

LETTER

Insertion DNA Promotes Ectopic Recombination during Meiosis in *Arabidopsis*

Xiaoqin Sun,^{*1} Yuanli Zhang,^{*1} Sihai Yang,^{*} Jian-Qun Chen,^{*} Barbara Hohn,[†] and Dacheng Tian^{*}

^{*}State Key Laboratory of Pharmaceutical Biotechnology, Plant Molecular Institute, Nanjing University, Nanjing, China; and [†]Friedrich Miescher Institute, Basel, Switzerland

Nucleotide insertion/deletions are common polymorphisms in living organisms; however, little is known about their genetic behavior during meiosis. Here, the recombination frequency (RF) of isogenic strains of transgenic *Arabidopsis thaliana*, that differ in the presence or absence of an insertion, was compared. We screened over 6 million seedlings and found that during meiosis the unpaired DNA insertions paired with ectopic homologues demonstrated a 13.8 times higher RF than that of noninsertion DNA. The direct measurement of recombination events provided the first evidence that a large piece of insertion DNA had a unique genetic behavior during meiosis. This pattern was consistently observed in different lines varying in overlapping sequence, construct orientation, chromosome location, and crossing direction. We suggest that higher ectopic recombination is promoted by DNA insertions and that this mechanism exists commonly in plants. Therefore, insertion DNA plays a nontrivial role in shaping genetic variation, chromosome instability, and genome evolution.

Introduction

It is often expected that a DNA sequence present in one individual of a species will be found in other members of that species (Fu and Dooner 2002). This “symmetric” genome structure assumption, which is generally confirmed to be true, supports this expectation. For example, human haplotypes possess highly identical sequences (Innan et al. 2003). However, insertion/deletion (indel) events result in polymorphisms and are commonly observed when 2 genomic sequences are compared (Wicks et al. 2001; Britten et al. 2003). DNA insertions can be a common component of genome sequences. For example, more than 20% of rice genomes consist of DNA insertions (Ding et al. 2007). In eukaryotic genomes, a small indel (<300 bp) in a heterozygous individual is found to increase mutation rate in the regions close to the insertion (Tian et al. 2008), whereas a large indel may reduce the crossover frequency in regions nearby (McKim et al. 1993; Hammarlund et al. 2005). Additionally, in regions where indels are abundant, highly divergent sequences leading to genetic isolation have been observed (Yang et al. 2007). The literature suggests that asymmetric DNAs such as insertion sequences affect heritable characters of adjacent nonindel sequences.

Heritable effects of insertions on adjacent symmetric DNA have been demonstrated. However, little is known about these genetic effects of insertion sequences on their own, during meiosis. For example, because an insertion has no counterpart at its allelic locus during meiosis, a large insertion will result in “unpaired DNA loops” (the insertion DNA in fig. 1*b*), which can be visualized cytologically (Poorman et al. 1981). The unpaired DNA could potentially pair with homologous sequences at nonallelic loci on any chromosome, either from a male or female parent. In other words, the DNA unpaired at allelic loci is completely free to

pair during meiosis at nonallelic loci. Therefore, a higher potential exchange with an ectopic (nonallelic) homologue is expected in the unpaired DNA than the paired symmetric sequence (fig. 1*a*).

The higher recombination potential in unpaired DNA can be directly detected by comparing the frequency of meiotic recombination between unpaired and paired DNA. The common recombination substrates in transgenic *Arabidopsis thaliana* lines (Puchta et al. 1995; Gherbi et al. 2001) are used for such detection. These substrates contain a disrupted β -glucuronidase (GUS) gene in a special arrangement, either in direct or in indirect orientation (GUUS and U'G'US in fig. 1*c*), designed specifically to reveal recombination events. These constructs consist of 2 fragments of GUS gene, one in line A651 possessing a 566-bp overlap; another in line A11 with a 1,213-bp overlap; and a 618-bp overlap in lines A1445 and A1406. A functional GUS gene can be restored when recombination occurs between the overlapping homologous sequences of the marker gene. All these lines have undergone several generations of self-pollination from the original transgenic plants and have been confirmed to be homozygous and represented by a single copy in the genome (or 2 copies in a diploid plant) (Molinier et al. 2004). Therefore, during meiosis, alleles can pair with allelic counterparts. Each line is crossed with its wild accession (or with another line; see table 1) to generate single or double insertion heterozygous plants and examine the unpaired DNA recombination frequency (RF).

The number of fully stained seedlings in a histochemical assay (fig. 2*a*) is considered evidence of meiotic recombination (Ries et al. 2000). The germinal recombinants are relatively rare in plants but can be detected. To detect such rare recombinants, we screened more than 6 million seedlings from the seeds of homozygous lines or F₁ plants (table 1). The germinal RF in unpaired recombination substrates was significantly higher ($P < 0.001$ in *t*-test) than that in the paired counterparts by a factor of 13.8 in total (table 1). Although the relative rates of RF in the unpaired compared with the paired sequences varied, rates were always greater than one in all the pairwise comparisons (ranging from 5.32 to 23.6). Each of these 4 substrates and their heterozygote progenies represented differences in the length of

¹ These authors contributed equally to this work.

Key words: insertion/deletion, ectopic recombination, genetic variation, *Arabidopsis*.

E-mail: dtian@nju.edu.cn.

Mol. Biol. Evol. 25(10):2079–2083. 2008

doi:10.1093/molbev/msn158

Advance Access publication July 18, 2008

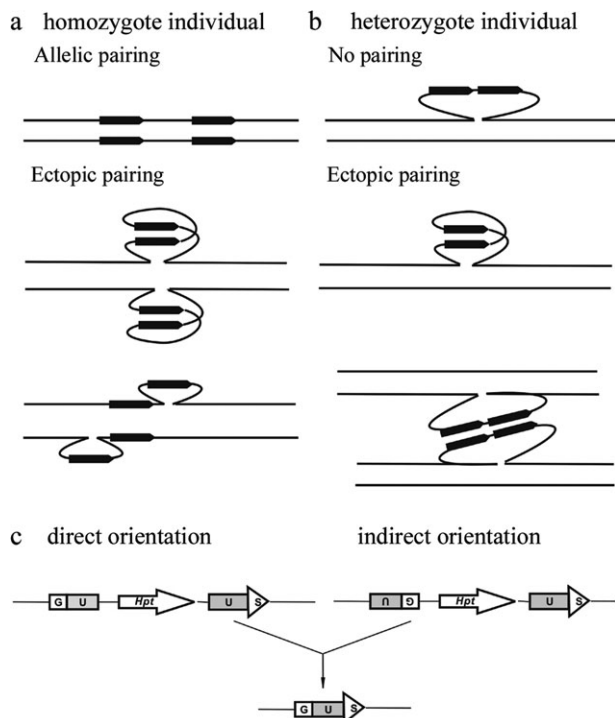


FIG. 1.—Schematic representation of a pairing model during meiosis in noninsertion DNA (a, symmetric sequences), insertion (b, asymmetric), and recombination substrates (c) integrated in transgenic lines. The 2 parallel lines represent 2 homologous chromosomes, and the gray boxes depict the homologous sequences. The ectopic pairing, defined as nonallelic pairing, is shown either in the same (center left and right) or in the different chromosomes (bottom left and right between homologous and nonhomologous chromosomes, respectively). In (c), the T-DNA of “G,” “U,” “S,” and Hpt represent part of GUS (uidA gene) and a marker gene (hygromycin phosphotransferase gene), and the arrows denote orientation of the genes. The recombination substrates carry the partially overlapping uidA region. The intact GUS gene can be restored via all types of ectopic pairing displayed above (Gherbi et al. 2001).

overlapping sequences, sequence orientation, sequence location on chromosomes, and crossing directions. These results were consistent in demonstrating a higher rate of RF in unpaired sequences. For instance, the shortest overlapping substrate 651 had a recombination rate of 13.5 when generated by crossing $651 \times \text{Col}$ or 23.6 by crossing $\text{Col} \times 651$. In addition, the relative rate value was influenced by some of the factors discussed above. Statistical analyses, employing an analysis of variance, revealed significantly ($P < 0.05$) different rates influenced by the length of overlapping sequences and the orientation of recombination substrates. However, our results were unable to detect significant differences on the relative rates affected by transgene position on the chromosome, direction of the cross, and copy numbers of the transgenes.

To confirm that these fully stained seedlings are products of meiotic recombination, at 4-leaf stage we assayed one leaf from each of 49,200 plants in the progenies of $\text{Col} \times 11$ and continued to maintain 9 plants with fully stained leaves (table 1 and fig. 2c). These assays displayed similar RF ($166.7 \sim 208.3 \times 10^{-6}$, similar to 238.3×10^{-6}). The plants maintained were checked via polymerase chain reaction (PCR) and subsequent sequencing to verify

that they were true recombinants, which contained only 1 copy of 2 fragments of GUS gene. The results were also supported by the fact that their selfed progenies segregate at roughly a 3:1 ratio, expected from the plants with fully stained leaves. Additionally, the stained patterns on leaves and the relative rates were clearly different between meiotic and mitotic recombination. Usually, only small spots on leaves were stained in the events of somatic recombination (fig. 2d), and low rates (0.26 on average in table 2) of somatic recombination between the asymmetric and the symmetric DNA were observed. These results indicated that the full-stained plants are less likely to be products of somatic recombination. Unlike meiotic division, somatic cell division lacks chromosome pairing (Griffiths et al. 2000). Therefore, the behavior of asymmetric DNA does not differ from that of symmetric DNA during mitotic recombination. Consequently, the greater the total copy number in a diploid cell, the higher the somatic RF (Swoboda et al. 1993; Molinier et al. 2004). Obviously, the sum of copy numbers in homozygotes is 2 times higher than that in heterozygotes (table 1). Our results indeed demonstrated a higher frequency of somatic recombination in homozygotes (1.33 stained spots per plant) than (0.21 spots) in heterozygous plants, which is consistent with reported results (e.g., by Swoboda et al. [1993] and Molinier et al. [2004]).

The RF in isogenic strains with or without the presence of an insertion provides the first direct evidence that a large piece of asymmetric DNA itself has unique genetic behaviors. The higher frequency of meiotic recombination in unpaired DNA indicates that an asymmetric architecture between homologous chromosomes could promote ectopic sequence exchange and copy-number variation, the primary mechanisms that shape genetic variation (Smith 1998; Arber 2000). Homogenization among homologues at different loci has been observed, and the increased flow of genetic information may in fact be the cause (Hipeau-Jacquotte et al. 1989; Vetsigian and Goldenfeld 2005). An asymmetric architecture among certain gene families results in an increased exchange of genetic information among homologues at nonallelic locations, compared with those families possessing a symmetric arrangement. Insertion DNAs that are comprised of one or more gene members may act as “bridges” to homogenize nucleotide diversity among nonallelic members. On the other hand, exchange events of genetic information from nonallelic locations will introduce novel nucleotide variations into new loci. An asymmetric arrangement can act as a “genetic introducer” from one locus to another that initiates a higher level of genetic diversity for alleles of a new locus. Therefore, asymmetric architectures can influence the evolution of gene families, having the capacity to generate both diversity and homogeneity. In addition, a nonallelic recombination will usually lead to a structural variation, for example, a copy-number variation in a gene cluster (or family), and a chromosomal rearrangement (Bailey and Eichler 2006). Thus, the asymmetric architecture functions as a genomic “reconstructor.” Via selfing such as in *A. thaliana*, any beneficial arrangement reconstructed in a heterozygote can be stably maintained in a homozygote.

The literature reports these expected consequences. By investigating the position of genes on chromosomes for 2

Table 1
Meiotic RF in Homozygous and Heterozygous Plants

Line	Features of Lines (copy no./orientation/ overlap[bp]/location)	No. of Full-Stained Seedlings	Total Screened Seedlings	RF ($\times 10^{-6}$) ^a	Relative Rate
Homozygote					
1406	2/DR/618/chr3_15367918	7	844,425	8.29	
1445	2/IR/618/chr5_8633790	7	633,825	11.04	
651	2/IR/566/chr5_10455625	6	1,060,700	5.66	
11	2/DR/1213/chr2_11765197	18	678,600	26.53	
Subtotal		38	3,217,550	11.81	
Heterozygote^b					
1406 × Col	1/DR/618/chr3_15367918	26	589,875	44.08	5.32
1445 × Col	1/IR/618/chr5_8633790	24	194,100	123.65	11.20
1445 × 1406	2/IR&DR/618/chr3&5	45	323,975	138.90	14.4
Col × 651	1/IR/566/chr5_10455625	15	112,575	133.24	23.56
651 × Col	1/IR/566/chr5_10455625	25	328,650	76.07	13.45
Col × 11	1/DR/1213/chr2_1176519	365	1,531,775	238.29	8.98
651 × 11	2/IR&DR/566&1213/chr5&2	19	109,225	173.95	10.81
Subtotal		519	3,190,175	162.7	13.78
Col × 11 ^c	1/DR/1213/chr2_1176519	5	30,000	166.67	
Col × 11 ^c	1/DR/1213/chr2_1176519	4	19,200	208.33	

^a Relative rate is calculated as the heterozygote RF divided by the corresponding homozygote RF. The seedlings used for GUS assay were F₂ seeds of heterozygous plants.

^b The first line in a cross is female and the second represents the male plant, which provides pollen.

^c Two independent tests of plants at 4-leaf stage.

rice genomes, the proportion of “asymmetric genes” in a gene family is positively correlated with the level of allelic diversity (Yang et al. 2007). Also, a significant difference in the proportion of asymmetric genes among gene families

has been observed (Ding et al. 2007), suggesting that natural selection is responsible for maintaining such asymmetric architecture. As further support of our results, a direct measurement of the recombination events at the Cf locus of tomato resistance gene, a higher frequency of recombination was clearly revealed in heterozygotes (Parniske et al. 1997; Takken et al. 1998; Wulff et al. 2004). Near isogenic lines of Cf0, Cf4, and Cf9 contain 1, 5, and 5 gene members as a gene cluster, respectively (Parniske et al. 1997). Frequent recombination events are identified among these members. Meanwhile, these genes and their intergenic sequences are highly divergent from their allelic counterparts. As a result, they are unable to pair well in heterozygotes during meiosis. In contrast, some regions in Cf4 are nearly identical (e.g., >98% nucleotide identity) with nonallelic counterparts in Cf9. As illustrated in figure 1, the unpaired (or less well-paired) DNA is free (or freer) to find a highly identical counterpart with which to pair. Therefore, more recombination events are expected in freer DNAs, particularly when the recombining partners are nearby on the same chromosome. Compared with heterozygotes, homolog slippage between sister chromatids (unequal crossing over showed in the

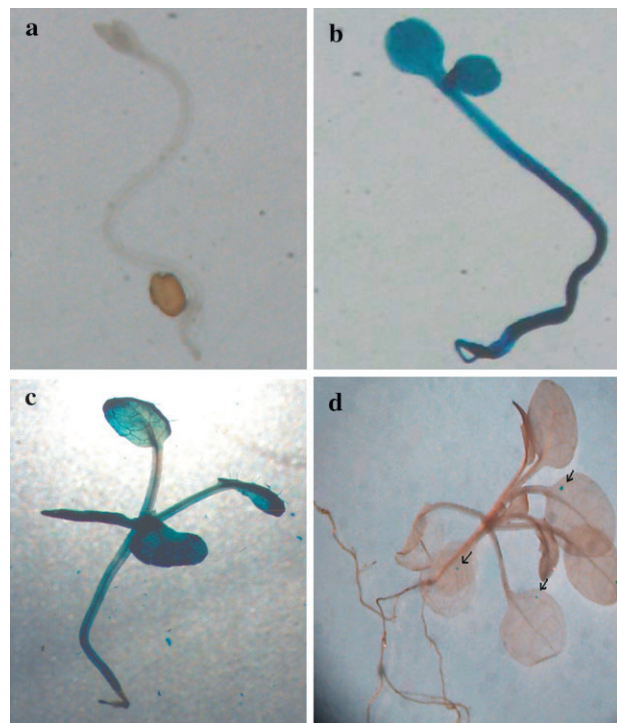


FIG. 2.—Visualization of recombination events in whole plants using histochemical GUS assay. (a) Seven-day-old seedlings of transgenic *Arabidopsis* (A11) exhibiting no recombination event, (b) Fully stained 7-day-old seedling (F₂) exhibiting a GUS⁺ phenotype, (c) Full-stained 14-day-old plants (F₃), and (d) Stained spots (arrows) exhibiting independent somatic recombination events in a 23-day-old F₂ plants.

Table 2
Somatic RF in Homozygous and Heterozygous Plants

Line	Heterozygote RF	Line	Homozygote RF	RF _{hetero} / RF _{homo}
Col × 651	0.11 ± 0.01	651	0.5 ± 0.02	0.22 ± 0.03
Col × 11	0.62 ± 0.1	11	3.27 ± 0.36	0.19 ± 0.04
1406 × Col	0.06 ± 0.01	1406	0.21 ± 0.03	0.29 ± 0.06
1445 × Col	0.06 ± 0.02	1445	0.17 ± 0.04	0.35 ± 0.01

NOTE.—The heterozygous and homozygous plants were assayed for somatic RF in independent experiments with 3 replicates by using >30 plants per replicate. RF_{hetero}/RF_{homo} is the relative rate between heterozygous and homozygous RF.

bottom of fig. 1a) is suppressed in homozygotes due to the sequence differences between intergenic regions that only allow alignments between alleles (Parniske et al. 1997). Indeed, 5 events out of 7,500 screened plants, 9 out of 17,601, and 3 out of 20,000 were found in the heterozygotes Cf4/Cf9 (Parniske et al. 1997), Cf4/Cf9 (Wulff et al. 2004), and Cf0/Cf4 (Takken et al. 1998), respectively. In all these events, the resulting gametes have not only altered copy numbers in a cluster but also reconstructed combinations of gene members within each resulting haplotype (Parniske et al. 1997). On the contrary, no events were found from 3,847 and 12,000 assayed homozygote Cf4/Cf4 (Wulff et al. 2004) and Cf9/Cf9 (Parniske et al. 1997) plants, respectively.

The consistent results demonstrated among different species suggest that the higher rate of ectopic exchange and copy-number variation promoted by asymmetric DNA is a common mechanism. Viewing asymmetric DNA as a bridge for nonallelic homologues, an introducer for alleles, or a reconstructor for chromosomes has interesting consequences. An insertion itself is a common polymorphism. Variations in the proportion, length, or location of asymmetric DNA for a gene family (or a genome) can modulate the level of genetic exchange or copy-number variation required for a family (or a genome). In fact, the proportion of asymmetric genes varies largely among gene groups (Ding et al. 2007). For example, 55.7% of rice disease resistance genes are in asymmetric DNA segments, which is 11.6 times higher than that of transcription-related genes. Clearly in a heterozygote, some of these resistance genes will have a greater chance to pair with nonallelic homologues on any chromosome. Consequently, a much higher level of nucleotide diversity and copy-number variation is achieved in disease resistance genes (Yang et al. 2007). These results suggest that asymmetric DNA plays a nontrivial role in shaping genetic variation and in maintaining chromosome instability. The outcrossed individual plants may have an added contribution to variability due to asymmetric DNA, whereas the inbred individuals can stably maintain the beneficial variants owing to symmetric sequences.

Methods

The transgenic *Arabidopsis* lines A11, A651, A1445, and A1406 were constructed by Puchta et al. (1995) and Gherbi et al. (2001). Their genome locations were determined by anchored PCR. Each line was self-pollinated for several generations, and homozygosity of the insertion was genotyped each generation by 3-primer PCR (Gherbi et al. 2001). The heterozygote plants were generated by manually crossing transgenic lines with their wild line Col or between line 1445 and 1406 or 651 and 11. After crossing, siliques were harvested separately. One hundred siliques were randomly sampled to verify hybridization by 3-primer PCR designed to confirm hybrids. Our results showed that 99% of crossed seeds were truly hybrids.

After F₂ or homozygous seeds were germinated on wet filter paper for 14 days, the seedlings were histochemically stained. The germinal RF was calculated as the number of fully stained individuals divided by the total number of

seedlings (=the total seeds × germination rate). The somatic RF was calculated as the number of blue spots per plant (Molinier et al. 2004), which were grown in vitro until 6- to 8-leaf stages and stained histochemically. Three replicates (typically >30 plants/replicate) per treatment were used to compute a mean and standard error for each line.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (30470924) to D.T. and by the Novartis Research Foundation to B.H.

Literature Cited

- Arber W. 2000. Genetic variation, molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev.* 24:1–7.
- Bailey JA, Eichler EE. 2006. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet.* 7:552–564.
- Britten RJ, Rowen L, Williams J, Cameron RA. 2003. Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci USA.* 100(8):4661–4665.
- Ding J, Araki H, Wang Q, Zhang P, Yang S, Chen J, Tian D. 2007. Highly asymmetric rice genomes. *BMC Genomics.* 8:154.
- Fu H, Dooner HK. 2002. Intraspecific violation of genetic colinearity and its implications in maize. *Proc Natl Acad Sci USA.* 99:9573–9578.
- Gherbi H, Gallego ME, Jalut N, Lucht JM, Hohn B, White CI. 2001. Homologous recombination in planta is stimulated in the absence of Rad50. *EMBO Rep.* 2(4):287–291.
- Griffiths A, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. 2000. An introduction to genetic analysis. New York: W.H. Freeman. p. 67–73.
- Hammarlund M, Davis MW, Nguyen H, Dayton D, Jorgensen EM. 2005. Heterozygous insertions alter crossover distribution but allow crossover interference in *Caenorhabditis elegans*. *Genetics.* 171(3):1047–1056.
- Hipeau-Jacquotte R, Brutlag DL, Bregegere F. 1989. Conversion and reciprocal exchange between tandem repeats in *Drosophila melanogaster*. *Mol Gen Genet.* 220(1):140–146.
- Innan H, Padhukasahasram B, Nordborg M. 2003. The pattern of polymorphism on human chromosome 21. *Genome Res.* 13(6A):1158–1168.
- McKim KS, Peters K, Rose AM. 1993. Two types of sites required for meiotic chromosome pairing in *Caenorhabditis elegans*. *Genetics.* 134(3):749–768.
- Molinier J, Ries G, Bonhoeffler S, Hohn B. 2004. Interchromatid and interhomolog recombination in *Arabidopsis thaliana*. *Plant Cell.* 16(2):342–352.
- Parniske M, Hammond-Kosack KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BB, Jones JD. 1997. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. *Cell.* 12. 91(6):821–832.
- Poorman PA, Moses MJ, Russell LB, Cacheiro NLA. 1981. Synaptonemal complex analysis of mouse chromosomal rearrangements. I. Cytogenetic observations on a tandem duplication. *Chromosoma.* 81(4):507–518.
- Puchta H, Swoboda P, Hohn B. 1995. Induction of intra-chromosomal homologous recombination in whole plants. *Plant J.* 7(2):203–210.

- Ries G, Heller W, Puchta H, Sandermann H, Seidlitz HK, Hohn B. 2000. Elevated UV-B radiation reduces genome stability in plants. *Nature*. 406(6791):98–101.
- Smith JM. 1998. *Evolutionary genetics*. Oxford: Oxford University Press. p. 267–273.
- Swoboda P, Hohn B, Gal S. 1993. Somatic homologous recombination in plants: the recombination frequency is dependent on the allelic state of recombining sequences and may be influenced by genomic position effects. *Mol Gen Genet*. 237(1–2):33–40.
- Takken FL, Schipper D, Nijkamp HJ, Hille J. 1998. Identification and Ds-tagged isolation of a new gene at the Cf-4 locus of tomato involved in disease resistance to *Cladosporium fulvum* race 5. *Plant J*. 14(4):401–411.
- Tian D, Wang Q, Zhang P, et al. (11 co-authors). 2008. Single-nucleotide mutation rate increases close to indels in eukaryotes. *Nature*. doi 10.1038/nature07175.
- Vetsigian K, Goldenfeld N. 2005. Global divergence of microbial genome sequences mediated by propagating fronts. *Proc Natl Acad Sci USA*. 102:7332–7337.
- Wicks SR, Yeh RT, Gish WR, Waterston RH, Plasterk RH. 2001. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat Genet*. 28(2):160–164.
- Wulff BB, Thomas CM, Parniske M, Jones JD. 2004. Genetic variation at the tomato Cf-4/Cf-9 locus induced by EMS mutagenesis and intralocus recombination. *Genetics*. 167(1):459–470.
- Yang S, Ke J, Hitoshi A, Jing D, Yang YH, Tian D. 2007. A molecular isolation mechanism associated with high intra-specific diversity in rice. *Gene*. 394:87–95.

Hideki Innan, Associate Editor

Accepted July 10, 2008