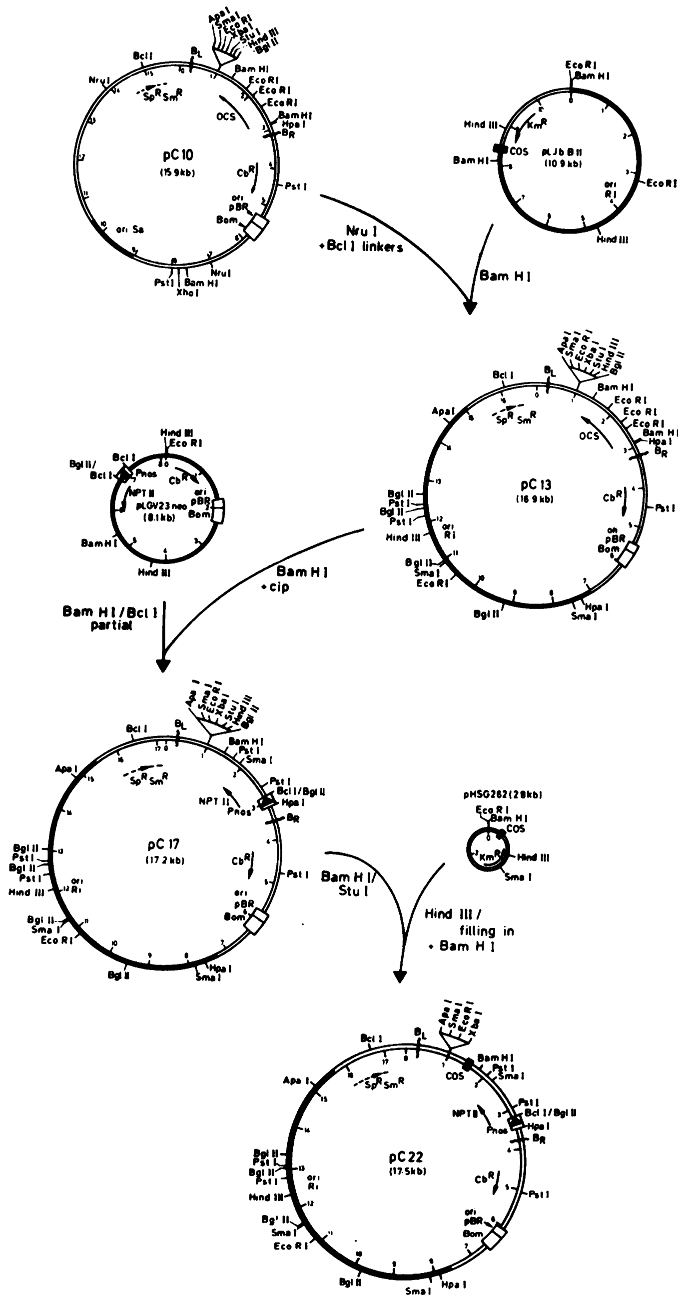
Leaf disc transformation

The leaf disc transformation method orginally described by [34] was modified. The infection medium was prepared by adding 0.5 ml of an overnight culture in LB medium to 10 ml liquid LS medium (hormone-free). Aseptically grown plants of *N. tabacum* cv. Gatersleben (= wild-type) served as source of leaf discs. Discs of about 6 mm x 5 mm were placed upside down on infection medium and kept at 26°C under low light intensity. After two days, the leaf discs were carefully washed with liquid LS medium and transferred on solid LS medium supplemented with 1 mg/l benzyl-6-aminopurine, 0.1 mg/l naphthalene acetic acid, 500 mg/l cefotaxime, and 50 mg/l kanamycin (shoot-inducing medium). Cultures were kept at 26°C under low light intensity.

After 4 weeks, the number of green shoot primordia formed from the leaf disc was determined. Six weeks after infection, the largest shoots were removed from the selection plates and rooted on LS medium supplemented with 0.1 mg/l indolyl butyric acid and 500 mg/l cefotaxime. Rooted plants were potted and grown to maturity in the greenhouse.

Seedling test for kanamycin resistance

Seeds were sterilized (70% EtOH, 3 minutes) and sown onto agar medium containing 1/4 concentration Murashige and Skoog salts [35] and 280 mg/l kanamycin. Unsealed petri dishes were cultured in containers at 16 hours light (2000 lux)/8 hours dark, by 26°C. Resistant and sensitive seedlings could be distinguished two weeks after sowing.


pC10
(15.9kb)
pLJbB11
(10.9kb)
Nru I
+Bcl I linkers
Bam H I
pC13
(16.9kb)
pLGV23neo
(8.1kb)
Bam H I
+cip
Bam H I/Bcl I
partial
pC17
(17.2kb)
pHSG262(2.8kb)
Bam H I/
Stu I
Hind III /
filling in
+ Bam H I
pC22
(17.5kb)

RESULTS

Components of pC22

The cosmid vector pC22 (Figure 1) is a binary vector especially designed to introduce genomic libraries into plant cells. We can distinguish two regions in pC22, one for bacterial maintenance and another for plant cell transformation. The first region allows pC22 to replicate stably both in *E. coli* and *A. tumefaciens*. This is due to the presence of the *ori*V of pBR322 and the Ri origin, respectively. The latter is derived from the Ri plasmid, pRiHR1, present in *Agrobacterium rhizogenes* and is compatible with the *ori*V of the Ti plasmid [36,37]. Mobilisation of pC22 from *E. coli* to *A. tumefaciens* is possible because of the presence of the pBR322 *bom* site [30]. Furthermore, pC22 harbours a carbenicillin and spectinomycin/streptomycin resistance gene, derived from pBR322 and R702, respectively [38].

The second region of pC22 is delimited by two borders derived from the octopine plasmid pTiB6S3 [39]. A chimeric gene, consisting of the coding region of the Tn*5* neomycin phosphotransferase II (*npt*-II) gene fused to the promoter of the nopaline synthase (*nos*) gene, allows selection for plant cell transformants [22]. Furthermore, there are two unique restriction sites (*Bam*HI and *Xba*I) and a λ *cos* sequence. The *Bam*HI site allows cloning of partial *Sau*3A genomic fragments. Since the vector is 17.5 kb in size, it is possible to clone, via *in vitro* packaging, DNA fragments ranging from 20 to 33 kb [40].

Figure 1. Construction of pC22.
pC10 is a binary vector containing both the origin of replication of the W-type plasmid Sa and the origin of replication of pBR322 (ori pBR) (unpublished data). The black lines indicate the origin of replications which are functional in *Agrobacterium* (oriSa, oriRi). The waved lines represent the left (B_L) and right (B_R) T-DNA borders. The position of the antibiotic resistance genes, abbreviated as in Table 1, are indicated with an arrow. Furthermore, this vector contains the octopine synthase (*ocs*) gene and a multilinker between the T-DNA borders. pC10 was digested with *Nru*I and these blunt-ends were converted to *Bcl*I sites by *Bcl*I linkers. The resulting *Bcl*I fragments were ligated to *Bam*HI-digested pLJbB11 DNA. In this way, we could substitute the Sa origin, present in pC10, for the origin of the Ri plasmid pRiHR1, resulting in pC13. A selectable marker for plant cell transformation was introduced into pC13 by replacing the octopine synthase (*ocs*) gene with a chimeric kanamycin resistance gene encoding neomycin phosphotransferase II (*npt*-II) under the control of the nopaline synthase promoter (Pnos). pC13 was digested with *Bam*HI, treated with calf intestinal alkaline phosphatase (CIP) and ligated to a partial *Bam*HI/*Bcl*I digestion of pLGV23neo. The resulting recombinant, called pC17, was finally adapted as genomic cloning vehicle by introduction of a λ *cos* sequence. The *Hind*III site of pHSG262 was filled in with DNA polymerase; then, the 0.8-kb blunt end/*Bam*HI fragment was ligated to *Bam*HI/*Stu*I-digested pC17 DNA. The resulting recombinant, called pC22, was chosen for further study.

Properties of pC22

In the E. coli strain K803 pC22 was shown to be perfectly stable even after 3 cycles of overnight growth in non-selective medium (see Materials and Methods). The plasmid pRK2013 is able to mobilize pC22 from E. coli to A. tumefaciens with an average frequency of 5×10^{-3}. No differences in transfer efficiencies were observed when the Agrobacterium strains, C58C1Rif^R, C58C1Rif^R(pGV2275), C58C1Rif^R(pMP90), were used as acceptor.

Since pC22 contains the origin of replication of an Ri plasmid, it was expected that this vector would be 100% stable in A. tumefaciens. Indeed, even after 3 cycles of overnight growth in non-selective medium of C58C1 Rif^R(pC22; pGV2275), C58C1Rif^R(pC22;pMP90) or C58C1Rif^R(pC22), no loss of pC22 or the Ti plasmid was observed. This is in contrast with the low stability of binary vectors, such as a precursor to pC22, pC10, that contain the Sa origin of replication. When following C58C1Rif^R(pC10) during 3 cycles of 24 hours non-selective growth, only 35% of the bacteria maintained the resident plasmid.

We transformed Nicotiana tabacum cv. SR1 mesophyll protoplasts with C58C1Rif^R(pC22;pGV2275) and C58C1Rif^R(pC22;pMP90), selecting transformed calli for kanamycin resistance. As a positive control C58C1Rif^R(pGV2260::pGV831) was used, a comparable construction except that in this case a cointegrate is formed between the Ti plasmid and the vector containing the T-region [10]. The results of the cocultivation experiments are shown in Table 2. The transformation frequencies can be considered normal when compared to the ones usually obtained by cocultivation, namely between 1 and 10%.

Leaf disc infection was used to obtain transformed N. tabacum plants. Transformation with C58C1Rif^R(pC22; pGV2275) and subsequent cultivation on shoot-inducing medium led to the formation of about 20 kanamycin-resistant

Table 2. Transformation frequencies obtained with pC22

Strain	Number of calli tested	Number of Km^R calli	Transformation frequency
C58C1Rif^R(pC22;pGV2275)	630	6	1%
C58C1Rif^R(pC22;pMP90)	420	20	5%
C58C1Rif^R(pGV2260::pGV831)	610	31	5%

The cocultivations were performed as described in Materials and Methods.

shoot buds per leaf disc. Comparable transformation efficiencies were obtained with the strain C58C1Rif^{R}(pGV2260::pGV831). Kanamycin-resistant shoots were rooted under non-selective conditions and grown to mature plants.

Of 16 plants tested, 12 were found to be kanamycin-resistant whereas the other 4 regenerated plants were sensitive to kanamycin. The high incidence of kanamycin-sensitive escapes has also been reported for leaf disc transformation with other Ti plasmid-derived vectors [34].

For one of the pC22-transformed plants, designated as T3, the level of kanamycin resistance was determined. The data given in Figure 2 show that the leaf tissue of this plant was approximately 13 times more resistant to kanamycin than that of wild-type plants.

Crossing analysis of a transformed plant

The kanamycin-resistant plant T3 was morphologically normal and fully fertile. The seed progeny obtained from self-fertilized flowers as well as those obtained from reciprocal back-crosses to a wild-type plant segregated into sensitive and resistant seedlings, which could be distinguished when germinated and grown on minimal medium containing 280 mg/l kanamycin. The sensitive seedlings had etiolated cotyledons and did not grow, whereas the resistant seedlings had green cotyledons and grew (although their growth rate was much reduced by the high concentration of kanamycin).

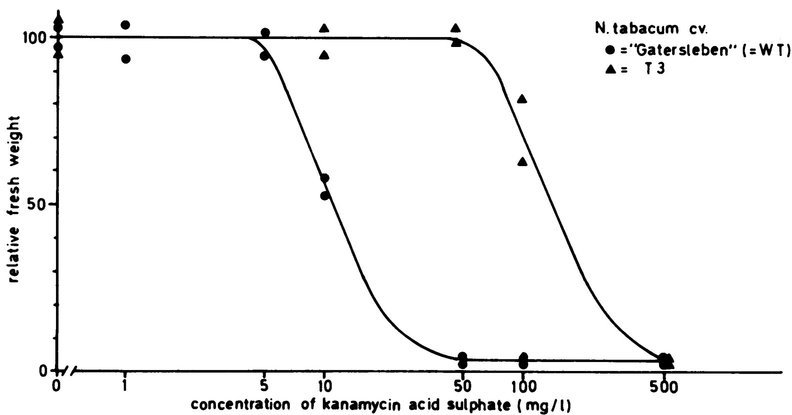


Figure 2. Determination of kanamycin resistance level.
Leaf discs were grown on medium with various concentrations of kanamycin for 28 days. The final fresh weights compared to a control grown without kanamycin were plotted against the corresponding kanamycin level.

Table 3. Segregation for kanamycin resistance among progeny of crosses involving the transformed *Nicotiana tabacum* plant T3

Cross	Number of seedlings			Expected segregation ratio
	total	resistant	sensitive	
T3 selfed	316	240	76	3:1 (237:79)
WT x T3	442	221	221	1:1 (221:221)
T3 x WT	320	164	156	1:1 (160:160)

The differences between the observed and expected ratios are not significant at the 5% level.

The segregation ratios observed were very close to a 3:1 (self-fertilized progeny) or a 1:1 ratio (back-cross progeny) (Table 3). Thus, the transformed phenotype of T3 was inherited as a hemizygous dominant Mendelian trait. This shows that the chimeric *npt*-II gene was stably integrated at a single chromosomal locus, and that it was transmitted in a regular manner to the sexual offspring.

Construction and analysis of a genomic library of *Arabidopsis thaliana*

A genomic library of *Arabidopsis thaliana* (variety Columbia) was made using total DNA prepared from greenish callus sprouting numerous shoots. This callus tissue was grown axenically on LS medium supplemented with 1 mg/l 2,4-D and 0.05 mg/l kinetin. *E. coli* K803 was chosen as acceptor strain since it has been reported that this strain gave a high transduction efficiency for λ clones containing plant DNA inserts [41]. With K803 we reached an efficiency of 6.5 x 10^5 colonies per µg DNA, 13 times more than with HB101. Two possible explanations are that HB101 selects more strongly against certain structural motifs (such as inverted-repeats), or that heavily modified (methylated) DNAs are more easily established in K803 than in HB101, thus resulting in a higher plating efficiency.

In order to determine the mean insert size, 50 genomic clones present in K803 were digested with *Hin*dIII and *Eco*RI. Clones which contain sequences that cross-hybridize to the large ribosomal RNAs (18S and 26S) were excluded since a high frequency of rearrangements of the insert DNA was observed (see further). The mean insert size was calculated to be 23.6 kb. Seven clones had an insert between 15 and 20 kb; 30 clones ranged from 20- to 25 kb insert size; 8 clones reached an insert size between 25 and 30 kb; and 5 clones had an insert from 30 kb up to 34 kb. Taking into account the

Table 4. Stability of A. thaliana genomic clones in E. coli and A. tumefaciens, and mobilization frequencies to A. tumefaciens

Plasmid	Insert size (kb)	Stability (%) in E. coli				Stability (%) in A. tumefaciens				Mobilization frequency
		+ Sp	-Sp 1	-Sp 2	-Sp 3	+ Sp	-Sp 1	-Sp 2	-Sp 3	
in HB101										
pAGC4	30	87	72	12	8	70	21	5	1	2×10^{-6}
pAGC7	20	100	96	89	91	99	99	97	96	3×10^{-3}
pC22	--	100	100	99	99	100	98	99	99	7×10^{-3}
in K803										
pAGC9	24	72	53	20	7	33	10	0	1	3×10^{-6}
pAGC11	33	90	84	65	47	95	69	59	44	2×10^{-5}
pAGC12	16	99	97	95	94	100	98	100	100	2×10^{-3}
pC22	--	100	99	99	98	100	100	99	100	6×10^{-3}

Stability tests were performed as described in Materials and Methods. The presence of the spectinomycin and carbenicillin resistance marker of pC22 was scored after an overnight culture in LB medium supplemented with 100 µg/ml Sp and three subsequent cycles of overnight growth in non-selective medium (indicated as 1, 2, and 3). The mobilization frequencies were determined as described in Materials and Methods. In all conjugations HB101(pRK2013) was used as a helper strain and C58C1Rif^R as an acceptor strain.

number of clones (10%) which contain chloroplast DNA (data not shown), only approximately 15,000 individual clones in E. coli K803 are required to represent the complete Arabidopsis genome with a probability of 99% [42].

One hundred recombinant clones with varying insert sizes were tested for stability in E. coli during one cycle of overnight growth as described in Materials and Methods. The results led to the conclusion that approximately 40% of the clones are stable, whereas the other 60% showed various degrees of instability. This instability resulted always in the loss of both the spectinomycin and carbenicillin resistance marker of pC22. Furthermore, we could not observe any relation between insert size and stability. Table 4 shows the stability of five typical recombinant clones. It is clear that although pAGC11 has a larger insert than pAGC9, it is never-

theless more stable. This observation suggests that the nature of the sequences which are inserted is a more important factor to restrict stability than the insert size.

An extensive restriction analysis of several hundreds of both stable and unstable recombinant clones, did not reveal any DNA rearrangements. For example, EcoRI digests always resulted in three fragments that cross-hybridize to vector DNA : an internal 7.4-kb fragment and two composite fragments with a minimum size of 9.3 kb and 0.8 kb (data not shown). However, clones that cross-hybridize to the large ribosomal DNA repeat unit (18S and 26S) of A. thaliana showed an extensive deletion formation of the insert DNA.

It has been demonstrated [16] that there are approximately 570 copies fo the large ribosomal DNA repeat unit per haploid A. thaliana genome. Each repeat unit is about 10 kb in length and the repeats are largely arranged in tandem array. Approximately 1.75% of the genomic pC22 clones (20 out of 1145 analyzed) cross-hybridized to the 7-kb SalI fragment of λbAT002, which contains the 18S and 26S rDNA. Upon restriction analysis of individual, purified colonies originating from one original recombinant clone, large differences in the insert DNA were noticed. Most subclones contain maximally one repeat unit of the ribosomal DNA, which is characterized by the presence of a 3.8-kb, 2.35-kb, 1.6-kb, 1.5-kb, and a 0.65-kb EcoRI fragment (data not shown). This suggests that the insert DNA has been rearranged by homologous recombination of the tandemly repeated ribosomal DNA units.

From Table 4 we can also conclude that the instability observed with some clones is independent of the chromosomal background. Clones which are unstable in E. coli are approximately as unstable in Agrobacterium. Furthermore, there is a direct relationship between stability in E. coli and A. tumefaciens and mobilization frequency from E. coli to A. tumefaciens. In general one can say that depending on stability the mobilization frequencies can vary from 10^{-3} to 10^{-6}. The mobilisation of the whole library (50,000 pooled clones) from E. coli to A. tumefaciens was possible with a frequency of 10^{-4}.

Statistical dispersion of the library during transfer to Agrobacterium and plants

Each time the library is pooled and amplified, a statistical dispersion causes that more clones must be collected in order to have a high chance either to maintain every clone of the initial library or to avoid

the loss of a single clone that is to be selected after the consecutive transfers.

(a) The problem of sampling was first analyzed for the case that a complete librabry has to be transferred from E. coli to Agrobacterium. Only the theoretical aspects of sampling are considered : biological problems of clone stability and of differences in transfer efficiencies of individual clones are not taken into account. By definition, K is the number of clones in E. coli (= size of the library) and N the number of Agrobacterium transconjugants. It is assumed that all K clones have the same probability to be selected out of an unlimited E. coli population. The problem is to find the minimal number N necessary to have a probability of 95% that every clone is at least once transferred to Agrobacterium. If χ_i is the event "clone i is not transferred", then the probability that at least one clone is not transferred is : $P(\chi_1 + \chi_2 + \ldots + \chi_K) =$

$$\Sigma P(\chi_i) - \Sigma P(\chi_i\chi_j) + \chi P(\chi_i\chi_j\chi_s) - \ldots. + (-1)^{K+1}(P\chi_1\chi_2 \cdot\cdot \chi_K)$$

(development of the general rule of addition in probability calculation)

$$= \Sigma_{j=1}^{K} P(j)(-1)^{j+1}$$

where P(j) is the probability that j clones are not transferred

$$P(j) = C_K^j \left(\frac{K-j}{K}\right)^N$$

The probability that in N independent transfers to individual acceptor bacteria a subset of j clones out of the original K clones is not transferred is $\left(\frac{K-j}{K}\right)^N$; C_K^j is the number of all possible subsets of size j (coefficient of the binomial law). The P(j) are the terms of a decreasing series and were calculated by computer in the case of K varying between 10^3 and 10^5 and N varying between 5K and 20K. The results (data not shown) can be summarized as follows. The $N_{0.95}$ values, i.e. the values of N that assure with a probability of 95% that a complete set of K different individuals are collected out of a pool, are 10K for $K = 10^3$, 12K for $K = 10^4$, and 14K for $K = 10^5$. For K = 15,000, thus the size of the Arabidopsis library in E. coli, $N_{0.95}$ was calculated to be 1.8×10^5. The above theoretical calculations were confirmed by computer-assisted simulation experiments (J. Coppieters, unpublished data).

(b) The requirement that no clone is lost during the transfer of the library from E. coli to A. tumefaciens is very stringent and demands a large amplification. This can especially be applied to the situation where one pool of Agrobacterium transconjugants has to serve as a basis for many shotgun transformations. However, most experiments do not require transfer

of the complete library, in particular each cocultivation experiment between the pool of Agrobacterium transconjugants and plant cells is performed with the purpose of selecting a specific gene. The probability that one particular clone is present in a sample taken at random of size N out of a pool containing K clones is :

$$1 - \left(\frac{K-1}{K}\right)^N$$

This probability was calculated for different K and N values. The results are that for N = 1K the probability is 0.63, for 3K 0.95, and for 5K 0.99.

We can conclude that for $K = 1.8 \times 10^5$ (i.e. the size of the representative library in Agrobacterium) 5.4×10^5 independently transformed plant cells are needed to have a 95% probability of a specific clone being recovered.

Throughout the statistical calculations, we always assumed that every clone had the same probability to be transferred from E. coli to Agrobacterium as from Agrobacterium to plants. Since, 60% of the clones in the library have various degrees of instability, the obtained minimal numbers of N are underestimated.

DISCUSSION

In yeast many genes have been successfully isolated via complementation of mutants [43,44]. In this paper, we described the basis of adapting this method for plants. In order to make shotgun cloning of plant genes by complementation of mutants feasible, it is necessary to limit the number of transformed plant cells required to ensure that any plant DNA sequence present in the genomic library is transferred. Arabidopsis thaliana is a good model plant for attempting this method. The haploid genome of this plant is extremely small (70,000 kb) which consequently reduces the number of clones needed to represent it. Furthermore, the Arabidopsis genome contains only 10 to 14 % of highly repetitive sequences and the unique sequences are arranged in large continuous regions (up to 140 kb) [16].

We constructed a partial Sau3A genomic library of A. thaliana (variety Columbia) in the binary vector pC22. This vector plasmid was especially designed to clone large genomic plant DNA fragments and subsequently to introduce these fragments via the Agrobacterium system into plant cells. Basically, this vector differs from other previously described binary vectors by the presence of the origin of replication of the Ri-plasmid pRiHR1. In contrast to binary vectors with the Sa or RK2 origin of replication, pC22 replicates perfectly stably in Agrobacterium [14; unpublished

results]. Since the original Ri-plasmid is 200 kb in size we assumed that pC22 recombinants with large inserts would be stable in Agrobacterium. The pC22 vector plasmid can be mobilized from E. coli to A. tumefaciens at a frequency as high as 7×10^{-3}. To transfer subsequently plant DNA fragments from Agrobacterium into plant mutants, pC22 contains a left and a right T-DNA border, and in addition a chimeric kanamycin resistance gene. We have shown, both by cocultivation and leaf disc transformation, that pC22 transforms plant cells at normal frequencies. A preliminary study of ten Nicotiana tabacum plants transformed with pC22 shows the presence of 1 to 3 independently segregating T-DNA loci [Müller et al., in preparation]. The presence of a λ cos sequence and two unique restriction sites (XbaI and BamHI) allows the cloning of large genomic DNA fragments between the T-DNA borders. The mean insert size of the Arabidopsis genomic library we constructed was estimated to be 23.6 kb.

Theoretically, taking into account the mean insert size and the number of clones (10 %) which contain chloroplast DNA, only approximately 15.000 individual clones in E. coli K803 are needed to represent the complete genome with a probability of 99 %. The strategy for plant gene isolation by shotgun transformation requires that every genomic clone is transferred with comparable efficiencies from E. coli to A. tumefaciens, and subsequently from Agrobacterium to plants. Therefore, we studied the stability and transfer efficiencies of a large number of recombinant clones in pC22. We estimated that approximately 40% of the clones are stable, whereas the other 60% showed various degrees of instability in the absence of selection. Clones which were unstable in E. coli were approximately as unstable in Agrobacterium. Under selective conditions, unstable clones can be easily maintained in E. coli and A. tumefaciens. Preliminary results have shown that the stability is unaffected when unstable recombinant plasmids were introduced in the recB, recC, sbcB E. coli mutant CES200 [45]. The stability was shown to be independent of the insert size, and seemed to be related to the nature of the insert DNA. Some very stable clones contain inserts larger than 30 kb, whereas some unstable clones contain inserts below 20 kb. The observed plasmid instability likely results from the complete loss of the plasmid during replication since the carbenicillin and spectinomycin resistance markers of pC22 are always simultaneously lost and since no DNA rearrangements were found up to now. An exception are clones which contain ribosomal DNA (18S and 26S). These clones rearrange very frequently, due to homologous recombination of the tandemly repeated ribo-

somal DNA units. In every case the vector remained intact. The nature of the Arabidopsis sequences which cause instability is not yet known. Although selection for recombinants in E. coli and A. tumefaciens ensures their maintenance, any manipulation of the library under nonselective conditions will favour the presence of some stable clones.

There is a direct relationship between the stability and the mobilization frequency of the genomic clones from E. coli to A. tumefaciens. The mobilization frequencies can vary between 10^{-3} and 10^{-6}. Consequently, when the library as a whole is transferred from E. coli to A. tumefaciens, some clones will transfer much more efficiently than others. We have analyzed the statistical dispersion of the library during transfer to Agrobacterium and plants. When the library (represented by 15,000 clones) is mobilized from E. coli to Agrobacterium every clone has to be present in the pool of Agrobacterium transconjugants since this serves as a basis for further plant cell transformation experiments. It was calculated that, assuming every clone transfers with the same efficiency, 180,000 transconjugants in Agrobacterium are needed to have a probability of 95% that each of the initial 15,000 clones is present.

For a given shotgun transformation experiment the requirements are less stringent since in this case only one specific clone has to be selected. It was estimated that it is sufficient to select three times more plant cell transformants (5.4×10^{5}) than Agrobacterium transconjugants to obtain a 95% probability that a given sequence is transferred to plants. Thus in conclusion, the two consecutive transfers demands that the original small number of 15,000 has to be multiplied by 36. This number is still feasible to work with. However, since only 40% of the clones in E. coli transfer with high efficiency to Agrobacterium, and since the remaining 60% of the clones have a lower transfer efficiency, one cannot be sure that the unknown sequence (gene) that has to be selected will be present in the pool of Agrobacterium transconjugants. Thus, shotgun transformation of plant cells remains risky, as long as the problem of the differences in transfer efficiencies has not been solved.

The statistical analysis has also clearly demonstrated that shotgun transformation with genomic libraries made from other plants than Arabidopsis will be extremely difficult. When following the same reasoning as above, we calculated that for example for a library of Nicotiana tabacum, consisting of 3.7×10^{5} clones, approximately 1.7×10^{7} plant cell trans-

formants are needed to have a probability of 95% that a particular sequence is transferred.

Acknowledgements

The authors wish to thank Dr. A. Depicker and Dr. A. Caplan for critical reading of the manuscript, D. Valvekens for technical assistance, and A. Verstraete, K. Spruyt, and M. De Cock for help in preparing the manuscript. We appreciated very much the help of Prof. Dr. H. Steyaert, Dr. N. Baeck, Dr. P. Verbeek (Sterrekundig Observatorium, Rijksuniversiteit Gent, Belgium), and of J. Coppieters with the statistical analysis. This work was supported by grants of the "ASLK-Kankerfonds", the "Fonds voor Wetenschappelijk Geneeskundig Onderzoek" (F.G.W.O. 3.0001.82), and the Services of the Prime Minister (O.O.A. 12.0561.84), and partially by L'Air Liquide S.A. (France). CS is a Research Assistant and DI a Senior Research Assistant of the National Fund for Scientific Research (Belgium).

REFERENCES

1. Gheysen, G., Dhaese, P., Van Montagu, M. and Schell, J. (1985) In Genetic flux in plants (Advances in Plant Gene Research, Vol. 2), Hohn, B. and Dennis, E.S. Eds., pp. 11-47, Springer Verlag, Wien.
2. Simpson, R.B., O'Hara, P.J., Kwok, W., Montoya, A.L., Lichtenstein, C., Gordon, M.P. and Nester, E.W. (1982) Cell 29, 1005-1014.
3. Yadav, N.S., Vanderleyden, J., Bennett, D.R., Barnes, W.M. and Chilton, M.-D. (1982) Proc. Natl. Acad. Sci. USA 79, 6322-6326.
4. Zambryski, P., Depicker, A., Kruger, K. and Goodman, H. (1982) J. Mol. Appl. Genet. 1, 361-370.
5. Shaw, C.H., Watson, M.D., Carter, G.H. and Shaw, C.H. (1984) Nucl. Acids Res. 12, 6031-6041.
6. Wang, K., Herrera-Estrella, L., Van Montagu, M. and Zambryski, P. (1984) Cell 38, 455-462.
7. Leemans, J., Deblaere, R., Willmitzer, L., De Greve, H., Hernalsteens, J.P., Van Montagu, M. and Schell, J. (1982) EMBO J. 1, 147-152.
8. Joos, H., Inzé, D., Caplan, A., Sormann, M., Van Montagu, M. and Schell, J. (1983) Cell 32, 1057-1067.
9. Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M. and Schell, J. (1983) EMBO J. 2, 2143-2150.
10. Deblaere, R., Bytebier, B., De Greve, H., Deboeck, F., Schell, J., Van Montagu, M. and Leemans, J. (1985) Nucl. Acids. Res. 13, 4777-4788.
11. Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J. and Schilperoort, R.A. (1983) Nature (London) 303, 179-181.
12. Bevan, M. (1984) Nucl. Acids Res. 12, 8711-8721.
13. An, G., Watson, B.D., Stachel, S., Gordon, M.P. and Nester, E.W. (1985) EMBO J. 4, 277-284.
14. Klee, H.J., Yanofsky, M.F. and Nester, E.W. (1985) Bio/technology 3, 637-642.
15. van den Elzen, P., Lee, K.Y., Townsend, J. and Bedbrook, J. (1985) Plant Mol. Biol. 5, 149-154.
16. Pruitt, R.E. and Meyerowitz, E.M. (1986) J. Mol. Biol. 187, 169-183.
17. Boyer, H.W. and Roulland-Dussoix, D. (1969) J. Mol. Biol. 41, 459-472.

18. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning, a laboratory manual, pp. 1-545, Cold Spring Harbor Laboratory, New York.
19. Van Larebeke, N., Engler, G., Holsters, M., Van den Elsacker, S., Zaenen, I., Schilperoort, R.A. and Schell, J. (1974) Nature (London) 252, 169-170.
20. Brady, G., Jantzen, H.M., Bernard, H.U., Brown, R., Schütz, G. and Hashimoto-Gotoh, T. (1984) Gene 27, 223-232.
21. Jouanin, L. (1984) Plasmid 12, 91-102.
22. Herrera-Estrella, L., De Block, M., Messens, E., Hernalsteens, J.-P., Van Montagu, M. and Schell, J. (1983) EMBO J. 2, 987-995.
23. Koncz, C. and Schell, J. (1986) Mol. Gen. Genet., in press.
24. Figurski, D. and Helinski, D.R. (1979) Proc. Natl. Acad. Sci. USA 76, 1648-1652.
25. Vervliet, G., Holsters, M., Teuchy, H., Van Montagu, M. and Schell, J. (1975) J. Gen. Virol. 26, 33-48.
26. Miller, J.H. (1972) Experiments in Molecular Genetics, pp. 1-466, Cold Spring Harbor Laboratory, New York.
27. Leemans, J., Shaw, C., Deblaere, R., De Greve, H., Hernalsteens, J.-P., Maes, M., Van Montagu, M. and Schell, J. (1981) J. Mol. Appl. Genet. 1, 149-164.
28. Linsmaier, E.M. and Skoog, F. (1965) Physiol. Plantarum 18, 100-127.
29. Ditta, G., Stanfield, S., Corbin, D. and Helinski, D.R. (1980) Proc. Natl. Acad. Sci. USA 77, 7347-7351.
30. Van Haute, E., Joos, H., Maes, M., Warren, G., Van Montagu, M. and Schell, J. (1983) EMBO J. 2, 411-418.
31. Birnboim, H.C. and Doly, J. (1979) Nucl. Acids Res. 7, 1513-1523.
32. Lemmers, M., De Beuckeleer, M., Holsters, M., Zambryski, P., Depicker, A., Hernalsteens, J.-P., Van Montagu, M. and Schell, J. (1980) J. Mol. Biol. 144, 353-376.
33. De Block, M., Herrera-Estrella, L., Van Montagu, M., Schell, J. and Zambryski, P. (1984) EMBO J. 3, 1681-1689.
34. Horsch, R.B., Fry, J.E., Hoffmann, N.L., Eichholtz, D., Rogers, S.G. and Fraley, R.T. (1985) Science 227, 1229-1231.
35. Murashige, T. and Skoog, F. (1962) Physiol. Plantarum 15, 473-497.
36. Costantino, P., Hooykaas, P.J.J., den Dulk-Ras, H. and Schilperoort, R.A. (1980) Gene 11, 79-87.
37. White, F.F. and Nester, E.W. (1980) J. Bacteriol. 144, 710-720.
38. Hedges, R.W. and Jacob, A. (1974) Mol. Gen. Genet. 132, 31-40.
39. Holsters, M., Villarroel, R., Gielen, J., Seurinck, J., De Greve, H., Van Montagu, M. and Schell, J. (1983) Mol. Gen. Genet. 190, 35-41.
40. Feiss, M., Fisher, R.A., Crayton, M.A. and Egner, C. (1977). Virology 77, 281-293.
41. Fedoroff, N. (1983) Plant Mol. Biol. Reporter 1, 27-29.
42. Clarke, L. and Carbon, J. (1976) Cell 9, 91-99.
43. Petes, T.D. (1980) Ann. Rev. Biochem. 49, 845-876.
44. Rine, J., Hansen, W., Hardeman, E. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 6750-6754.
45. Wyman, R.A., Wolfe, L.B. and Botstein, D. (1985). Proc. Natl. Acad. Sci. USA 82, 2880-2884.