nature biotechnology

T-DNA–mediated transfer of *Agrobacterium tumefaciens* chromosomal DNA into plants

Bekir Ülker[1,4], Yong Li[2,4], Mario G Rosso[2,4], Elke Logemann[1], Imre E Somssich[1] & Bernd Weisshaar[2,3]

Besides the well-documented integration of DNA flanked by the transfer DNA borders, occasional insertion of fragments from the tumor-inducing plasmid into plant genomes has also been reported during *Agrobacterium tumefaciens*–mediated transformation. We demonstrate that large (up to ~18 kb) gene-bearing fragments of *Agrobacterium* chromosomal DNA (*A*chrDNA) can be integrated into *Arabidopsis thaliana* genomic DNA during transformation. One in every 250 transgenic plants may carry *A*chrDNA fragments. This has implications for horizontal gene transfer and indicates a need for greater scrutiny of transgenic plants for undesired bacterial DNA.

Agrobacterium tumefaciens is a soil-borne bacterial pathogen of plants. In nature, *Agrobacterium* transfers a defined segment of the tumor-inducing (Ti) plasmid (T-DNA) into the host, leading to the formation of crown gall tumors controlled by T-DNA–encoded oncogenes[1]. *Agrobacterium*-mediated DNA transfer has been exploited to introduce transgenes into plants and to transform other organisms such as yeast, fungi and even human cells[2]. Sometimes, part of the Ti plasmid outside the T-DNA borders may be integrated into plant genomes[3]. The *A. tumefaciens* strain C58 genome of 5.7 megabases has been completely sequenced and comprises four replicons: a linear chromosome, a circular chromosome and the two large plasmids AtC58 and TiC58 (refs. 4,5).

While characterizing a T-DNA insertion locus named PM within the fully sequenced *A. thaliana* genome, we discovered a 322-bp DNA fragment of nonplant origin associated with the right border (RB) of the T-DNA. Our discovery that this sequence is identical to a region on the sequenced linear chromosome of *A. tumefaciens* led us to determine whether this was a unique event or whether it is an intrinsic property associated with T-DNA transfer in general. Therefore, we analyzed databases that contain *A. thaliana*–flanking sequence tags (FSTs)[6–9], the sequences that flank T-DNA insertion sites in populations of insertion lines generated to saturate the genome with mutations. Fragments of *A*chrDNAs were detected in all tested T-DNA insertion databases, and *A*chrDNAs were found much more frequently in FSTs recovered from the RB (**Table 1**). Based on these data, gathered from >375,000 T-DNA–tagged *A. thaliana* lines, we estimated that ~0.4% (from the RB FSTs of GABI-Kat) of the insertion sites contain bacterial chromosomal DNA. The different populations studied have been generated with different T-DNA vectors and *A. tumefaciens* strains, indicating that fragments of *A*chrDNA are transferred to the plant genome irrespective of the binary vector or *A. tumefaciens* strain used. In addition, we have studied rice FST collections and also detected *A*chrDNA sequences, indicating that transfer of *A*chrDNAs through *Agrobacterium* happens in rice as well.

To determine the source of the *A*chrDNA and how it is transferred to plants, we analyzed in detail several bacterial DNA–carrying *A. thaliana* lines from the GABI-Kat T-DNA–tagged population (**Fig. 1a–c**). As the *bla* gene, which confers ampicillin resistance in *Escherichia coli*, is present in the T-DNA inserts in these plants, we rescued the T-DNA and flanking sequences from these lines using the restriction enzyme *Spe*I (which does not cut within the T-DNA or the vector backbone) and sequenced through the insertion sites. Genomic DNA from GABI-Kat line 086C02 contains an 18.5-kb *A*chrDNA insert at the RB. It comprises 2 partial and 16 complete bacterial open reading frames (ORFs; Atu4208–Atu4226) from the linear chromosome. Segregation analyses in 086C02 T3 plants demonstrated linkage and mendelian inheritance of the T-DNA and the *A*chrDNA (**Supplementary Fig. 1** online). For GABI-Kat line 052H10, the insertion was

Table 1 Number of FSTs containing sequences from the *A. tumefaciens* linear or circular chromosome

	FLAGdb		SAIL		GABI-Kat		SIGnAL	
	LB	RB	LB	RB	LB	RB	LB	RB
No. of FSTs analyzed	30,379	5,908	57,429	0	122,078	3,028	168,081	0
No. of linear chromosome hits	33	31	11	–	34	13	27	–
No. of circular chromosome hits	2	3	2	–	7	1	3	–

The sequences of the GABI-Kat FSTs included in the table are provided as a sequence file, for the other FSTs the EMBL/GenBank accession numbers or SALK IDs are listed in **Supplementary Data** online. FST, flanking sequence tag; LB, left border; RB, right border.

[1]Max Planck Institute for Plant Breeding Research, Department of Plant Microbe Interactions, Carl-von-Linne-Weg 10, D-50829 Cologne, Germany. [2]Institute for Genome Research and Systems Biology, Bielefeld University, Universitaetsstrasse 25, D-33594 Bielefeld, Germany. [3]Max Planck Institute for Plant Breeding Research, Department of Molecular Plant Genetics, Carl-von-Linne-Weg 10, D-50829 Cologne, Germany. [4]Present addresses: School of Biological and Biomedical Sciences, Durham University, Science Site, South Road, Durham DH1 3LE, UK (B.U.); University of Freiburg, FRISYS Core Facility Data Management, Habsburgerstrasse 49, D-79104 Freiburg, Germany (Y.L.); Leibniz-Institute of Plant Genetics and Crop Plant Research, Department of Molecular Genetics, Corrensstrasse 3, D-06466 Gatersleben, Germany (M.G.R.). Correspondence should be addressed to B.W. (bernd.weisshaar@uni-bielefeld.de) or B.Ü. (dr.bulker@gmail.com).

Received 11 June; accepted 31 July; published online 31 August 2008; doi:10.1038/nbt.1491

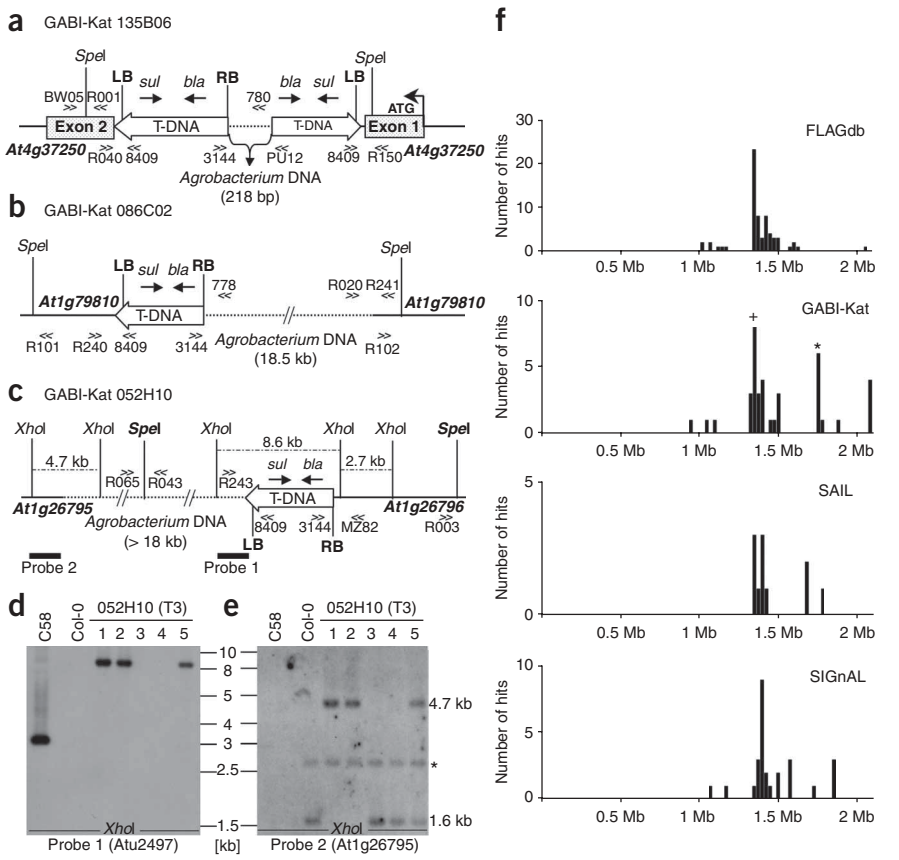

Figure 1 Determination of T-DNA insertion site structures within the *A. thaliana* genome using plasmid rescue. (**a**) An inverted-repeat T-DNA insertion site in the GABI-Kat line 135B06. (**b**) T-DNA insertion site in the GABI-Kat line 086C02. (**c**) Insertion site from GABI-Kat line 052H10. LB, left border; RB, right border; *bla*, *beta-lactamase* gene conferring resistance to ampicillin in *E. coli*; *sul*, plant selectable marker gene conferring resistance to sulfadiazine. Dotted lines indicate the *A*chrDNA fragment. Crucial primer annealing sites for sequencing are indicated; 8409 anneals close to LB, 3144 anneals close to RB, the other primers anneal at the indicated positions (**Supplementary Methods** online). >> and << show the direction of the primers used for sequencing reactions. *Spe*I restriction sites were used for plasmid rescue. (**d,e**) Blot analyses of DNA isolated from *A. tumefaciens* strain C58, wild-type *A. thaliana* Col-0 plants and five T_3 progeny of line 052H10 digested with *Xho*I. Locations of the specific ^{32}P-labeled probes are indicated in c as thick, solid lines. The blot in d was hybridized with probe 1 detecting *A*chrDNA locus Atu2497 and then subsequently reprobed (**e**) with probe 2 detecting the interrupted plant locus At1g26795. The T_3 052H10 plants 1 and 2 are homozygous for the insertion (8.6 kb and 4.7 kb respective fragments), whereas plant 5 is heterozygous. T_3 052H10 plants 3 and 4 only carry the 1.6 kb endogenous wild-type locus also detected in Col-0 plants. The asterisk in e marks the cross-hybridizing band of the neighboring *A. thaliana* locus (At1g26796) sharing 93% sequence identity to probe 2. The 3.0 kb *Xho*I DNA fragment detected with probe 1 in *A. tumefaciens* C58 is consistent with the expected size as deduced from the bacterial genome. (**f**) Distribution of FST hits on the linear chromosome of *A. tumefaciens*. The *y* axes represent the number of FST hits from each FST collection to the *A. tumefaciens* linear chromosome (*x* axes in million base pairs (Mb)). The numbers of hits are 64 for FLAGdb[6], 38 for GABI-Kat[8], 11 for SAIL[7] and 27 for SIGnAL[9]. +, the *A*chrDNA fragment found in GABI-Kat line 086C02 and several other lines in both GABI-Kat and other populations. *, the IS element sequences found in GABI-Kat lines 135B06, 146B12 and 343H01. See **Supplementary Data** for the individual FSTs plotted.

associated with a >18-kb *A*chrDNA originating from the circular chromosome. It contains 2 partial and 18 complete bacterial ORFs (Atu2496-Atu2516) at the T-DNA left border (LB).

Genomic DNA blot analyses of selected *A. thaliana* T-DNA insertion lines unequivocally demonstrated physical linkage of the *A*chrDNA to specific plant loci. Hybridization of DNA extracted from individual T3 plants of the 052H10 line with probes specific to the circular chromosome of *Agrobacterium* and to the *A. thaliana* locus At1g26795 revealed linkage of the *A*chrDNA both to the T-DNA and to plant genomic DNA (**Fig. 1c–e**). Similar results were obtained with the GABI-Kat lines 086C02 and 101C08, both harboring nearly identical-sized fragments derived from the same linear bacterial chromosomal region but inserted into different *A. thaliana* chromosomes (**Supplementary Fig. 1** and data not shown). Taken together, the data demonstrate integration of >18-kb fragments of *A. tumefaciens* genomic DNA into the genome of a dicotyledonous plant.

Based on these results, as well as supporting evidence discussed below, we hypothesize that transfer of *A*chrDNA fragments into plants is mediated by T-DNA. We postulate that the single-stranded T-DNA transfer intermediate generated during the normal transformation event, called the T-strand, infrequently integrates into the chromosomes of *Agrobacterium*, and that occasionally it can be remobilized (recleaved) from such sites together with adjacent chromosomal DNA segments (**Supplementary Fig. 2** online). This model resembles the observed F′-mediated chromosomal DNA transfer occurring in and between conjugative bacteria species[10]. Activation of virulence genes as well as insertion sequences (IS) leads to nicks in the DNA and these nicks are highly recombinogenic[11,12]. We envisage that such nicks possibly generated by VirD2 at T-DNA border repeats or border-like sequences on the *A*chrDNAs, as well as other nicks and breaks resulting from IS transpositions, serve as hot spots for T-DNA integration into *A*chrDNAs. T-DNA insertions into essential genes would be lethal, but insertions into duplicated genes or into nonessential regions would be tolerated. Remobilization of *A*chrDNA would require intact border sequences or border-like sequence homologies at or near the insertion site on the *A*chrDNA. Intact border sequences may be generated in the chromosomal DNA upon insertion of more than one copy of the T-DNA directly adjacent or in close proximity to each other. Alternatively, T-DNAs with improperly processed borders might insert into *A*chrDNA and these sites may be used for retargeting the T-DNA into the plant genome. Additionally, T-DNA backbone sequences highly similar to *A*chrDNA (such as IS and their remnants) could be involved in homologous recombination of the entire T-DNA vector into *A*chrDNA. T-DNA with intact border sequences, which was deliberately inserted into bacterial chromosomes by homologous recombination, can be transferred to plants[13].

In support of this model, we found that T-DNA-*A*chrDNA junction sequences resemble typical T-DNA junctions with plant DNA (**Supplementary Fig. 3** online). These include RB nicks between the third and fourth base from the junction and micro-homology between the

sequences at or around the RB and *A*chrDNAs as well as minor deletions or substitutions. We estimate that there is a tenfold higher frequency of *A*chrDNA at the RB than at the LB site (**Table 1**). This may be because the T-DNA that inserts into *A*chrDNA lacks a functional RB sequence, as this is left behind in the Ti plasmid upon cleavage with VirD2. Therefore, once integrated into the bacterial chromosome, the T-DNA usually cannot be remobilized. However, in cases where integration occurs at chromosomal sites with RB-like sequences, these sequences might compensate for the missing RB sequence. We found evidence of an RB-like sequence on the *A*chrDNA fragment of GABI-Kat line 086C02 in the corresponding linear *Agrobacterium* chromosome (**Supplementary Fig. 4** online). Besides the RB-like sequence, we also found an overdrive[14]-like region known to enhance T-DNA cleavage and plant transformation next to the RB-like sequence (**Supplementary Fig. 4**). Although the overall similarity to the RB repeat is not very high, the presence of the overdrive and findings by others[15] that several RB-like elements originating from plant genomic sequences can be recognized and cleaved by VirD2 support our hypothesis that this RB-like region was involved in transferring this *A*chrDNA fragment into plants.

The origin of *A*chrDNA sequences found in plants is not random. Mapping the *A*chrDNA sequences found in *A. thaliana* T-DNA FSTs to the *A. tumefaciens* genome sequence showed that the same sequences from the linear chromosome appear in the FSTs at high frequency, with a similar distribution in all four T-DNA collections (**Fig. 1f** and **Supplementary Table 1** online). The sequence of the *A*chrDNA found in the GABI-Kat line 135B06 is identical to the insertion sequence IS426. This sequence was nearly identical to those found in the two independent T-DNA lines 146B12 and 343H01. These three lines contain at least two T-DNAs flanking parts of IS426. Interestingly, one of the T-DNAs was always associated with the right inverted repeat of the IS (**Supplementary Fig. 5** online). This finding is consistent with our model that chromosomal nicks and breaks generated by IS transpositions may serve as sites of T-DNA insertion into *A*chrDNA. Moreover, insertion of an additional adjacent T-DNA in the proper orientation facilitates reexcision of the complex T-DNA structure by providing LB sites at both ends, which typically lack only three nucleotides after VirD2 nicking and therefore can readily be reconstituted to functional border sequences.

The large bacterial *ABC* transporter gene family may be over-represented among the FSTs containing *A*chrDNAs. For example, the *A*chrDNA fragment obtained from the RB in the GABI-Kat line 086C02 contains a disrupted coding sequence of the *A. tumefaciens* *ABC* transporter Atu4209 and several additional coding sequences from the *A. tumefaciens* linear chromosome. Four other GABI-Kat FSTs (CU462787, CU462792, CU462794, CU462800) and 19 FLAGdb FSTs from either the RB or the LB contain parts of the same *ABC* transporter gene sequence. Therefore, certain *A. tumefaciens* genomic fragments appear to be transferred more frequently than others.

Although the 25-bp RB sequence has no perfect matches in the *A. tumefaciens* genome, 21 and 160 RB-like sequences are present when four and five mismatches are allowed, respectively. Because of the high plasticity of VirD2-mediated recognition of border-like sequences[15], these sites could serve as appropriate motifs. As there is no obvious correlation between the location of the border-like sequences and the frequency of transferred *A*chrDNAs, insertion into active IS or duplicated gene regions such as ABC transporters or multiple adjacent T-DNA insertions may determine whether or not these RB-like sequences are used.

Currently, we cannot rule out the possibility that *A*chrDNA is transferred independently of T-DNA. Such transfer could account for occasional co-integration and hence linkage to T-DNAs. Although independent *A*chrDNA integrations would not have been detected in our screen, this is highly unlikely. *A*chrDNAs found in T-DNA–tagged plants are mostly associated with the T-DNA RB border, whereas one would expect equal distribution of *A*chrDNA to both T-DNA borders in such a scenario.

Lack of sufficient sequences from transgene insertion sites and only recent availability of the entire *A. tumefaciens* genome sequence[4,5] are the likely reasons why the transfer of *A. tumefaciens* genomic DNA has not been reported before. Another complication is that *A. tumefaciens* DNA is present in transformed plant tissue owing to the presence of residual bacteria soon after transformation. As detection of *A. tumefaciens* sequences in PCR experiments is expected, these are usually filtered against *A. tumefaciens* sequences to remove such potential contaminations. This is why access to true raw data is crucial to detect *A*chrDNAs in FST collections; the lack thereof hindered our numerical analysis of rice FST data.

Our analysis of T-DNA insertion sequence databases indicates that transfer of bacterial chromosomal DNA fragments other than the T-DNA and Ti vector backbone is possible. It appears that horizontal transfer of genomic information from bacteria to plants is intimately associated with the process of crown gall formation in nature and might have played a role in the evolution of plant genomes[16]. Our results highlight not only the need for more research to understand the mechanisms of horizontal gene flow through bacteria in the evolution of higher organisms, but also the need for increased vigilance in testing for superfluous bacterial DNA in transgenic plants. As unintentional transfer of bacterial chromosomal DNA sequences into plants is not unique to *Agrobacterium*-mediated transformation but can occur even using direct DNA transformation methods that rely on plasmid DNA purified from bacteria[17], occasional transfer of *A*chrDNA to plants should not detract from the utility of *Agrobacterium* as a tool in basic and applied plant research.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank the students Pino Cagna and Alexander Boes for initial inverse PCR experiments leading to the discovery of the *Agrobacterium* chromosomal DNA in the transgenic locus named PM (PaperMaker) in a *WRKY40* promoter::*GUS* line generated by Hikaru Seki, Prisca Viehoever for technical assistance and Csaba Koncz and Steven Spiker for suggestions and critical reading of the manuscript. This work was supported in part by the Max-Planck Society, the EU-funded REGIA project and the BMBF-funded GABI-Kat project.

Published online at http://www.nature.com/naturebiotechnology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Gelvin, S. *Microbiol. Mol. Biol. Rev.* **67**, 16–37 (2003).
2. Lacroix, B. *et al. Trends Genet.* **22**, 29–37 (2006).
3. Smith, N. *Trends Plant Sci.* **3**, 85 (1998).
4. Wood, D.W. *et al. Science* **294**, 2317–2323 (2001).
5. Goodner, B. *et al. Science* **294**, 2323–2328 (2001).
6. Samson, F. *et al. Nucleic Acids Res.* **30**, 94–97 (2002).
7. Sessions, A. *et al. Plant Cell* **14**, 2985–2994 (2002).
8. Rosso, M.G. *et al. Plant Mol. Biol.* **53**, 247–259 (2003).
9. Alonso, J.M. *et al. Science* **301**, 653–657 (2003).
10. Porter, R.D. Conjugation. in *Modern Microbial Genetics* 2nd edn. (eds. Streips, U.N. & Yasbin, R.E.) 463–506 (Wiley-Liss, New York, 2002).
11. Timmerman, B. *et al. J. Mol. Biol.* **203**, 373–384 (1988).
12. Mahillon, J. & Chandler, M. *Microbiol. Mol. Biol. Rev.* **62**, 725–774 (1998).
13. Miranda, A. *et al. J. Bacteriol.* **174**, 2288–2297 (1992).
14. Van Haaren, M.J. *et al. Plant Mol. Biol.* **13**, 523–531 (1989).
15. Rommens, C.M. *et al. Plant Physiol.* **139**, 1338–1349 (2005).
16. White, F.F. *et al. Nature* **301**, 348–350 (1983).
17. Ülker, B. *et al. Transgenic Res.* **11**, 311–313 (2002).