In-depth molecular characterization of T-DNA and *Tos17* integration patterns in the Genoplante rice insertion collection

D. MIEULET, G. DROC, M. BÈS, C. ROUVIÈRE, D. MEYNARD, E. BOURGEOIS, C. SALLAUD, E. GUIDERDONI and P. PIFFANELLI
CIRAD, UMR PIA 1096, TA40/03,
34398 Montpellier Cedex 5 , France guiderdoni@cirad.fr

In the framework of the Génoplante project we have produced a collection of rice lines (Sallaud *et al.,* 2004) containing insertions of both T-DNA and *Tos17* retrotransposon (Hirochika *et al.*, 1996).

A set of 3,907 transformants harbouring the pC4978 T-DNA containing a *gusA* enhancer trap and newly transposed copies of *Tos17* was analyzed in detail to determine the potential of walk-PCR to characterize most of the inserts existing in this representative sample. This pilot study will help developing the most efficient strategy for the in-depth characterization of the entire rice mutant collection.

❑ Sequence analysis of T-DNA FST

Genomic regions flanking the left (LB) and right (RB) borders of T-DNA inserts were amplified using an efficient walk-PCR method, then sequenced. Sequence analysis using a dedicated bio-informatics pipeline revealed that the percentage of informative sequences (anchored to the 12 rice pseudomolecules) was strongly dependent upon which enzyme and T-DNA border was analysed (**Figure 2**), ranging from 31% to 52%.

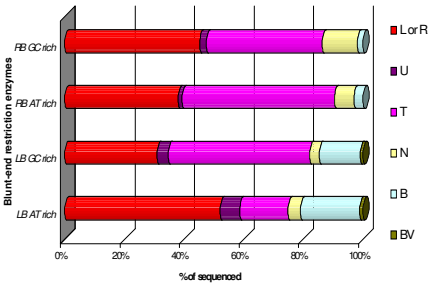

Figure 2. Graph showing the distribution of T-DNA FST-derived sequence classes as a function of the blunt-end restriction site used AT rich : DraI and SspI, GC Rich: EcoRV) and of the T-DNA border analysed (LB and RB).
L = good quality sequence > 30 bp obtained from Left Border; R = good quality sequence > 30 bp obtained from Right Border; T = tandem T-DNA sequence; N = good quality sequence < 30 bp; U = good quality sequence but no T-DNA tag identified; BV = binary vector backbone sequence; B = poor quality sequence.

In 275 lines good sequences were retrieved from PCR fragments amplified from both the left and right borders of T-DNA insert(s) at a single insertion locus. The deletion putatively generated by the insertion of the T-DNA ranged from 2 to 1,694 bp (**Table 2**) and averaged 67 bp as illustrated in **Figure 3.**

Deletion (bp)	Rice insertion lines		Average (bp)
2 à 200	258	94%	39
200 à 500	14	5%	312
500 à **1694**	3	1%	1344
Total	**275**	100%	**67**

Table 2

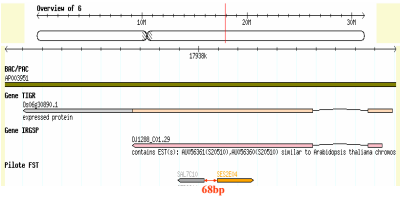

Figure 3

❑ T-DNA and *Tos17* integration patterns

Southern blot analyses conducted on a sub sample of 384 primary transformants and selected progenies revealed that:

✓an average of 2.2 copies of the T-DNA was integrated and stably transmitted to the T1 progeny.

✓the binary vector backbone was found in 46% of the analysed plants. and its presence positively correlated with the number of integrated T-DNA copies.

✓an average of 3.4 newly-transposed copies of *Tos17* were present in the genome of the regenerated plants in addition to the two native copies [on chromosome 7 and 10] and were stably and independently transmitted to the T1 generation (**Table 1**).

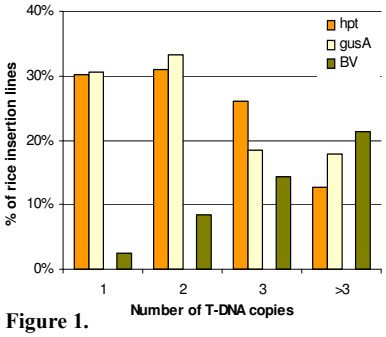

Figure 1.

	Number of spot in Southern blot					
Probes	0	1	2	3	>3	Average
hpt	14	109	112	94	46	**2,2**
gusA	16	111	121	67	65	**2,2**
BV	208	125	33	12	2	**0,6**
T-DNA	2	59	106	107	91	**2,7**
Tos17	9	37	68	84	170	**3,4**

Table 1. Summary of Southern blot hybridization of genomic DNA with hpt, gusA, binary vector (BV) and *Tos17* probes.

❑ Sequence analysis of *Tos17* FST

A highly efficient protocol of selective amplification of newly-inserted copies of *Tos17* was developed enabling the recovery of at least one *Tos17* FST in **70%** of the primary transformants .

Sequence analysis of *Tos17* flanking regions using and ad-hoc bio-analysis pipeline indicated that **92%** were informative and anchored to the rice pseudomolecules (**Table 3**).

FST sequence classes	Rice insertion lines	
TR	1761	92%
TU	58	3%
TN	36	2%
TS	22	1%
TB	43	2%

Table 3. Distribution of *Tos17* FST-derived sequence classes.

TR = good quality sequence > 30 bp obtained from newly-inserted *Tos17* copies; TN = *Tos17*-like sequence; TS = good quality sequence < 30 bp; TU = good quality sequence but no *Tos17* tag identified; TB = poor quality sequence.

❑ Preferential insertion in intergenic vs genic regions

Restriction enzyme		AT-rich				GC-rich					
		T-DNA insertions				T-DNA insertions			*Tos17* insertions		
Intragenic Regions	Average Size (TIGR; pb)	Frequencies		Corrected frequencies		Frequencies		Corrected frequencies	Frequencies		Corrected frequencies
Promoter	750	351	21%	13,2%		117	16%	12,7%	98	6%	6,5%
5'UTR	142	195	**11%**	**38,7%**		**44**	**6%**	**25,2%**	**31**	**2%**	**10,8%**
Exon	1374,87	460	**27%**	**9,4%**		**233**	**32%**	**13,8%**	**817**	**48%**	**29,5%**
Intron	918,2	405	**24%**	**12,4%**		**239**	**32%**	**21,2%**	**629**	**37%**	**34,0%**
3'UTR	318	296	**17%**	**26,2%**		**106**	**14%**	**27,1%**	**123**	**7%**	**19,2%**
TOTAL	3503,07	1707	**100%**	**100%**		**739**	**100%**	**100%**	**1698**	**100%**	**100%**
Intragenic		1707	61%			739	86%		1698	90%	
Intergenic		1090	39%			124	14%		197	10%	

Table 4. Distribution of T-DNA and *Tos17* FSTs as a function of the restriction enzyme used to generate the sequenced walk-PCR fragments; Frequencies were corrected to take into account the variable size interval of the regions delimited by TIGR annotation.

As shown in **Table 4**, the preference of insertion may be slightly biased by the restriction enzyme used to generate DNA fragments that have served to ligate adapters and specifically amplify T-DNA flanking regions. GC-rich restriction enzymes tend to reveal more frequently insertions in intragenic regions (86%) than AT-rich enzymes (61%) but less insertions in 5'UTR regions. Previously reported preference of insertion of *Tos17* for intragenic regions (90%), and exons (64%) was confirmed.