

C. Vitte · T. Ishii · F. Lamy · D. Brar · O. Panaud

## Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.)

Received: 3 June 2004 / Accepted: 16 September 2004 / Published online: 16 October 2004  
© Springer-Verlag 2004

**Abstract** The origin of rice domestication has been the subject of debate for several decades. We have compared the transpositional history of 110 LTR retrotransposons in the genomes of two rice varieties, Nipponbare (Japonica type) and 93-11 (Indica type) whose complete sequences have recently been released. Using a genomic paleontology approach, we estimate that these two genomes diverged from one another at least 200,000 years ago, i.e., at a time which is clearly older than the date of domestication of the crop (10,000 years ago, during the late Neolithic). In addition, we complement and confirm this first in silico analysis with a survey of insertion polymorphisms in a wide range of traditional rice varieties of both Indica and Japonica types. These experimental data provide additional evidence for the proposal that Indica and Japonica rice arose from two independent domestication events in Asia.

**Keywords** LTR-retrotransposons · Rice · Domestication · Indica/Japonica · Retrotransposon-Based Insertion Polymorphism (RBIP)

### Introduction

Despite the importance of Asian rice (*Oryza sativa* L.) as the major staple food crop worldwide, its origin, i.e. the place and period of its domestication, is still debated (Khush 1997). The majority of rice cultivars can be classified as either Japonica or Indica type, based on agromorphological traits (Oka and Morishima 1997). This classification has been supported by several studies based on the use of molecular markers, such as isozymes (Glaszmann 1987), RFLPs (Wang and Tanksley 1989), ISSRs (Blair et al. 1999) and most recently SINE insertions (Cheng et al. 2003). All these studies show a clear genetic differentiation between the two varietal groups, which has led some authors to propose that Asian rice originated from two geographically distinct gene pools of the wild progenitor *Oryza rufipogon* (Kato et al. 1928; Second 1982). We refer to this as “the double domestication” hypothesis. However, the only archaeological records that indicate rice cultivation in the late Neolithic have been found in the Yangtze river basin in China (Chen 1999). This fact supports “the single domestication” hypothesis for the origin of rice. According to this latter hypothesis, cultivated rice originated 10,000 years ago in China, and the differentiation of Indica and Japonica types must have resulted from adaptive selection following domestication (Oka and Chang 1962; Chang 1976). Although both hypotheses can explain the diphyletic structure of Asian rice germplasm, from an evolutionary perspective, they essentially differ with regard to the date of the radiation between Indica and Japonica subgroups: according to the single domestication hypothesis, this radiation occurred within the last 10,000 years, whereas the double domestication hypothesis posits

**Electronic Supplementary Material** Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00438-004-1069-6>

Communicated by M.-A. Grandbastien

C. Vitte · F. Lamy  
Laboratoire Ecologie Systematique et Evolution,  
Université Paris-Sud,  
Batiment 360, 91405 Orsay Cedex, France

T. Ishii  
Laboratory of Plant Breeding, Faculty of Agriculture,  
Kobe University, Nada-ku,  
657-8501 Kobe, Japan

D. Brar  
Plant Breeding Genetics and Biochemistry Division,  
International Rice Research Institute,  
DAPO Box 7777, Metro Manila, Philippines

O. Panaud (✉)  
Laboratoire Genome et Developpement des Plantes,  
Université de Perpignan,  
68060 Perpignan Cedex, France  
E-mail: panaud@univ-perp.fr

that Indica and Japonica types arose from two distinct gene pools of *O. rufipogon* that diverged much earlier than this.

LTR retrotransposons are ubiquitous components of plant genomes (Kumar and Bennetzen 1999). These transposable elements, because of their copy-and-paste mode of replication via mRNA intermediates, increase in copy number during the retrotransposition phase. At the time of insertion, the two LTRs of a given element are identical in sequence. Over time, however, inserted elements accumulate mutations that lead to the divergence of their two LTRs. The extent of this divergence is proportional to the time elapsed since the insertion. Using an estimate of the divergence rate of these particular sequences, it is thus possible to translate the extent of divergence of the two LTRs of a given element into an insertion date (SanMiguel et al. 1998).

The International Rice Genome Sequencing Project (IRGSP; <http://rgp.dna.affrc.go.jp>) has generated an almost complete genomic sequence from the Japonica rice variety Nipponbare. In addition, a draft sequence of the genome of the Indica rice variety 93-11 has been released (Yu et al. 2002). The availability of the genomic sequences of both cultivar types provides the opportunity to test the *single domestication* and the *double domestication* hypotheses by comparing the transpositional histories of LTR retrotransposons in both Indica and Japonica genomes. We show that the date of divergence between these two gene pools is at least 0.2 million years ago (Mya), which is unambiguously older than the date of domestication of the crop (10,000 years ago). Our data therefore provide strong molecular evidence for two independent domestications of Asian rice.

---

## Materials and methods

### Identification of retrotransposon insertions in the Nipponbare genome sequence

The BlastN procedure (<http://www.ncbi.nlm.nih.gov/Blast>) was used to retrieve copies of four previously described LTR retrotransposon families: *hopi* (Panaud et al. 2002), *houba* (Panaud et al. 2002), *Retrosat1* (GenBank Accession No. AF111709) and *RIRE8* (Kumekawa et al. 1999) from the genome of the Japonica variety Nipponbare. The following Genbank accessions were used as queries: AF537364 for *hopi*, AF537365 for *houba*, AF111709 for *Retrosat1* (note that AF111709 contains the sequences of both *Retrosat1* and *Retrosat2* retrotransposons; only the sequence of *Retrosat1* was used as a query), AB014740 (internal region) and AB014742 (LTR) for *RIRE8*. At the time of this analysis, only 50% of the Nipponbare genome sequence was available. Details of the insertions detected, and their locations are given in Supplementary Table 1.

Test for the presence of insertions at orthologous positions in the Indica genome

For each insertion identified in the Japonica genome, 200 bp of the flanking sequence was used as the query for a BlastN search of the 93-11 genomic sequence (<http://btn.genomics.org.cn/rice/>). If this Blast search did not yield any hits in the Indica sequence, or if it yielded more than two hits – indicating that the query sequence corresponded to a repeated sequence – the insertion was considered to be non-informative and was not kept for further analyses. If the Blast search allowed identification of a clear ortholog of the 200-bp flanking sequence, a gross alignment of the Nipponbare BAC sequence containing the insertion with the sequence of the corresponding Indica contig was performed using the BlastN2 procedure (<http://www.ncbi.nlm.nih.gov/Blast>). This second step allowed us to check for the presence of the insertion in the Indica genome. If the element was absent from the Indica genome (i.e., polymorphic between Indica and Japonica), then we checked for the presence of the target site duplication (TSD) in both sequences, in order to eliminate the possibility that the Indica contig sequence corresponded to an assembly artifact (we identified the TSDs in both sequences for all cases).

### Estimation of the insertion dates of the retrotransposons

The insertion date ( $T$ ) of each element identified in the Japonica genome was computed using the formula  $T = d/2s$ , where  $d$  is the substitution rate between the two LTRs and  $s$  is the average synonymous substitution rate of the *adh1* and *adh2* loci in the Family Poaceae. The latter has been estimated to be  $6.5 \times 10^{-9}$  substitutions per synonymous site per year (Gaut et al. 1996). Because the use of this value for dating retrotransposon insertions has been subject to much criticism and debate, we also computed the percentage of synonymous substitutions in the *ADHI* gene between Indica (cv. 93-11) and Japonica (cv. Nipponbare) genomes. The coding sequences (CDS) of both *ADHI* and *ADH2* genes for Nipponbare were first retrieved from the genomic sequence (Genbank Accession Nos. AC123521 and AC123515 for *ADHI* and *ADH2*, respectively). We then searched for orthologs of both genes in the Indica genome using the BlastN procedure. A clear ortholog was found for *ADHI* only (Genbank Accession No. AAAA01017984, nt 3454-3561). The p-distance based on synonymous substitutions between Indica and Japonica was therefore computed for the *ADHI* gene only. We used MEGA version 3.0 software (Kumar et al. 2004) following the method of Nei and Gojobori (1986). The standard error was computed using 500 bootstrap replicates.

Confirmation of the in silico analysis by PCR assays on a collection of traditional varieties of Japonica and Indica rice

We tested for the presence/absence of 13 polymorphic insertions (see Supplementary Table 2 for detailed description of these elements) in 66 accessions of traditional rice varieties of both Indica and Japonica types were obtained from the Germplasm Center at the International Rice Research Institute (IRRI, Manila, Philippines). Accession numbers can be found in Supplementary Table 3. Plants were grown at IRRI. Aliquots (100 ng) of total genomic DNA were extracted from fresh leaves (using the CTAB method) and used as templates for PCR. For each of the 13 insertions tested, two PCRs were performed. For the first reaction, we used primers that were homologous to the sequences flanking the inserted retrotransposon. Amplification could only be obtained if no insertion was present, since the large size of the elements used for the study prevents any amplification in PCRs using conventional *Taq* polymerase. In the second PCR, one primer was homologous to one flanking sequence and the second primer was designed to recognize the LTR sequence of the corresponding retrotransposon. In this second test, an amplification product could only be obtained if the insertion was present. This PCR test was first proposed by Flavell et al. (1998) and termed RBIP (for Retrotransposon-Based Insertion Polymorphisms). PCR conditions were identical to those given in Panaud et al. (1996), except that the annealing step was performed at a temperature of 60°C for all PCRs. Primer sequences are given in Supplementary Table 2.

## Results

### Dating of retrotransposon insertions in Nipponbare and 93-11 genomes

We first retrieved from the Nipponbare genome 179 retrotransposon sequences for the four families *hopi*, *houba*, *Retrosat1* and *RIRE8* (Table 1 and Supplementary Table 1). For 69 copies, either no ortholog was found in the Indica sequence, or the sequence flanking the retrotransposon corresponded to a repeated sequence. Therefore, genomic paleontology was conducted based on the analysis of the remaining 110 copies. Of these, 28 were also found in orthologous regions of the

Indica genome, whereas the remaining 82 were absent (Table 1). In addition, the date of insertion in the Nipponbare genome was estimated for the 110 elements by determining the degree of sequence divergence between the LTRs of each element (Sanmiguel et al. 1998). This allowed us to tentatively date the radiation between the Indica and Japonica gene pools. Figure 1 shows the results of this survey.

While the majority of recently inserted Nipponbare elements are absent from the 93-11 genome, the most ancient elements present in the Nipponbare genome are also found in the 93-11 genome. This finding shows that the two genomes have undergone distinct retrotranspositional events during a certain period, and thus supports the diphyletic origin of Asian rice. Moreover, our data allow us to date the radiation between the two gene pools at 0.9–2.1 Mya, which is unambiguously earlier than the domestication of rice (10,000 years ago).

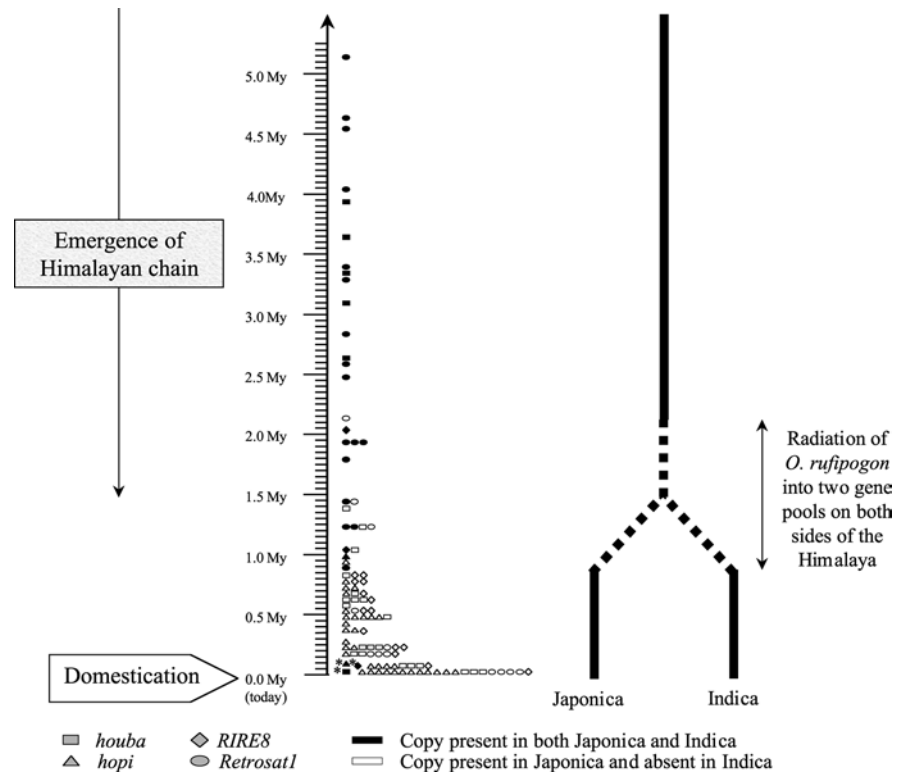
### RBIP analysis of insertions in traditional rice varieties of Indica and Japonica types

Since the previous analysis is based on the comparison of the genomic sequences of only two rice cultivars, we applied the RBIP assay (Flavell et al. 1998) to test for the presence/absence of some of the insertions in a large sample of traditional rice varieties of both Indica and Japonica types (Fig. 2). Sixty-six landraces originating from 11 Asian countries were analyzed using this assay for 13 insertion events (Table 2). These 13 elements exhibit a degree of base substitution between their LTRs ranging from 0% to 0.09% (100% indicates complete sequence identity between the two LTRs) which can be translated into a date of insertion ranging from 0 Mya to 0.7 Mya (see Materials and methods). In addition, the Nipponbare variety was used as a positive control. Of the 13 insertions tested, 11 (i.e., insertions T, G, B, L, M, V, N, D, X, O and F; see Supplementary Table 2 for a complete description of these insertions) show a dichotomy between the Indica and Japonica gene pools. This shows that the Nipponbare and 93-11 lines are typical for the Japonica and Indica types, respectively, suggesting that the date of divergence between the genomes of the two varieties 93-11 and Nipponbare inferred from the in silico analysis indeed corresponds to the date of divergence between Indica and Japonica gene pools, and therefore supports the double domestication hypothesis. For insertions H and I, however, the

**Table 1** Data mining of the retrotransposons in the Japonica rice Nipponbare and comparative analysis of the Indica rice 93-11 genome

Retroelement family	Not informative		Informative		Total
	No ortholog found in Indica	Repeated sequence	Insertion present in Japonica only	Insertion present in Indica and Japonica	
<i>hopi</i>	3	32	35	2	37
<i>houba</i>	1	5	21	6	27
<i>Retrosat1</i>	5	8	12	17	29
<i>RIRE8</i>	0	15	14	3	17
Total	9	60	82	28	110

**Fig. 1** Timing of insertions of the four families of LTR retrotransposons and comparative study of Indica and Japonica genomes. The *line on the right* shows the evolutionary history of Indica and Japonica rice as inferred from our study. The *dotted portion* of the line represents the time period during which divergence between the two gene pools may have occurred. We positioned the node of the radiation arbitrarily in the middle of this period. The *asterisks* indicate the insertions that may result from recent introgressions of Japonica into the 93-11 genome



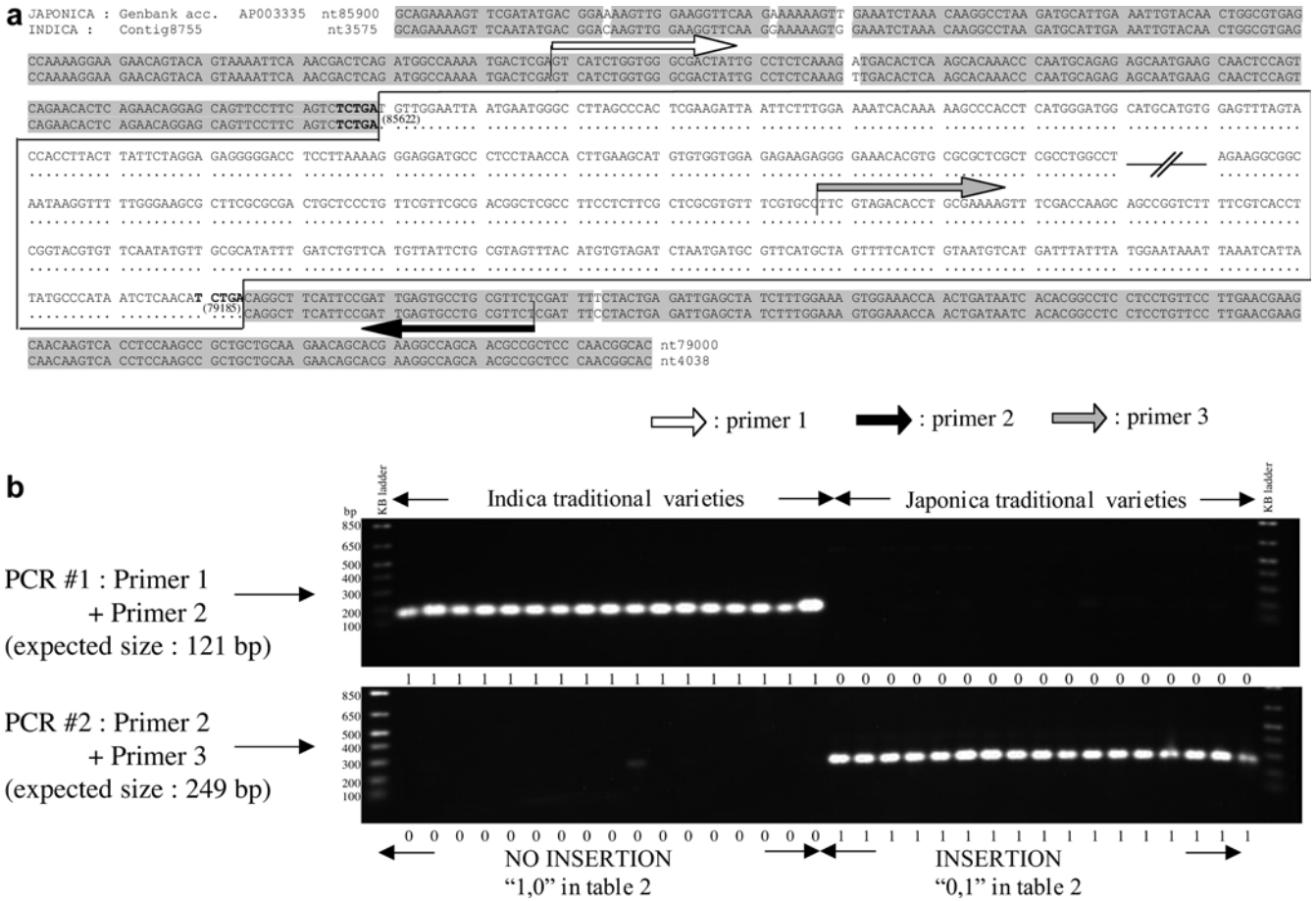
majority of the rice landraces are of Indica type (i.e., lacking the insertions), which may indicate that these two copies have transposed recently—after the domestication of the crop—and are thus found in only a few Japonica cultivars. Interestingly, these two insertions show a 100% sequence identity between their two LTRs, which is consistent with this hypothesis.

## Discussion

Both our *in silico* analysis and the PCR assay confirm that Japonica and Indica genomes diverged much earlier than 10,000 years ago. This provides molecular evidence that there have been at least two domestications of rice in Asia. In addition, our estimated date of divergence between Indica and Japonica is coincident with the rise of the Himalayas, which could have separated *O. rufipogon*, the wild ancestor of cultivated Asian rice, into two distinct gene pools (Audley-Charles et al. 1981). The gene pool to the North of the mountain range would then be at the origin of the Japonica type, whereas the one in the South would be at the origin of the Indica type (Second 1982). In order to confirm this hypothesis, we now need to extend our survey of retrotransposon insertions to a larger sample of *O. rufipogon* accessions covering the range of distribution of the species in Asia. At the time this study was conducted, the contigs of genomic sequences available for the 93-11 variety were of small size, compared to those available for Nipponbare. Consequently, the approach described above could not be applied reciprocally, i.e. first searching the 93-11 sequence for retrotransposon inser-

tions and then checking for their presence at orthologous positions in the Nipponbare sequence. This should, however, be possible as the sequencing of the 93-11 genome progresses and larger sequence blocks become available (<http://rise.genomics.org.cn>). Only then will we be able to tentatively identify the origins of both Indica and Japonica rice types.

Our estimate of the date of the Indica-Japonica radiation relies upon the rate of divergence between the LTRs of the retrotransposons that we retrieved from the Nipponbare genomic sequence. As this rate is not known, we used the average nucleotide substitution rate for the *adh1* and *adh2* loci of the Family Poaceae, which has been estimated at  $6.5 \times 10^{-9}$  substitutions per synonymous site per year (Gaut et al. 1996). This rate has been applied to the dating of insertions of LTR retrotransposons in the maize (SanMiguel et al. 1998) and rice (Jiang et al. 2002) genomes. One could, however, reasonably argue that LTR sequences might diverge at a higher rate than the synonymous sites of the genic regions. SanMiguel et al. (1998) have indeed suggested that the molecular clock might run up to three times faster for the LTRs of the retrotransposons than for the synonymous sites in coding sequences such as the ADH genes. Recent studies have shown that LTR retrotransposons in *Arabidopsis thaliana* (Devos et al. 2002) and *O. sativa* (Ma et al. 2004) indeed mutate at a high rate, but through small deletions rather than substitutions. A similar observation has been made for LINES (Long Interspersed Nuclear Elements) in the *Drosophila* genome (Petrov et al. 1996). However, we have only taken substitutions into account in dating the insertion events.



**Fig. 2a, b** Results of a representative RBIP assay to test for presence/absence of a particular retrotransposon in the sample of traditional varieties of rice analyzed. The example shows the results for insertion N (retrotransposon *houbu*). **a** Sequence alignment of BAC clone AP003335 (Japonica) and contig8755 (Indica) in the vicinity of the retrotransposon insertion site. The shaded portions indicate complete sequence identity between Japonica and Indica. The white, grey and black arrows indicate the positions of the primers used for the PCR assays. The retrotransposon present in the Japonica sequence is boxed. The sequences shown in bold indicate the target site that is duplicated as a result of the insertion. **b** Results of the PCR assays for insertion N from a subset of rice landraces of Indica and Japonica types (see Supplementary Table 3 for the correspondence between the genotypes and the lanes on the gel)

Nevertheless, we have also computed the p-distance between the *ADH1* genes of Indica and Japonica based on synonymous substitutions (see Materials and methods) and found a value of  $0.008 \pm 0.005$ , which can be translated to a date of  $0.62 \pm 0.38$  Mya using a substitution rate of  $6.5 \pm 10^{-9}$  substitutions/site/year. If this value represents the true date of the divergence between the two gene pools, then it would suggest that the average substitution rate for the retrotransposons is indeed twice as high as that of the *ADH* gene, i.e. in the order of  $13 \times 10^{-9}$ . In any case, we could consider that the radiation between the Indica and Japonica gene pools took place at least 200,000 years ago—still long before the domestication of the crop.

Another possible source of error in our estimate of the radiation date is gene flow that may have occurred between Indica and Japonica subgroups after their domestication. In our case, introgression of Japonica DNA into the 93-11 genome would perturb the estimation of the radiation date. We identified three discrepancies in our dataset (marked with asterisks in Fig. 1) that could be the result of such events. These three elements have inserted into the Nipponbare genome within the last 100,000 years, and are also found in orthologous regions of 93-11. Furthermore, the results of the RBIP assay show that such introgressions between both gene pools may indeed have occurred after domestication: In Table 2, 17 cases of discrepancies (9 in Japonica landraces and 8 in Indica landraces) can be observed, e.g. the presence of the O insertion in the Indica landrace Kalajira or the absence of the T insertion in the Japonica landrace Chow-sung. However, such discrepancies represent only 2.3% of the dataset (17 out of 726 data-points) obtained for the 11 insertions that occurred before the domestication of the crop, and therefore do not affect the conclusions with regard to the clear separation of the two gene pools and the date of radiation drawn from our analysis.

Among the 13 insertions tested with the RBIP assay, 4, i.e. I, H, T and G, correspond to retrotransposons that show 100% identity between their two LTRs (see

**Table 2** Summary of the results of the PCR assay on the 66 traditional landraces and the variety Nipponbare for 13 retrotransposon insertions

Insertion code <sup>a</sup>			I	H	T	G	B	L	M	V	N	D	X	O	F	
LTR divergence <sup>b</sup>			0	0	0	0	0.001	0.002	0.002	0.002	0.003	0.005	0.005	0.008	0.009	
Chromosomal location			I	I	X	I	X	IV	VII	III	I	IV	I	I	I	
Variety	Type <sup>c</sup>	Origin <sup>d</sup>	PCR data <sup>e</sup>													
Barah Baghlan	Ind	AFG	1,0	1,0	1,0	1,0	1,0	<b>0,1</b>	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Basmati Kunduz	Ind	AFG	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Dehraduni Kunduz	Ind	AFG	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Luk Herat	Ind	AFG	1,0	1,0	1,0	1,0	1,0	<b>0,1</b>	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Laldhan	Ind	BGD	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Pang Bara	Ind	BHU	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Sheng-li-hsien	Ind	CHN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Cina	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Fache Dogo	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Fawayaso	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Ga Falu	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Masia	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Nata	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Sari Lawa	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Serigi Manggala	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Si Adulo	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Siumene	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Sua	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Taria Faigi	Ind	IDN	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Kalajira	Ind	IND	1,0	1,0	1,0	<b>0,1</b>	1,0	<b>0,1</b>	<b>0,1</b>	1,0	1,0	1,0	<b>0,1</b>	<b>0,1</b>	1,0	1,0
Eravapandy	Ind	IND	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Basmati	Ind	IND	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Latisail	Ind	IND	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Sudubalawee	Ind	LKA	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Gangala	Ind	LKA	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Babawee	Ind	LKA	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Bathkirieli	Ind	LKA	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Batapola-al	Ind	LKA	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Baheria	Ind	NPL	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Bansbareli	Ind	NPL	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Basmati Nokhi	Ind	NPL	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Dewsar	Ind	NPL	1,0	1,0	1,0	1,0	1,0	<b>0,1</b>	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Dhusuni	Ind	NPL	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Dumsikalām	Ind	NPL	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Balibud	Ind	PHL	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Burik	Ind	PHL	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Sekiyama 2	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Take-nari	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Bozu	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Oba	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Shinshu	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Goriki 2	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Husaku-shirazu	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Kibi-ho	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Shiratama	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Shinriki 11	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Ben-Kei	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Asahi 1	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Kuhei 2	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Kuro mochi	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Gin bozu chusei	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Hattan 10	Jap	JPN	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Chow-sung	Jap	KOR	1,0	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Mansaeng	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Chanarak	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Cheonjudo	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Daegujo	Jap	KOR	<b>0,1</b>	<b>0,1</b>	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Damajo	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Doyajichal	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Dujo	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Hwinbe	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Jokjebichal	Jap	KOR	1,0	1,0	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	1,0	1,0	1,0

**Table 2** (Contd.)

Insertion code <sup>a</sup>	I	H	T	G	B	L	M	V	N	D	X	O	F	
LTR divergence <sup>b</sup>	0	0	0	0	0.001	0.002	0.002	0.002	0.003	0.005	0.005	0.008	0.009	
Chromosomal location	I	I	X	I	X	IV	VII	III	I	IV	I	I	I	
Variety	Type <sup>c</sup>	Origin <sup>d</sup>	PCR data <sup>e</sup>											
Maekjo	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Mongeunsare	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Mongeunchal	Jap	KOR	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
Yeongdalchal	Jap	KOR	1,0	1,0	1,0	1,0	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	1,0
Nipponbare	Jap	JPN	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>

<sup>a</sup>The correspondence between the insertion codes and the copy used in the in silico analysis can be found in the Electronic Supplementary Table 2

<sup>b</sup>The LTR divergence is the ratio of the number of substitutions between the two LTRs over the total size of the LTR

<sup>c</sup>Ind, Indica; Jap, Japonica

<sup>d</sup>The country codes for the origin of the varieties are : AFG (Afghanistan), BGD (Bangladesh), BHU (Bhutan), CHN (China), IDN (Indonesia), IND (India), LKA (Sri Lanka), NPL (Nepal), PHL (Philippines), JPN (Japan) and KOR (Korea). Accession

numbers from the Rice Germplasm Center (IRRI) are given in Supplementary Table 3

<sup>e</sup>The PCR results are tabulated as follows. 1,0 indicates that PCR #1 gave a clear amplification product, whereas the PCR #2 did not give any product, implying that the insertion is absent from the corresponding genome. 0,1 (*bold*) indicates that the PCR #1 did not give any amplification product, whereas PCR #2 gave a clear amplification signal, implying that the insertion is present in the corresponding genome (see Fig. 2 for an illustration of the data)

Table 2). I, H and T correspond to *houba* elements, whereas the G insertion represents a *Retrosat1* element (see Supplementary Table 2). Since the LTRs of *houba* and *Retrosat1* are 1000 bp and 400 bp long, respectively, this complete sequence identity translates into an insertion date ranging from 0 to 77,000 years ago in the case of *houba* and from 0 to 192,000 years ago in the case of *Retrosat1* (when a substitution rate of  $6.5 \times 10^{-9}$  substitution/site/year is used; see Materials and Methods and the discussion above). The results of the PCR assay obtained for both insertions T and G are identical to those obtained with older insertions, and therefore confirm the results of the in silico analysis. However, in the case of the insertions I and H, our PCR assays showed that they are present only in Nipponbare (our positive control) and the variety Daegujo, which is an indication that these two elements might have transposed *after* the domestication of Japonica rice, i.e. within the last 10,000 years. The insertions H and I are therefore not exploitable for the present study. Nevertheless, this shows that recently inserted retrotransposons could be a source of polymorphic markers within Japonica (and Indica) varietal groups.

Rice has now become one of the major model crop species in plant genomics. The debate over its origin, i.e. a single versus a double domestication, may be of interest not only to historians. Indeed, knowledge of the diphyletic nature of the rice gene pool and, most importantly, of the time of divergence between the Indica and Japonica types will facilitate characterization of the molecular diversity of rice germplasm. In particular, the ancient radiation of Indica and Japonica gene pools suggests new ways to approach these two genomes and the utilization of Indica and Japonica rices in genomic research and breeding. Moreover, the differentiation of Indica and Japonica ancestral genomes through the activity of transposable elements may not only be

structural, but also gives new insights for future research in functional genomics. A better understanding of the molecular bases of the phenotypic diversity that Asian rice exhibits is one of the biggest challenges awaiting rice genomics.

**Acknowledgements** The authors thank A. Frary for her helpful comments on the manuscript and S. Yanagihara for his valuable help in choosing the traditional Japonica varieties used for the analysis.

## References

- Audley-Charles MG, Hurley AM, Smith AG (1981) Continental movements in the mesozoic and cenozoic. In: Whitmore TC (ed) Wallace's line and plate tectonics. Clarendon Press, Oxford, UK, pp 9–23
- Blair MW, Panaud O, McCouch SR (1999) Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.) Theor Appl Genet 98:780–782
- Chang TT (1976) The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. Euphytica 25:435–485
- Chen B (1999) Origin of 8000-year-old cultivated rice in Henan's Jia Lake site. Agric Archaeol 1:55–57
- Cheng C, Motohashi R, Tsuchimoto S, Fukuta Y, Ohtsubo H, Ohtsubo E (2003) Polyphyletic origin of cultivated rice: based on the interspersed pattern of SINEs. Mol Biol Evol 20:67–75
- Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075–1079
- Flavell AJ, Knox MR, Pearce SR, Ellis THN (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 16:643–50
- Gaut BS, Morton BR, McCaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci USA 93:10274–10279
- Glazmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 74:21–30

- Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler SR (2002) A recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* 161:1293–1305
- Kato S, Kosaka H, Hara S (1928) On the affinity of the cultivated varieties of rice plants, *Oryza sativa* L. *Bull Sci Fac Agric Kyushu Univ, Fukuoka, Japan* 3:132
- Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* 35:25–34
- Kumar A, Bennetzen JL (1999) Plant retrotransposons. *Annu Rev Genet* 33:479–532
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
- Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E (1999) Identification and characterization of novel retrotransposons of the *gypsy* type in rice. *Mol Gen Genet* 260:593–602
- Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14:860–869
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Oka HI, Chang TT (1962) Rice varieties intermediate between wild and cultivated forms and the origin of the Japonica type. *Bot Bull Acad Sinica* 3:109–131
- Oka H, Morishima H (1997) Wild and cultivated rice. In: Matsuo T, Futsuhara Y, Kikushi F, Yamaguchi H (eds) *Science of the rice plant* (vol 3: Genetics). Nobunkyo, Tokyo, pp 88–111
- Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet* 252:597–607
- Panaud O, Vitte C, Hivert J, Muzlak S, Talag J, Brar DS, Sarr A (2002) Characterization of transposable elements in the genome of rice (*Oryza sativa* L.) using Representational Difference Analysis. *Mol Genet Genomics* 268:113–121
- Petrov D, Lozovskaya E, Hartl D (1996) High intrinsic rate of DNA loss in *Drosophila*. *Nature* 384:346–349
- SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20:43–45
- Second G (1982) Origin of the genetic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. *Jpn J Genet* 57:25–57
- Wang ZY, Tanksley SD (1989) Restriction fragment length polymorphism in *Oryza sativa* L.. *Genome* 32:1113–1118
- Yu J, et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica*). *Science* 296:79–92