On the estimation of number of events required for saturation mutagenesis of large genomes

T-DNA and transposon-based mutagenesis are being used as powerful tools for insertional mutagenesis in plants[1–3]. In T-DNA-based mutagenesis, lines with T-DNA insertions are produced by genetic transformation. The method has been used extensively in model species *Arabidopsis thaliana*, due to possibilities of both seed[4] and *in planta*[5] transformation. However, the applicability of T-DNA insertional mutagenesis to other plant genomes is extremely limited, as development of upward of 10^4 independent transformants (with single copy integrations) by *Agrobacterium*-mediated transformation of *in vitro* regenerating explants would be extremely time consuming and difficult. With transposons, i.e. a Ds element, initial two single-copy lines would suffice for initiating transpositions by crossing Ds-containing lines with a line containing Ac transposase[6]. Populations carrying transpositions can be enriched for plants containing unlinked transpositions by using a negative selection marker[6]. Obviously, the latter method is more generally applicable and superior to T-DNA insertion method of saturation mutagenesis, as only a few transgenics will have to be developed.

Calculations have been made[3,7] for the number of plants that would have to be screened for saturation mutagenesis. These calculations are based on Poisson distribution, expressed as $f(i) = e^{-m}m^i/i!$, where $f(i)$ = frequency of targets with i insertions, $m = r/n$ (r = number of insertions and n = total number of targets). Based on this algorithm it was postulated that for model plant species *A. thaliana* (n = 5, 125 Mb haploid genome size), if around 20,000 target genes are present, for 95% of the targets to have at least one insertion (either of T-DNA or a transposon), around 60,000 lines with independent random insertions will have to be generated (Table 1). The same formulae have been used to make estimates for the maize genome[3]. However, these seem to be underestimations.

Saturation mutagenesis is a classical 'occupancy problem' wherein one seeks the probability of finding exactly 'i' number of cells empty, which in the case of saturation mutagenesis is zero as all targets/genes are to be hit at least once. Thus, the appropriate algorithm for calculating the number of hits required for saturation mutagenesis would be a Poisson approximation suggested by Feller[8] which is represented as $f(i) = e^{-m}m^i/i!$, where $m = ne^{-r/n}$ and $f(i)$ represents frequency of i targets remaining without any hit.

We assessed and compared the values for small as well as large number of targets with the commonly used Poisson distribution formulae and the algorithm suggested by Feller[8], both at 0.95 probability. Significant differences were observed in the number of events required that were estimated to be needed for saturation mutagenesis by the two theorems (Table 1). As the number of targets increased, the differences in the estimated number of events also increased. As an example, for 1000 targets the number of events required by Poisson distribution algorithm was only 3000, whereas by Feller's algorithm it was 9878 (Table 1). The numbers predicted by Feller's theorem are also supported by numbers predicted by a computer program based on multiplicative congruential random number generator, wherein, given a set of targets, the algorithm calculated the number of hits required to hit all the targets at a probability of 0.95. The numbers estimated by the random hit generator were comparable to the numbers determined by Feller's formulae (Table 1). Therefore, for saturation mutagenesis of *Arabidopsis*, considering 20,000 gene targets, 2,57,474 insertion lines would be required (at 0.95 probability level) and not 60,000 as calculated by Poisson distribution formulae[3,7] (Table 1). Reaching such a large number of hits (independent lines) would be difficult with the best of the tagging system that has been developed[6]. The success of hitting

Table 1. Number of events required, at a probability of 0.95, to hit all target genes with increasing number of genes. Calculations are based on the assumption that insertions are random

Number of events (plants) required to be screened at 0.95 probability*			
Number of targets (x)	Using $f(i)^{\#} = e^{-m}m^i/i!$, where $m = r/n$	Using random hit generator♦	Using $f(i)^{\clubsuit} = e^{-m}m^i/i!$, where $m = ne^{-r/n}$
5	15	21	23
10	30	50	53
15	45	84	86
25	75	155	155
50	150	339	345
75	225	543	547
100	300	751	758
125	375	970	975
250	750	2131	2123
500	1500	L♠	4593
1000	3000	L	9878
2000	6000	L	21,143
5000	15,000	L	57,437
10000	30,000	L	1,21,806
20000	60,000	L	2,57,474
40000	1,20,000	L	5,46,674

*0.95 probability in case of Poisson's theorem suggests that 95% of the targets are hit. In case of Feller's theorem, it means that there is 95% chance that all the targets are hit.

#$f(i)$ in Poisson theorem represents the probability of hitting a given target i times amongst a variable no. of (x) targets.

♦Average of 1000 runs. These numbers were generated using multiplicative congruential random number generator. Similar numbers were also observed using additive generator.

♣$f(i)$ in Feller's theorem represents the probability of not hitting exactly i number of targets.

♠L: Limitation as the random hit generator programme could calculate only till 250 targets.

unique sequences will be very high in the beginning of the program, but the process of generating unique hits will become highly arduous and wasteful after tagging of around 5000 unique sequences.

The problems related to generation of a large number of mutant lines can be resolved by the use of a four-element transposon system[9] which we had proposed earlier. This strategy would be applicable to both saturation mutagenesis and allele-specific tagging. Constructs for this system would carry two transposable systems (say dSpm and Ds) in conjunction. In the first step, one of the transposable elements will be used for unlinked transpositions, thereby scattering the second element throughout the genome. For the *Arabidopsis* genome, development of around 6000 lines would provide the second transposable element within 200 kb of every gene sequence. As transposable elements like Ds and dSpm show very high frequency (around 10^{-3}) of transpositions to linked sites[1,10], region-wise saturation mutagenesis would be possible by activating the second transposable element.

1. Sundaresan, V., *Trends Plant Sci.*, 1996, **1**, 184–190.
2. Krysan, P. J., Young, J. C. and Sussman, M. R., *Plant Cell*, 1999, **11**, 2283–2290.
3. Walbot, V., *Curr. Opin. Plant Biol.*, 2000, **3**, 103–107.
4. Feldmann, K. A. and Marks, M. D., *Mol. Gen. Genet.*, 1987, **208**, 1–9.
5. Bechtold, N., Ellis, J. and Pelletier, G., *C. R. Acad. Sci. Life Sci. Paris*, 1993, **316**, 1194.
6. Sundaresan, V. *et al.*, *Genes Dev.*, 1995, **9**, 1797–1810.
7. Bouchez, D. and Hofte, H., *Plant Physiol.*, 1998, **118**, 725–732.
8. Feller, W., *An Introduction to Probability and its Application*, John Wiley and Sons, New York, 1968, 3rd edn, p. 105.
9. Phogat, S., Burma, P. K. and Pental, D., *J. Biosci.*, 2000, **25**, 57–63.
10. Machida, C., Onouchi, H., Koizumi, J., Hamada, S., Semiarti, E., Torikai, S. and Machida, Y., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 8675–8680.

Received 17 June 2000; revised accepted 18 January 2001

SANJAY K. PHOGAT*
RAMNEEK GUPTA†
PRADEEP KUMAR BURMA*
KANWAR SEN‡
DEEPAK PENTAL*,#

*Department of Genetics,
University of Delhi South Campus,
Benito Juarez Road,
New Delhi 110 021, India
†Centre of Biological Sequence Analysis,
Bldg. 208, Department of Biotechnology,
The Technical University of Denmark,
Lyngby, DK-2000, Denmark
‡Department of Statistics,
University of Delhi,
Delhi 110 007, India
#For correspondence
e-mail: dpental@hotmail.com

Pollen germination on ovary wall and consequences thereof

Stigma is the seat where compatible pollen grains germinate and produce pollen tubes in all angiosperms. Only in a few species do pollen grains occasionally germinate *in situ*, e.g. *Trifolium dubium*[1], on extra floral parts and on the ovary wall, e.g. *Arabidopsis thaliana*[2,3]. In none of these exceptional cases is it known whether pollen tubes formed at sites other than stigma cause fertilization.

A unique case of pollen germination on the ovary wall was observed in *Thlaspi arvense* Linn., a crucifer. In this plant pollen germinates on ovary wall as well as on the stigma. At first, it seemed a stray event, but subsequent light and fluorescence microscopic observations revealed it to be common to all plants of the population.

T. arvense is a small winter annual, bearing white chasmogamous flowers in racemes. The flowers have cruciform corolla, 6 tetradynamous stamens and a bicarpellary syncarpous pistil. The anthers of 4 long stamens are positioned close to the stigma (Figure 1 *a*). Those of the 2 short stamens keep adpressed to the ovary (Figure 1 *b*). Flower opening begins around 6:30 h and continues throughout the day. Anthers dehisce one hour after anthesis when the stigma has already turned receptive. The stigma receives heavy pollen load within a couple of hours of anthesis. Sufficient quantity of pollen settles on the ovary wall as well. Pollens of the long stamens alight mostly on stigma and those of short stamens settle on the ovary wall (Figure 1 *a* and *b*). The plants practice autogamy as well as allogamy. The latter is effected by insects.

Pistils stained in Lewis's stain, differentiated in lactophenol and studied under light microscope revealed 76% pollen germination on the ovary wall. Germination of pollen grains on the ovary wall was also studied under fluorescence microscope following the technique described by Shivanna and Rangaswamy[4]. Fluorescence microscope helped to trace the entry and growth of pollen tubes through the stigma and ovary wall (Figure 1 *c* and *d*). Although the actual event of fusion of male and female nuclei could not be observed, presence of pollen tubes within the ovule and the development of seed are regarded as evidences of fertilization.

In order to explore functional significance of pollen germination on the ovary wall, three sets of experiments were conducted on plants raised in experimental beds in the University Botanic Garden at Jammu. In one set of plants, flowers were emasculated and bagged prior to anthesis and anther dehiscence to determine if seed develops without pollination and fertilization. When the bags were removed after 10 days, all flowers had withered and none set any fruit or seed, which indicates that the plants do not practice apomixis. The second set of plants was used as control to estimate fruit and seed set in nature, which are 58% and 55% respectively. In the third set of plants, pistils of still closed flowers were decapitated to severe the stigma. To ensure pollination, pollen was dusted manually.