New Phytologist

Research

Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes

Qihui Zhu and Song Ge

Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

Author for correspondence:
Song Ge
Tel: +86 10 62836097
Fax: +86 10 62590843
Email: gesong@ibcas.ac.cn

Received: *15 December 2004*
Accepted: *28 January 2005*

Summary

• The A-genome group in *Oryza* consists of eight diploid species and is distributed world-wide. Here we reconstructed the phylogeny among the A-genome species based on sequences of nuclear genes and MITE (miniature inverted-repeat transposable elements) insertions.

• Thirty-seven accessions representing two cultivated and six wild species from the A-genome group were sampled. Introns of four nuclear single-copy genes on different chromosomes were sequenced and analysed by both maximum parsimony (MP) and Bayesian inference methods.

• All the species except for *Oryza rufipogon* and *Oryza nivara* formed a monophyletic group and the Australian endemic *Oryza meridionalis* was the earliest divergent lineage. Two subspecies of *Oryza sativa* (ssp. *indica* and ssp. *japonica*) formed two separate monophyletic groups, suggestive of their polyphyletic origin. Based on molecular clock approach, we estimated that the divergence of the A-genome group occurred *c*. 2.0 million years ago (mya) while the two subspecies (*indica* and *japonica*) separated *c*. 0.4 mya.

• Intron sequences of nuclear genes provide sufficient resolution and are informative for phylogenetic inference at lower taxonomic levels.

Key words: A-genome, divergence time, miniature inverted-repeat transposable elements (MITEs), molecular phylogeny, nuclear intron sequence, *Oryza*.

New Phytologist (2005) **167**: 249–265

© *New Phytologist* (2005) **doi**: 10.1111/j.1469-8137.2005.01406.x

Introduction

The rice genus (*Oryza*) comprises approx. 23 species and is represented cytogenetically by 10 genome groups (i.e. the A-, B-, C-, BC-, CD-, E-, F-, G-, HJ- and HK-genomes; Ge *et al.*, 2001). The A-genome group, also called the *Oryza sativa* complex, consisting of eight diploid species (Vaughan, 1989), is distributed world-wide and the most recently diverged lineage in *Oryza* (Vaughan, 1989; Ge *et al.*, 2001). In this group, *O. sativa*, which is of the Asian origin and cultivated world-wide, has two main subspecies, *O. sativa* ssp. *indica* and *O. sativa* ssp. *japonica* (Nayar, 1973; Oka, 1988; Morishima *et al.*, 1992). The other cultivated species, *Oryza glaberrima*, was domesticated in West Africa (Vaughan, 1989). Of the remaining six wild species, the perennial *Oryza rufipogon* Griff. is distributed throughout tropical Asia and Oceania, whereas the annual *Oryza nivara* Sharma et Shastry is restricted to tropical continental Asia. The two species differ markedly in life-history traits and habitat preference (Oka, 1988; Vaughan, 1989) but intermediate types have been found in some *O. rufipogon* strains (Sano *et al.*, 1980). Of the two wild species endemic to Africa, *Oryza barthii* A. Chev. is the annual wild relative of the cultivated *Oryza glaberrima*, while the perennial and rhizomatous *Oryza longistaminata* is partly self-incompatible and allogamous (Ghesquiere, 1986). The Australian endemic *Oryza meridionalis* is a primarily inbred annual species and the South American endemic *Oryza glumaepatula* varies in perenniality (Vaughan, 1989). Because these wild species have played significant roles in rice breeding by contributing genes valuable for resistance and tolerance to

biotic and abiotic stresses (Chang, 1976; Vaughan, 1989), a better understanding of genetic diversity and relationships among them will facilitate more effective conservation and efficient utilization of the rice germplasm (Morishima *et al.*, 1992; Aggarwal *et al.*, 1999; Lu *et al.*, 2000; Ge *et al.*, 2001).

In the last decade, a wealth of data from morphological and cytological studies (Nayar, 1973; Lu *et al.*, 2000) and molecular markers such as isozyme (Second, 1982), nuclear restriction fragment length polymorphism (RFLP) (Wang *et al.*, 1992; Bautista *et al.*, 2001; Lu *et al.*, 2002), random amplified polymorphic DNA (RAPD) (Ishii *et al.*, 1996; Bautista *et al.*, 2001; Ren *et al.*, 2003), amplified fragment length polymorphism (AFLP) (Aggarwal *et al.*, 1999; Park *et al.*, 2003), simple sequence repeat (SSR) (Ishii *et al.*, 2001; Ren *et al.*, 2003), short interspersed elements (SINEs) and miniature inverted-repeat transposable elements (MITEs) insertions (Mochizuki *et al.*, 1993; Iwamoto *et al.*, 1999; Cheng *et al.*, 2002) have significantly enhanced our understanding of the evolution of the A-genome species. However, taxonomy of this group of species is particularly problematic because most of them lack clear morphological distinguishing characteristics (Vaughan, 1989; Lu *et al.*, 2000) and their phylogenetic relationships have not been fully resolved or are inconsistent between studies. One example involves the basal split in the A-genome group. Wang *et al.* (1992) and Park *et al.* (2003) indicated that *O. meridionalis* was the most distinct species in the *O. sativa* complex (A-genome group) based on nuclear RFLP and MITE-AFLP data. By RAPD and SSR analyses, however, Ren *et al.* (2003) found that the *O. longistaminata* accessions differentiated significantly from all other A-genome species in agreement with previous investigations (Iwamoto *et al.*, 1999; Cheng *et al.*, 2002). Another long-standing debate concerns whether the Asian cultivated rice is originated monophyletically or polyphyletically. Oka & Morishima (1982) considered that the subspecies *indica* was domesticated primarily from its wild relatives in South or South-east Asia, and subspecies *japonica* was a type derived from *indica* and adapted to high elevation and high latitude. Recent studies based on the interspersion pattern of SINEs demonstrated that *O. sativa* had been derived polyphyletically from *O. rufipogon* (Cheng *et al.*, 2003; Yamanaka *et al.*, 2003).

In recent decades, molecular sequence data have been increasingly used for phylogenetic studies at varying scales from major lineages of land plants to closely related species (Soltis & Soltis, 2000; Sang, 2002). Nonetheless, phylogenetic reconstruction at the lower taxonomic levels is limited by the availability of sequences with sufficient variation (Doyle *et al.*, 1996). Noncoding DNA sequences such as the chloroplast *trn*L-*trn*F, *atp*B-*rbc*L and nuclear rDNA internal transcribed spacer region (nrITS) represent a potential source of markers for resolving phylogenetic relationships of closely related species (Baldwin *et al.*, 1995; Soltis & Soltis, 1998). However, in some cases, there is no sufficient resolution for them to distinguish the groups that have undergone rapid or recent radiations (Vargas *et al.*, 1998; Kelchner, 2000; Oh & Potter, 2003). Accordingly, intron sequences of nuclear genes have been increasingly used for reconstruction of phylogeny at lower taxonomic levels. Doyle *et al.* (1996) found that sequences from the nuclear histone H3 intron provided sufficient resolution among the species in the genus *Glycine*, whose relationships were unsolved in previous studies. Using the second intron of a homoeotic gene, *LEAFY*, Oh & Potter (2003) elucidated the phylogenetic relationships within the genera *Neillia* and *Stephanandra*, and a comparison of the *LEAFY* data with ITS and cpDNA data demonstrated that the *LEAFY* intron was the most variable and useful for phylogenetic analysis at the species level. Nuclear intron sequences have also been successfully used in the genera *Sphaerocardamum* (Bailey & Doyle, 1999), *Scaevola* (Howarth & Baum, 2002), *Amorphophallus* (Grob *et al.*, 2004) and paleotropical moss (Wall, 2002).

In this study, we sequenced introns of four nuclear genes on different chromosomes for multiple accessions of each of eight A-genome species in the genus *Oryza*. Based on the intron sequences, we reconstructed the phylogenetic relationships of the A-genome species. We are particularly interested in: (1) Which species is the basal lineage of the A-genome group? (2) How did the Asian cultivated rice originate, monophyletically or polyphyletically? (3) What is the appropriate taxonomic treatment for the perennial *O. rufipogon* and annual *O. nivara*? (4) Which group is the most closely related relative to the American *O. glumaepatula*? (5) When did the A-genome species begin to diverge and radiate? We also explore the utility of nuclear introns for phylogenetic studies within the A-genome group and discuss the implication for their use at lower taxonomic level in general.

Materials and Methods

Plant materials

Thirty-seven accessions of eight A-genome species were used, including eight accessions of *O. sativa* (four *indica* and four *japonica*), two accessions of *O. glaberrima*, and 27 accessions of wild species (seven *O. rufipogon*, five *O. nivara*, three *O. glumaepatula*, three *O. meridionalis*, six *O. longistaminata*, and three *O. barthii*). We also included one accession each of the B-genome species (*O. punctata*) and the E-genome species (*O. australiensis*) as the outgroups, because previous studies showed that the B- and E-genome species were closely related to the A-genome group (Ge *et al.*, 1999a, 2001). All the accessions sequenced in this study, with their scientific names, geographic origins and GenBank accession numbers are listed in Table 1.

Intron selection and primer design

In order to select applicable introns, we searched for target genes based on the criteria that they (1) should be single copy and not members of a gene family and thus reduced paralogy concerns, (2) should include one or more sizeable introns so that conserved primers could be developed based on their

Table 1 *Oryza* plant material used in this study

Taxa	Code[a]	Accession No.[b]	Origin	GenBank Accession No.			
				Adh1-i3	*Os1283*	*Os9971*	*Os17357*
O. sativa ssp. indica	*indica*-PHL	30416	Philippines	AY749209	AY749247	AY749285	AY749323
O. sativa ssp. indica	*indica*-CHN	ZH249	China	AY749210	AY749248	AY749286	AY749324
O. sativa ssp. indica	*indica*-IDN	10594	Indonesia	AY749211	AY749249	AY749287	AY749325
O. sativa ssp. indica	*indica*-IND	74716	India	AY749212	AY749250	AY749288	AY749326
O. sativa ssp. japonica	*japonica*-FRA	69317	France	AY749213	AY749251	AY749289	AY749327
O. sativa ssp. japonica	*japonica*-GRC	Au70140	Greece	AY749214	AY749252	AY749290	AY749328
O. sativa ssp. japonica	*japonica*-JPN	Nipponbare	Japan	AC123521	AP001366	AP005691	AL731638
O. sativa ssp. japonica	*japonica*-VNM	Au8134	Vietnam	AY749215	AY749253	AY749291	AY749329
O. glaberrima	*glaberrima*-WAF	T0999/90	West Africa	AY749232	AY749270	AY749308	AY749346
O. glaberrima	*glaberrima*-BFA	103474	Burkina Faso	AY749233	AY749271	AY749309	AY749347
O. rufipogon	*rufipogon*-IND1	101965	India	AY749216	AY749254	AY749292	AY749330
O. rufipogon	*rufipogon*-IND2	80506	India	AY749219	AY749257	AY749295	AY749333
O. rufipogon	*rufipogon*-TAI	100678	Taiwan, China	AY749217	AY749255	AY749293	AY749331
O. rufipogon	*rufipogon*-THA	105942	Thailand	AY749218	AY749256	AY749294	AY749332
O. rufipogon	*rufipogon*-CHN	GXHY	Guangxi, China	AY749220	AY749258	AY749296	AY749334
O. rufipogon	*rufipogon*-LAO	106161	Laos	AY749221	AY749259	AY749297	AY749335
O. rufipogon	*rufipogon*-MMR	105494	Myanmar	AY749222	AY749260	AY749298	AY749336
O. nivara	*nivara*-LKA1	103407	Sri Lanka	AY749223	AY749261	AY749299	AY749337
O. nivara	*nivara*-LKA2	105431	Sri Lanka	AY749225	AY749263	AY749301	AY749339
O. nivara	*nivara*-THA	105391	Thailand	AY749224	AY749262	AY749300	AY749338
O. nivara	*nivara*-CHN	103824	China	AY749226	AY749264	AY749302	AY749340
O. nivara	*nivara*-IND	106061	India	AY749227	AY749265	AY749303	AY749341
O. glumaepatula	*glumaepatula*-COL	105561	Colombia	AY749229	AY749267	AY749305	AY749343
O. glumaepatula	*glumaepatula*-BRA	105672	Brazil	AY749230	AY749268	AY749306	AY749344
O. glumaepatula	*glumaepatula*-CUB	100184	Cuba	AY749231	AY749269	AY749307	AY749345
O. meridionalis	*meridionalis*-AUS1	101147	Australia	AY749242	AY749280	AY749318	AY749356
O. meridionalis	*meridionalis*-AUS2	105306	Australia	AY749243	AY749281	AY749319	AY749357
O. meridionalis	*meridionalis*-AUS3	103317	Australia	AY749244	AY749282	AY749320	AY749358
O. longistaminata	*longistaminata*-MLI	101219	Mali	AY749238	AY749276	AY749314	AY749352
O. longistaminata	*longistaminata*-KEN	104977	Kenya	AY749234	AY749272	AY749301	AY749348
O. longistaminata	*longistaminata*-NGA1	104075	Nigeria	AY749239	AY749277	AY749315	AY749353
O. longistaminata	*longistaminata*-CIV	101207	Cote d Ivoire	AY749228	AY749266	AY749304	AY749342
O. longistaminata	*longistaminata*-ETH	104638	Ethiopia	AY749240	AY749278	AY749316	AY749354
O. longistaminata	*longistaminata*-NGA2	105061	Nigeria	AY749241	AY749279	AY749317	AY749355
O. barthii	*barthii*-SDN	100933	Sudan	AY749235	AY749273	AY749311	AY749349
O. barthii	*barthii*-MLI	106205	Mali	AY749236	AY749274	AY749312	AY749350
O. barthii	*barthii*-CMR	104140	Cameroon	AY749237	AY749275	AY749313	AY749351
O. punctata	*punctata*	105984	Cameroon	AY749246	AY749284	AY749322	AY749360
O. australiensis	*australiensis*	105263	Australia	AY749245	AY749283	AY749321	AY749359

[a]Three-letter abbreviations following species names represent the country of origin.
[b]All accessions are provided by the Genetic Resources Center of the International Rice Research Institute (IRRI) at Los Banos, Philippines, except for ZH249, GXHY, Nipponbare which were collected by the authors.

flanking coding regions (exons), and (3) should be unlinked preferably on different chromosomes, and thus represent historically independent markers.

Specifically, we first used the predicted rice cDNA sequences as queries for computer-assisted sequence similarity searches in the rice (*O. sativa* ssp. *japonica*) BGF (Beijing Gene Finding) gene database using the program BLAST (Altschul *et al.*, 1990). As a result, 420 predicted single-copy genes were obtained in this step. From them 58 target genes were extracted because these genes included introns of appropriate length (around 1 kb). Second, by referring to maize and wheat cDNA sequences in GenBank, we selected seven genes located on six different chromosomes to design primers based on the exon sequences conserved across rice and maize or wheat. Finally, three genes, *OsRFCD001283* on chromosome 1, *OsRFCD017357* on chromosome 4, and *OsRFCD009971* on chromosome 2 were chosen in the present study. *OsRFCD001283* contains three introns and the second intron (746 bp) was used in this study. Both *OsRFCD017357* and *OsRFCD009971* have nine introns and the fourth introns (702 and 490 bp, respectively) were selected. The primers designed for amplifying these introns are listed in Table 2.

Table 2 The primers for polymerase chain reaction (PCR) amplification and sequencing

Intron	Gene name	Chromosome	Primer	Sequence (5′–3′)
Adh1-i3	*Adh1*	11	*Adh1*-F5	TCCCgTgTTCCCTCggATCTTC
			Adh1-R3	gTCACACCCTCTCCAACACTCT
Os1283	*OsRFCD001283*	1	*Os1283*U	ACgATCTCCgTggTCCCTC
			*Os1283*L	CgATTgAAgCAATggCATTC
Os17357	*OsRFCD017357*	4	*Os17357*U	gCTTgTCCgTAgAAgAgTTg
			*Os17357*L	gCCATTgTATgTCgCATTTg
Os9971	*OsRFCD009971*	2	*Os9971*U	AgAgAAAgCggTgTCAgACg
			*Os9971*L	ggCTTgTTCCATCgACgATC

We also included the third intron of *Adh1* gene, which is the largest and the most variable intron of *Adh1* gene in grasses (Ge *et al.*, 1999a; Hass *et al.*, 2003). The primers for amplification of this intron (*Adh1*-i3) (Table 2) were designed based on sequences of the third and fourth exons that were conserved across rice and maize.

DNA extraction, amplification and sequencing

Total DNA was extracted from fresh or silica gel-dried leaves, following the method as described by Ge *et al.* (1999a). Polymerase chain reaction (PCR) was performed in a total volume of 25 μl which contained 5–50 ng of template DNA, 0.2 μM of each primer, 200 μM of each dNTP, 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, and 0.75 U Ex *Taq* DNA polymerase (TaKaRa, Shiga, Japan). Amplification was carried out in a Tgradient 96 U thermocycler (Biometre, Göttingen, Germany) as follows: 2 min at 94°C followed by 38 cycles of 30 s at 94°C, 30–50 s at 54°C, 90 s at 72°C and a final extension at 72°C for 10 min. The PCR products were kept at 4°C. Because of different melting temperatures T_m, GC content and priming efficiency of different primer pairs, PCR ingredients and amplification conditions were optimized separately for each primer pair (details are available upon request). Amplification products were separated by electrophoresis on 1.5% agarose gels stained with ethidium bromide using a 100-bp DNA ladder, and gel-purified with a Pharmacia purification kit (Amersham Pharmacia Biotech, Piscataway, NJ, USA). Sequencing reactions were conducted with the forward and reverse primers using the DYEnamic ET Terminator Kit (Amersham Pharmacia Biotech), following the manufacturer's protocol. Sequencing was done on a Megabase 1000 automatic DNA sequencer (Amersham Pharmacia Biotech) after the sequencing reaction product was purified through precipitation with 95% ethanol and 3 M sodium acetate (pH 5.2). The sequences reported in this paper are deposited in GenBank under accession numbers AY749209–AY749360 (Table 1).

Data analysis

Sequences were aligned with CLUSTALX version 1.81 (Thompson *et al.*, 1997) and refined manually. Mononucleotide repeats were excluded from all phylogenetic analysis. Indels (insertion or deletion) introduced into the alignment were coded in the following ways. Shared indels were treated as single characters. Indels of uniform length were coded as absence(1)/presence(0) characters independent of the indels length. The gapped regions in the alignment were excluded from subsequent analysis unless some positions included nucleotide variation. Gaps were treated as missing entries (Petersen & Seberg, 2002).

The GC content, pairwise divergence, pairwise transition/transversion ratio (Ti/Tv), and base frequency were calculated by DAMBE version 4.1.19 (Xia & Xie, 2001). This program was also used to investigate variable substitutions over sites. The percentage of phylogenetically informative characters was calculated from each aligned data set excluding outgroups. The maximum likelihood (ML) distance and sequence divergence were calculated in MEGA2 (Kumar *et al.*, 2001), using the MODELTEST 3.06 (Posada & Crandall, 1998) estimations of both the proportion of invariable sites and the gamma distribution shape parameter.

Phylogenetic analyses of the sequence data were performed using the parsimony and Bayesian Markov chain Monte Carlo (MCMC) methods. Maximum parsimony analyses were conducted in PAUP version 4.0b10 (Swofford, 2001). Heuristic search was performed with random taxon addition and 1000 replicates, with one tree held at each step. Other tree search options included tree bisection-reconnection (TBR) branch swapping, MULPARS option, ACCTRAN optimization. Topological robustness was assessed by bootstrap analysis with 1000 replicates, using the same parameters as above (Felsenstein, 1985). Bayesian inference (BI) was conducted using MRBAYES version 3.0 (Huelsenbeck & Ronquist, 2001), with MCMC estimation of posterior probability distributions. Four chains of the MCMC were run each for 1 000 000 generations and were sampled. For all analyses, first 300 samples from each run were discarded as burn-in to ensure that the chains reached stationarity. Phylogenetic inferences were based on the trees sampled after generation 30 000. Appropriate nucleotide substitution model for each data set was determined by MODELTEST 3.06. The models were chosen according to the hierarchical likelihood ratio test (LRT).

Congruence between data sets was evaluated using the partition homogeneity test (PHT) (Farris *et al.*, 1995), as implemented in PAUP with 1000 replicates, random taxon addition (10 replicates), and one tree saved per replicate. The resulting *P*-value was used to determine whether the data sets had significant incongruence. Following Cunningham (1997), we take 0.01 as a significance

threshold for this test. Invariant characters were deleted before performing these tests, which was taken as an especially important step when the original partitions differ in the number of variable characters (Cunningham, 1997).

Divergence time of the A-genome lineage was estimated by a molecular clock, using the formula $T = K/2r$, where r corresponds to the absolute rate of substitutions/site/year and K is the estimated numbers of substitutions per site between homologous sequences. Tajima's relative rate test was used to assess heterogeneity among different lineages (Tajima, 1993). Rates of nucleotide substitution for individual gene introns were estimated by method of Nei–Gojobori with the Jukes–Cantor correction as implemented in MEGA2 program. The transposon sequences in the introns were excluded from the estimation because they might insert into the genomes at different times and may evolve at different rates.

Results

Sequences characteristics

To reconstruct reliable phylogenetic relationships, it is important to compare orthologous genes. On one hand, the BLAST search ensured that the selected target genes did not have a counterpart with a similarity > 50% in the rice genome database. On the other hand, PCR amplification produced a single band in each accession for a specific intron and no sequence heterogeneity was found after directly sequencing the PCR products. Therefore, we have reason to assume that these four genes are single-copy and the intron sequences are orthologous in the group studied.

A polyG mononucleotide repeat (5–17 bp) between positions 360 and 361 in *Adh1*-i3 sequence and a polyT repeat (5–15 bp) in *Os1283* sequence were found. Some studies indicated that uncertainty of homology in these regions could be exacerbated by potential inaccuracies of enzymatic process during PCR and sequencing (Kelchner, 2000), therefore these polyG and polyT repeats in the data sets were excluded from the phylogenetic analyses. In addition, a 152-bp insertion in *Os17357* and a 130-bp insertion in *Os9971* were detected for all the species except *O. meridionalis* and the outgroups. The insertions have some distinctive structure features similar to transposable elements (i.e. target site duplications (TSDs) and terminal inverted repeats (TIRs)). Using these insertions as the queries, we executed the BLAST searches in the TIGR rice repeat database (http://www.tigr.org/tdb/e2k1/plant.repeats/) and identified these two insertions as MITEs, one type of class-2 transposons. The 152-bp insertion in *Os17357* was identified as MITE-*stow* and the 130-bp insertion in *Os9971* was MITE-*gaijin*. Similarly, in the *Os1283* data set, two long autapomorphous insertions into *O. meridionalis*, were identified as MITE-*adhB*, including one 249-bp element between positions 279 and 280 and one 241-bp element between positions 724 and 725. Another insertion (274 bp) into the *japonica* accessions and one *O. nivara* accession (Acc. 103824) was recognized as MITE-*kiddo*. These three insertions in *Os1283* were excluded from subsequent phylogenetic analyses after they were coded as binary characters.

The sequences of *Adh1*-i3 varied in length from 774 bp to 843 bp with the aligned length of 933 bp. Of them, 141 sites (15.1%) were variable including 61 (6.54%) informative characters and 12 binary absence/presence characters derived from gap coding when the outgroups were excluded. Sizes of the *Os1283* sequences varied from 702 bp to 748 bp, excluding the insertions mentioned above, and the final aligned sequences included 781 bp long. Of these, 43 characters were phylogenetically informative (5.5%). All the aligned sequences of *Adh1*-i3 and *Os1283* belonged to the intron regions. Sizes of the *Os17357* sequences varied from 535 to 760 bp, of which 477–702 bp belonged to the intron regions. Lengths of the *Os9971* sequences ranged from 478 to 607 bp, of which 361–490 bp belonged to the intron regions. Detailed information for each data set is presented in Table 3.

Models of sequence evolution

There was little concordance between the models of sequence evolution selected for each region, in part reflecting different

Table 3 Description of sequences from the individual and combined data sets

Locus	Aligned length (range)	GC (%)	Percentage of sequence divergence (range)	Number of variable sites (%)	Number of informative sites (%)	Number of MITEs	Number of informative indels (bp)
Adh1-i3	933 (774–843)	35.4	1.58 (0–4.57)	141 (15.1%)	61 (6.5%)	0	12
Os1283	781 (702–748)	38.6	0.84 (0–2.27)	48 (6.1%)	43 (5.5%)	3	15
Os17357	777 (535–760)	34.7	1.19 (0–2.85)	44 (5.7%)	31 (4.0%)	1	1
Os9971	757 (478–607)	39.6	1.98 (0–4.89)	62 (8.2%)	57 (6.6%)	1	4
Combined *Os1283–Os17357–Os9971*	2315 (1548–2088)	37.8	1.30 (0–2.68)	154 (6.7%)	131 (5.7%)	5	20

The outgroups were excluded from the calculation. Average of pairwise sequence divergences was estimated by the best-supported model resulting from MODELTEST 3.06 (Posada & Crandall, 1998). MITEs, miniature inverted-repeat transposable elements.

Table 4 Best-supported models of molecular evolution and estimated parameters for the individual and combined data sets

			Base composition						Substitution rates						
									Ti			Tv			
	Model	–ln *L*	a	c	g	t	*I*	α	a–g	c–t	a–c	a–t	c–g	g–t	Ti : Tv
Adh1-i3	HKY+*G*	2200.19	0.21	0.16	0.21	0.41	0	0.7685	–	–	–	–	–	–	1.3056
Os1283	HKY+*G*	1749.29	0.31	0.19	0.21	0.29	0	0.5918	–	–	–	–	–	–	1. 1863
Os17357	K81uf+*G*	1727.74	0.35	0.18	0.16	0.31	0	0.5435	1.62	1.62	1.00	0.46	0.46	1.00	–
Os9971	HKY+*G*	1778.06	0.31	0.22	0.16	0.31	0	0.8386	–	–	–	–	–	–	1.3924
Combined[a]	HKY+*G*	5417.84	0.32	0.20	0.18	0.30	0	0.5369	–	–	–	–	–	–	1.2619

I is the proportion of invariant sites; α, shape parameter for gamma distribution; HKY, Hasegawa-Kishino-Yano (Hasegawa *et al*., 1985); G, γ correction (Hasegawa *et al*., 1985); K81uf, K81 model with unequal base frequencies (Kimura, 1981).
[a]Combined *Os1283*, *Os17357* and *Os9971* data set.

Table 5 Summary of the statistics from parsimony analyses with the individual and combined data sets

	Adh1-i3	*Os1283*	*Os17357*	*Os9971*	Combined[a]
Number of MPTs	2	1	48	22	160
Tree length	222	143	115	190	468
Consistency index (CI)	0.95	0.916	0.939	0.916	0.900
Retention index (RI)	0.97	0.938	0.965	0.936	0.932
Number of clades > 70% bootstrap support	8	10	8	10	14

MPTs denote the maximum parsimony trees.
[a]Combined *Os1283*, *Os17357* and *Os9971* data set.

evolutionary patterns of different genes. The best evolutionary model for *Adh1*-i3 data set was HKY+*G* (HKY, Hasegawa–Kishino–Yano; G, γ correction; Hasegawa *et al*., 1985) with equal base frequencies, a transition–transversion ratio (Ti : Tv) of 1.3056 and a gamma-distributed rate variation (gamma shape parameter α = 0.7685). Both *Os1283* and *Os9971* data sets were also best explained by this two-parameter model with some modifications for Ti : Tv and gamma shape values (Table 4). The most fitting model for *Os17357* data set was K81uf (K81 model with unequal base frequencies, Kimura, 1981) with a sharp parameter shape α of 0.5435 (Table 4).

It is obvious from Fig. 1 that nucleotide substitutions are not distributed evenly over introns, which suggests among-site variation. For *Adh1*-i3 and *Os1283*, the 3′-end of the sequence appeared to be more variable than the 5′-end (Fig. 1a,b). For *Os17357* and *Os9971*, nucleotide substitutions occurred mostly in the intron regions. The gamma-distributed rate heterogeneity of the four data matrices is listed in Table 4. It is noteworthy that two high plateaux were observed in *Os17357* (Fig. 1c) and *Os9971* (Fig. 1d), and they correspond to regions with transposons, suggesting that the variation in transposon region is high and distributed evenly. Comparison of sequence divergence between the transposon and intron regions based on Jukes–Cantor distance revealed that nucleotide diversity in transposon regions was five to eight times higher than that in intron regions (data not shown).

Phylogenetic analysis of individual data sets

Parsimony analysis of *Adh1*-i3 yielded two equally most parsimonious trees and each was 222 steps long with a consistency index (CI) of 0.950 and a retention index (RI) of 0.970 (Table 5).The strict consensus tree of the two equally parsimonious trees is shown in Fig. 2a. The Bayesian inference based on HKY+*G* model generated exactly the same topology, with only a few differences in statistical supports for some clades (Fig. 2a). Similarly, heuristic searches of the individual *Os1283*, *Os9971* and *Os17357* data sets recovered 1, 48 and 22 equally parsimonious trees of 143, 115, 190 steps in length, respectively (Table 5). The Bayesian inference based on HKY+*G* model for *Os1283* and *Os9971* data sets, and K81uf model for *Os17357* data set all generated essentially the same topologies as those by MP analysis (Fig. 2b–d). For the *Os9971* data set, minor topological differences were examined between the MP and BI trees. The Bayesian inference weakly supported two subclades of Asian species to form a monophyly with 0.60 posterior probability (PP), and weakly supported the grouping of *O. glumaepatula*, and *O. glaberrima*/*O. barthii* clade with 0.61 PP (Fig. 2d).

The strict consensus trees for each of the four analyses were identical in the following aspects: (1) a monophyly of the A-genome species was highly supported with 100% bootstrap or 1.00 PP; (2) the Asian cultivated rice (*O. sativa*) and Asian

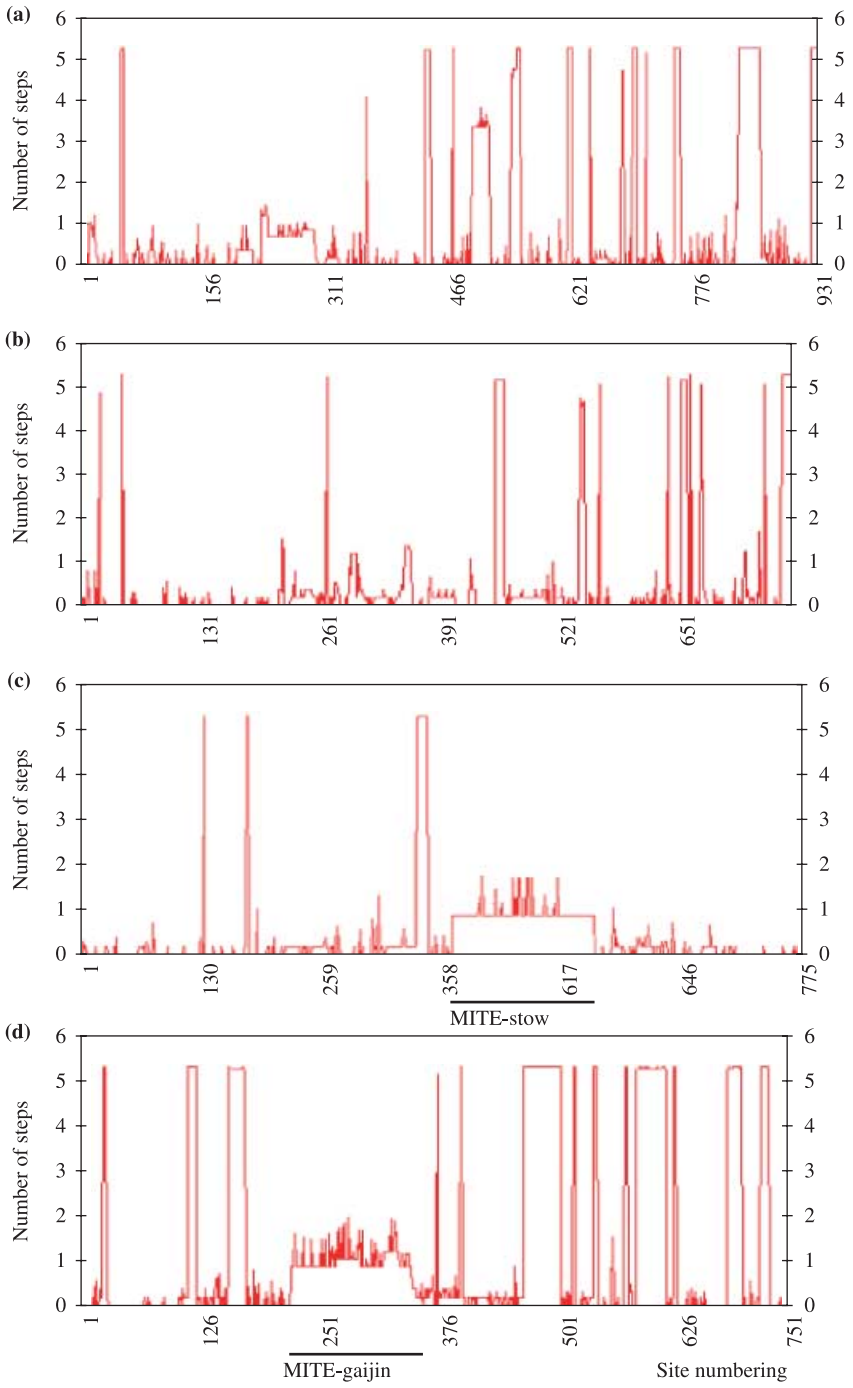

Fig. 1 The inferred number of nucleotide substitutions per site over the entire alignment. (a) *Adh1*-i3; (b) *Os1283*; (c) *Os17357*; (d) *Os9971*. The MITE (miniature inverted-repeat transposable elements) regions are noted with lines.

wild species (*O. rufipogon* + *O. nivara*) formed a monophyly with varying degrees of statistical supports; (3) five major clades were recognized including clade I consisting of the Asian cultivated rice and the *O. rufipogon*/*O. nivara* accessions, clade II of the African cultivated rice and *O. barthii* and one *O. longistaminata* (Acc. 100678), clade III of two *O. glumaepatula* accessions, clade IV of four *O. longistaminata* accessions and clade V of three *O. meridionalis* accessions (Fig. 2). Phylogenetic relationships among the five major clades are almost identical for all the four data sets with single exception that relates to the position of *O. longistaminata* (clade IV). As shown in Fig. 2, on the *Adh1*-i3 tree, *O. longistaminata* clade is sister to *O. meridionalis* clade and they form a monophyly with strong support (93% BS or 1.00 PP, Fig. 2a). By contrast, *O. longistaminata* clade does not form a group with any other A-genome species on the *Os1283*, *Os9971*, and *Os17357* trees.

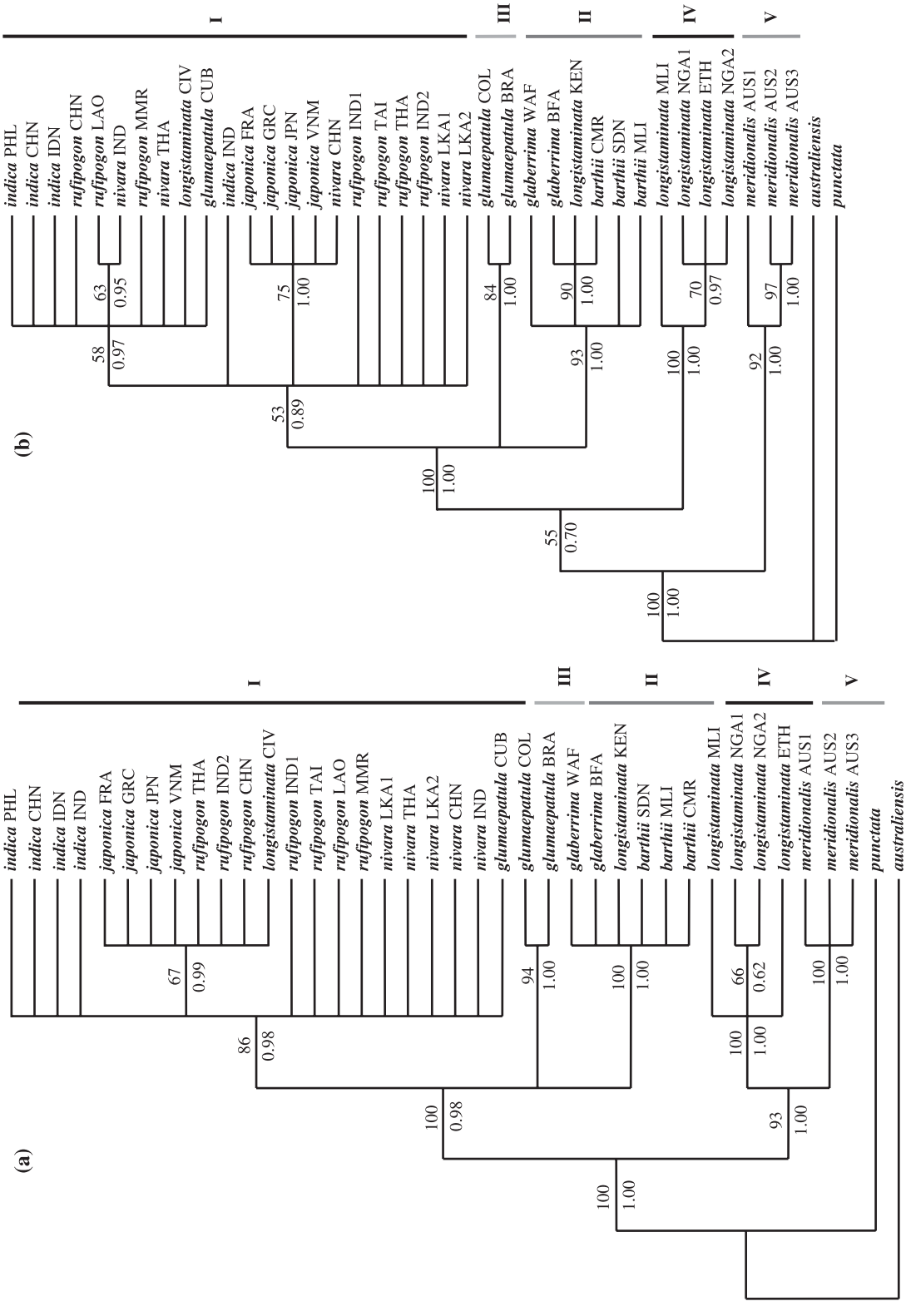

Fig. 2 Strict consensus trees resulted from the MP (maximum parsimony) analyses of individual data sets. The topologies obtained by Bayesian Inference are the same except for the nodes indicated in the figures. Bootstrap values greater than 50% are indicated above the branches and the Bayesian posterior probability are shown below the branches. Dashed lines indicated the nodes supported by Bayesian Inference. (a) Strict consensus tree of the two most-parsimonious trees (222 steps, CI = 0.950, RI = 0.970) and the best tree in Bayesian Inference (–ln *L* = 2200.19) based on the *Adh1-i3* data set. (b) The single most-parsimonious tree for the *Os1283* data set (143 steps, CI = 0.916, RI = 0.938). The same topology was obtained in Bayesian Inference (–ln *L* = 1749.29). (c) Strict consensus tree of the 48 most-parsimonious trees from the *Os17357* data set (115 steps, CI = 0.939, RI = 0.965). The same topology was obtained in Bayesian Inference (–ln *L* = 1727.74). (d) Strict consensus tree of the 22 most-parsimonious trees (190 steps, CI = 0.916, RI = 0.936) and the best tree in Bayesian Inference (–ln *L* = 1778.06) based on the *Os9971* data set.

 © *New Phytologist* (2005)

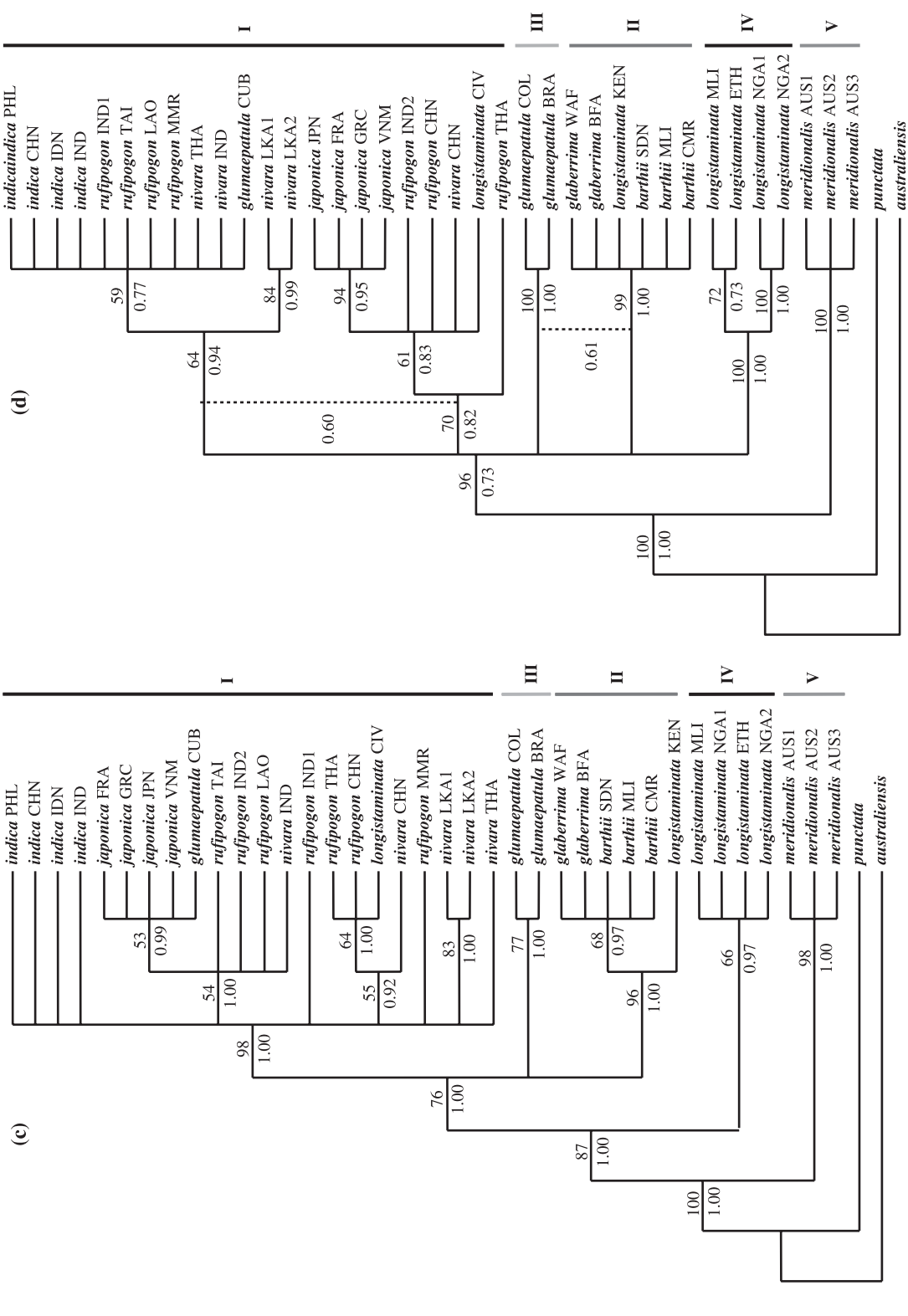

Fig. 2 *continued*

These three phylogenies show similar topologies and the *O. longistaminata* clade appears to be distantly related to *O. meridionalis* clade. Additional small differences among individual data sets affect the resolution, especially within the clade I, the Asian cultivated and wild accessions (Fig. 2).

It is noteworthy that three accessions, *O. glumaepatula*-CUB from Cuba (Acc. 100184), *O. longistaminata*-CIV from Cote d Ivoire (Acc. 101207), and *O. longistaminata*-KEN from Kenya (Acc. 104977), do not cluster with accessions from the same species. Instead, *O. glumaepatula*-CUB and *O. longistaminata*-CIV are nested into *O. sativa*/*O. rufipogon*/*O. nivara* clade (clade I) and *O. longistaminata*-KEN is grouped with *O. glaberrima*/*O. barthii* accessions (see the Discussion section).

Phylogenetic analyses of combined data sets

For four individual data sets, the partition homogeneity test (PHT) was executed across the entire data set. In general, there was no significant incongruence between any two partitions ($P > 0.03$) but the *P*-value was significant ($P = 0.01$) when the PHT was conducted for the combined four data sets, while the *P*-value for the combined *Os1283*, *Os17357* and *Os9971* data sets was 0.05, suggesting that the incongruence among data sets resulted from *Adh1*-i3. Owing to the obvious basal conflict between *Adh1*-i3 phylogeny and the phylogenies of the other three introns, we chose to combine the *Os1293*, *Os17357*, and *Os9971* data sets. The combined data set was 2315 bp long, including 131 (8.5%) informative characters and 20 binary (absence/presence) characters (Table 3). The best-fitting evolutionary model for the combined data matrix was HKY+*G* with a sharp parameter shape α of 0.5369 (Table 4).

The tree obtained from the combined phylogeny depicts very similar relationships whether using MP or BI methods, with only a few differences in bootstrap (BS) and Bayesian PP for some clades (Fig. 3). The combined data set also provides robust support for each of the five main clades with almost 100% BS or 1.00 PP supports, and indicates that *O. meridionalis* is the basal lineage (99% BS or 1.00 PP support). It is interesting that clade I is divided into two subclades, and two subspecies of *O. sativa* (*O. sativa* ssp. *japonica* and *O. sativa* ssp. *indica*) are not grouped in the same subclade. All the *japonica* accessions, three *O. rufipogon* accessions, each of *O. nivara* and *O. longistaminata* (Acc. 101207) accessions form one subclade with a high support (99% BS, 1.00 PP), which is sister to the other subclade consisting of all the *indica* accessions, four *O. rufipogon* and four *O. nivara* accessions with a slightly lower support (64% BS, 0.98 PP). Two *O. glumaepatula* accessions (clade III) are sister to the African *O. barthii*/*O. glaberrima* clade with weak support (0.77 PP) on the combined Bayesian phylogeny. It is obvious that the combined analysis both improves the resolution and increases the number of well-supported clades compared with the individual analyses of four data sets (Figs 2 and 3).

MITE insertions as phylogenetic markers

We detected several MITE insertions in three introns, including MITE-*gaijin*, MITE-*stow*, MITE-*kiddo* and MITE-*adhB*. Of these, MITE-*kiddo* and MITE-*adhB* in *Os1283* inserted one or two specific species and therefore were not used further in the analysis. It is interesting to note that MITE-*gaijin* in *Os9971* and MITE-*stow* in *Os17357* all insert into the A-genome species except for *O. meridionalis* (Fig. 3). In addition, we found another transposable element (MITE-*stow*) in intron 4 of nuclear gene *OsRFCD014662 (abbr. Os14662)* that is on chromosome 3. This insertion is also recognized in all the A-genome species except for *O. meridionalis* (Fig. 3). These last three insertions provide important information on the basal split in the A-genome group.

Based on previous and present studies, three hypotheses on the basal lineage of the A-genome species are presented in Fig. 4 and could be speculated as follows. (1) If *O. meridionalis* is the basal lineage (Fig. 4a), the most parsimonious explanation of the occurrence for a MITE requires one insertion event, as denoted by the arrow. Because one MITE-*gaijin* and two MITE-*stow* elements are present in all the A-genome species except for *O. meridionalis* and the outgroups, these three MITEs would insert into the common ancestor of the other seven species after its separation from *O. meridionalis* during early divergence of the A-genome group. (2) If *O. longistaminata* and *O. meridionalis* form a monophyletic group and are sister to all the other A-genome species (Fig. 4b), as revealed by *Adh1*-i3 phylogeny, the most parsimonious scenario requires at least two evolutionary events for each MITE: either two independent insertions in *O. longistaminata* and the progenitor of the other three clades including five species (closed, downward arrows in Fig. 4b) or one insertion in the progenitor of the A-genome species and one excision from *O. meridionalis* (open, upward arrows in Fig. 4b). (3) If *O. longistaminata* is the basal most divergent species, at least two evolutionary events are needed to explain the occurrence of each MITE (Fig. 4c). Taken together, the phylogeny demonstrated in Fig. 4a that *O. meridionalis* as basal lineage of the A-genome group is the most parsimonious explanation because it includes three evolutionary events in total for the occurrence of the three MITEs, while at least six events needed to be included if other phylogenetic relationships occur (Fig. 4b,c). Therefore, insertion events of three MITEs in *Os9971*, *Os17357* and *Os14662* introns strongly support the basal position of *O. meridionalis*, in agreement with the phylogenies generated by intron sequences of *Os1283*, *Os17357* and *Os9971* (Fig. 2b–d).

Discussion

MITE insertions and the basal lineage of the A-genome species

The A-genome species in the genus *Oryza* have the widest geographic distribution of any rice genome. Evidence shows

 © *New Phytologist* (2005)

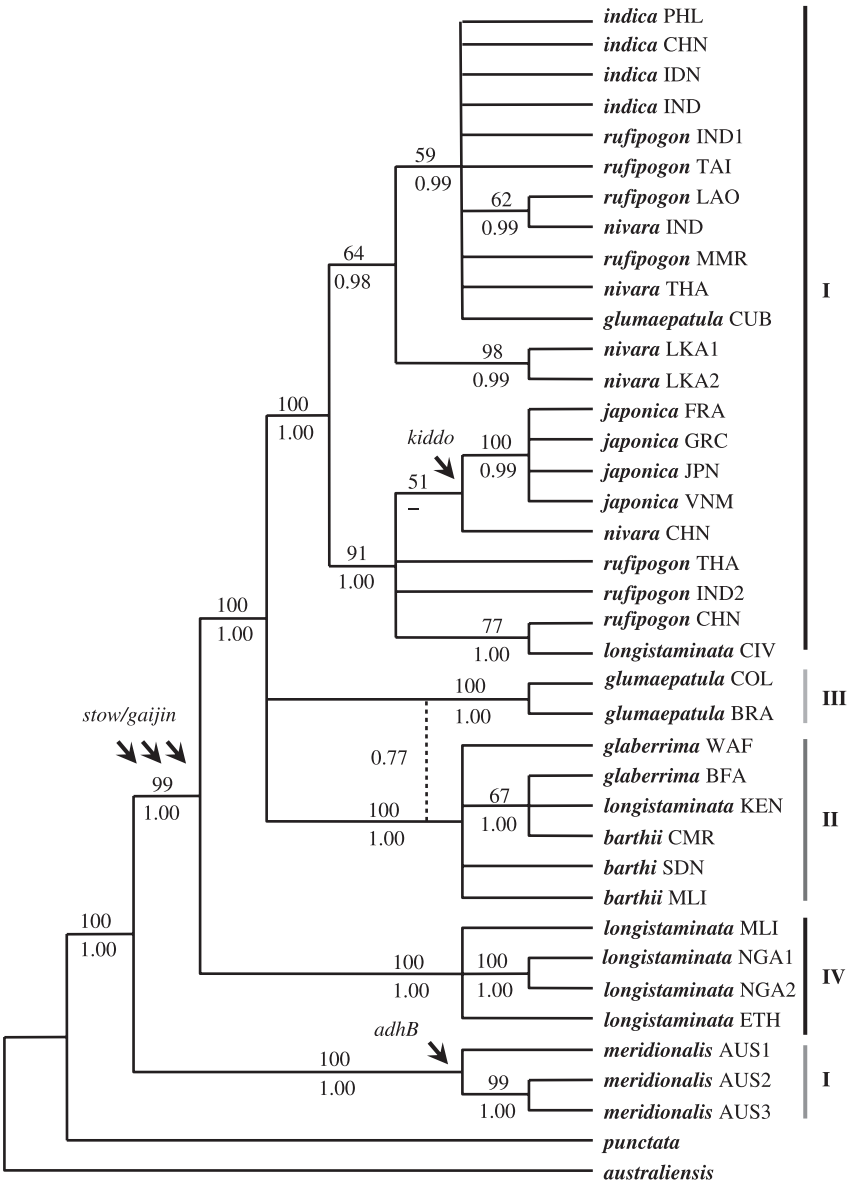

Fig. 3 Strict consensus tree of the 160 most-parsimonious trees generated from the combined *Os1283*, *Os17357* and *Os9971* data sets (468 steps, CI = 0.900, RI = 0.930). A similar topology was obtained in Bayesian Inference (–ln *L* = 5417.84). The evolutionary model used in the BI analysis was HKY+*G* (Hasegawa *et al*., 1985) with a transition–transversion ratio of 1.2619 and the sharp parameter α of 0.5369. Bootstrap percentages > 50% are indicated above the branches and the values below the branches are Bayesian posterior probability. Dashed lines indicate the nodes resulting from Bayesian Inference. The MITE (miniature inverted-repeat transposable elements) insertions are marked by thick arrows, with their types indicated.

that this group is a distinct and monophyletic lineage (Wang *et al*., 1992; Aggarwal *et al*., 1999; Ge *et al*., 1999a). This is further demonstrated by the present study. Although most previous studies demonstrated that *O. meridionalis* and *O. longistaminata* were two species distinctly related to other A-genome species (Wang *et al*., 1992; Ishii *et al*., 1996; Lu *et al*., 2000; Morishima, 2001; Cheng *et al*., 2003; Park *et al*., 2003), two different opinions have been proposed regarding the basal lineage in this group. Some authors proposed that *O. longistaminata* was the earliest divergent species in the A-genome group and thus suggested that the A-genome might have originated in Africa where *O. longistaminata* occurs (Iwamoto *et al*., 1999; Cheng *et al*., 2002; Ren *et al*., 2003). Conversely, others indicated that the Australian *O. meridionalis* was the basal-most lineage (Wang *et al*., 1992; Ge *et al*., 1999a; Park *et al*., 2003). Current phylogenetic analysis based on *Os1283*, *Os17357* and *Os9971* data sets all place *O. meridionalis* as the basal lineage whereas *O. longistaminata* and all the other species form a sister lineage to *O. meridionalis* (Fig. 2b–d) despite *Adh1*-i3 phylogeny showing a different topology with *O. longistaminata* and *O. meridionalis* forming one monophyletic clade at the basal position (Fig. 2a). This topological incongruence might reflect intrinsic evolutionary properties of genes, or it might instead reflect an important feature of the A-genome evolution, namely, a relatively rapid diversification after its origin.

Accumulating literature indicates that phylogenetic incongruence is almost the rule rather than the exception (Wendel & Doyle, 1998) and can be attributed to many different causes, such as gene choice and insufficient data, convergent

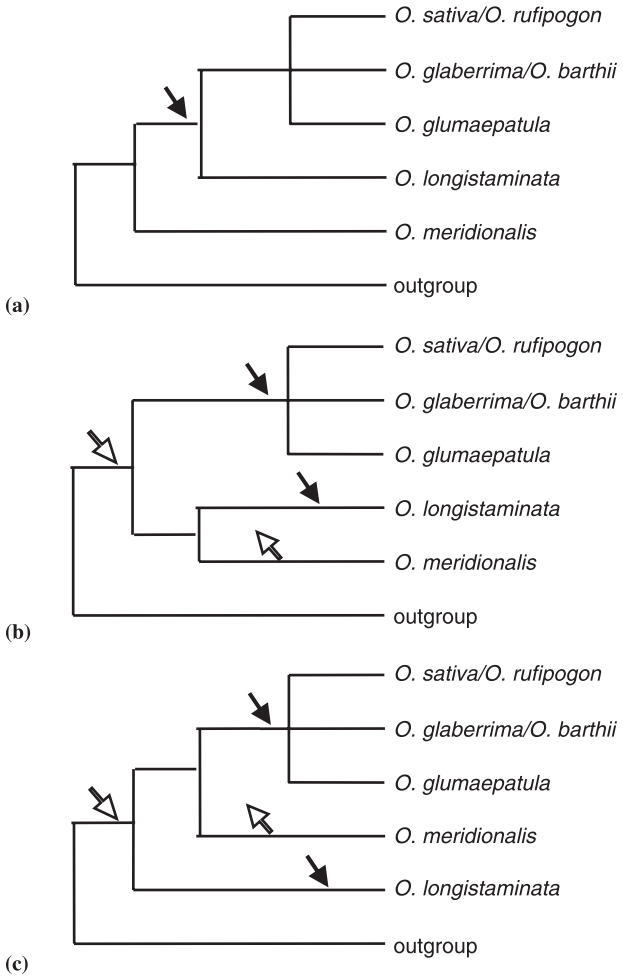

Fig. 4 Three phylogenetic hypotheses regarding the basal lineage of the A-genome species. Downward and upward arrows stand for the insertion and excision of MITEs (miniature inverted-repeat transposable elements), respectively. The closed and open arrows indicate different insertion or deletion events, respectively. Because three MITEs (one MITE-*gaijin* and two MITE-*stow*) were found to insert into all the A-genome species except for *Oryza meridionalis*, different insertion and deletion events are needed to account for the existence of each of the three MITEs under three phylogenetic hypotheses. (a) When *O. meridionalis* is at the basal position, the most parsimonious explanation of the occurrence for a MITE requires only one insertion event, as denoted by the arrow. (b) When *Oryza longistaminata* and *O. meridionalis* form a monophyletic group and is sister to all the other A-genome species, at least two evolutionary events requires for accounting for the occurrence of a MITE (i.e. two independent insertions in *O. longistaminata* and the progenitor of the other three clades (closed arrows), or alternatively one insertion into the progenitor of the A-genome species and one deletion from *O. meridionalis* (open arrows)). (c) When *O. longistaminata* is at basal, at least two evolutionary events needed to explain the occurrence of a MITE (i.e. either two independent insertions (closed arrows), or one insertion in the progenitor of the A-genome species and one deletion from *O. meridionalis* (open arrows)).

evolution, rapid diversification, hybridization and introgression and lineage sorting (Kellogg *et al.*, 1996; Wendel & Doyle, 1998). Because the sequences we used represent multiple loci from different chromosomes and evolve rapidly, insufficient data seems not be a factor for the incongruence. However, hybridization/introgression and lineages sorting are most likely to cause the discrepancy in the position of *O. longistaminata* because these processes often occur at lower taxonomic ranks (Wendel & Doyle, 1998). Whether this incongruence reflects underlying biological phenomena or some other unknown artefacts needs to be further investigated.

In this study, we found a number of insertion events of transposons that provide additional informative characters to resolve the inconsistency of basal lineage in the group. As shown in Fig. 3, three MITEs (one MITE-*gaijin*, two MITE-*stow*) insertions are detected in all the A-genome species except for *O. meridionalis*. More importantly, the three MITE insertions occur in three unlinked genes on different chromosomes (chromosomes 2, 3 and 4). Therefore, these insertion patterns can only be explained by *O. meridionalis* being the earliest divergent lineage that separated from all other A-genome species at the early stage of the A-genome radiation (Fig. 4). Growing evidence demonstrates that MITEs with a known position in the genome might provide phylogenetic signals as long as they are homologous (Iwamoto *et al.*, 1999; Petersen & Seberg, 2002). Because frequency of transposition of MITEs is generally low, the loci containing them are free from selection pressure and are useful markers to infer phylogenetic relationships based on the presence or absence of these elements (Wessler *et al.*, 1995; Kanazawa *et al.*, 2000). Kanazawa *et al.* (2000) indicated that the pattern of the presence/absence of MITEs was highly associated with ecotypes in *O. rufipogon*. Based on the MITE-transposon display, Park *et al.* (2003) evaluated the genetic variation and species relationships of the A-genome species. The present study further demonstrates that the presence/absence of MITEs provides important and sometimes critical information for phylogenetic inference.

It should be noted that two *O. longistaminata* accessions (101207 and 104977) do not cluster with other *O. longistaminata* accessions, and instead are nested in the *O. sativa*/*O. rufipogon* clade (clade I) and *O. glaberrima*/*O. barthii* clade (clade II), respectively (Figs 2 and 3). It is of interest to note that previous phylogenetic studies based on nuclear *Adh1* and *Adh2* sequences (Ge *et al.*, 1999a) and the *P-SINE1*-like intron of the *CatA* catalase homologues (Iwamoto *et al.*, 1999) also revealed that *O. longistaminata* 104977 was evolutionarily close to the *O. glaberrima*/*O. barthii* group. Therefore, it is most likely that accession 104977 was misidentified as *O. longistaminata* and should be renamed as *O. barthii*. Using MITE-AFLP markers, Park *et al.* (2003) also detected a misidentified *O. longistaminata* accession that proved to be *O. barthii*. However, as the most widely distributed *Oryza* species in Africa, *O. longistaminata* is partly self-incompatible perennial and can hybridize with *O. glaberrima* and *O. sativa*

(Ghesquiere, 1986; Vaughan, 1989). Therefore, hybridization and introgression between *O. longistaminata* and the other A-genome species may also result in the unexpected position of the two accessions.

Relationship between *O. rufipogon* and *O. nivara* and the origin of the Asian cultivated rice

Nomenclature of the Asian wild taxa with the A-genome has long been a subject of controversy (Nayar, 1973; Oka, 1988; Vaughan, 1989). The perennial and annual types of *O. rufipogon* were traditionally classified as two independent species (*O. rufipogon* and *O. nivara*) because they were morphologically different and exhibited contrasting life-history traits that characterize fecundity/survivorship of the individuals (Nayar, 1973; Morishima, 2001). The perennial *O. rufipogon* is widely distributed in southern China (up to Jiangxi Province), South and South-east Asia and northern Australia (Vaughan, 1989). It inhabits in stable water condition through the year, such as swamps, deep and big ponds or ditches. By contrast, the annual *O. nivara* is mainly found in South and South-east Asia, and occurs in seasonally dry/wet areas such as ponds, swamps and the vicinity of rice fields with shallow water (Chang, 1976). Despite the differences between the two species in their life-history traits, previous artificial hybridization suggested that the two species did not have apparent reproductive isolation (Oka, 1988; Lu *et al.*, 2000). In addition, life-history traits characterizing them vary continuously in nature and segregate in an F_2 population, suggesting that those traits are controlled by multiple factors (Morishima, 2001). Therefore, they were mostly treated as two different ecotypes under a single species (*O. rufipogon*) (Oka, 1988; Morishima *et al.*, 1992; Lu *et al.*, 2000; Cheng *et al.*, 2003).

In the present study, we sampled seven accessions of *O. rufipogon* and five accessions of *O. nivara* that cover the entire geographical distribution of the two species. Phylogenetic analyses based on individual and combined data sets show clearly that they are scattered in two subclades within clade I and mix with each other along with *O. sativa* accessions (Figs 2 and 3), which indicates their close genetic relationships. On each of phylogenetic trees, *O. rufipogon* and *O. nivara* accessions do not show any distinct differentiation at the species level or in geographical distribution. Therefore, our results support combining the two species in a single group (*O. rufipogon*) as two ecotypes of a large species complex (Oka, 1988; Morishima *et al.*, 1992).

The origin of the Asian cultivated rice (*O. sativa*) is another unsolved question for decades. There are two different hypotheses regarding its origin. Monophyletic hypothesis is mainly based on the fact that wild rice has a potentiality to evolve into *indica* and *japonica* types and differentiation between *indica* and *japonica* is not found in *O. rufipogon* (Chang, 1976; Wang *et al.*, 1992; Lu *et al.*, 2002). By contrast, polyphyletic hypothesis postulates that *indica* and *japonica* types originate from different lineages of *O. rufipogon* because a clear differentiation exists between them and they are closer to different wild accessions than to each other (Second, 1982; Wang *et al.*, 1992; Mochizuki *et al.*, 1993; Cheng *et al.*, 2002; Ren *et al.*, 2003; Yamanaka *et al.*, 2003). Our results are congruent with polyphyletic hypothesis because the *O. sativa*/*O. rufipogon*/*O. nivara* clade falls into two apparent subclades to which *japonica* and *indica* accessions belong separately (Fig. 3). Further sampling on a large scale would be necessary to clarify the subdivision and origin of specific subspecies or strains.

Origin of the African cultivated rice and phylogenetic position of *O. glumaepatula*

Compared with wide distribution of the Asian cultivated rice (*O. sativa*), the African cultivated rice (*O. glaberrima*) is cultivated only in local agricultural ecosystems in West Africa (Nayar, 1973). A wealth of investigations has shown that *O. barthii* is the immediate progenitor of *O. glaberrima* (Oka, 1988; Morishima *et al.*, 1992.; Wang *et al.*, 1992; Park *et al.*, 2003), which is strongly supported by the present study. Furthermore, our results suggest that *O. longistaminata*, another African wild species, is distantly related to *O. glaberrima* and *O. barthii*, though their distribution is sympatric in some areas. These three species from Africa do not form a monophyly (Fig. 2), implying different divergence times in history. Present multiple gene phylogenies suggest that *O. longistaminata* might diverge much earlier than the *O. glaberrima*/*O. barthii* separation that occurred long after the divergence between the *O. glaberrima*/*O. barthii* group and the *O. sativa*/*O. rufipogon*/*O. nivara* group (Fig. 3). Cultivated rice (*O. sativa* and *O. glaberrima*) was considered as an example of parallel evolution in crop plants with *O. rufipogon* and *O. barthii* as the wild progenitors of *O. sativa* and *O. glaberrima*, respectively (Oka, 1988). This parallel evolution was reported by many previous studies (Morishima *et al.*, 1992; Cheng *et al.*, 2002) and is confirmed in this study.

Oryza glumaepatula is endemic to America. Although no clear morphological characteristic distinguishes *O. glumaepatula* from other A-genome species from Asia and Australia (Juliano *et al.*, 1998), sterility barriers have been detected between this species and the Asian A-genome species (Naredo *et al.*, 1998). The RAPD analysis also revealed a great genetic differentiation between the Chinese *O. rufipogon* and the Brazilian *O. glumaepatula* populations (Ge *et al.*, 1999b). Of three *O. glumaepatula* accessions sampled, 105561 from Colombia and 105672 from Brazil form a clade (clade III) with high statistical supports (Figs 2 and 3). The combined Bayesian phylogeny indicates that *O. glumaepatula* clade (clade III) shows closer affinity with the African *O. barthii*/*O. glaberrima* clade than with the Asian *O. sativa*/*O. rufipogon*/*O. nivara* clade (Fig. 3), but the statistical support is slightly lower and this affinity need to be further justified with larger samples.

In this study, one *O. glumaepatula* accession (100184) from Cuba is nested into clade I rather than grouped with the other

two *O. glumaepatula* accessions. The unexpected position of one *O. glumaepatula* accession (100184) might arise from misidentification or hybridization between *O. glumaepatula* and the Asian species. Based on RAPD profiles, Martin *et al.* (1997) revealed that four *O. glumaepatula* accessions from Cuba (100184), Venezula (103810) and Brazil (101960 and 104387), respectively, should be classified as *O. rufipogon*. Wang *et al.* (1992) also found that one *O. glumaepatula* accession (W1185) was misidentified as *O. rufipogon*. Therefore, misidentification is common in this complex group. By contrast, well-documented cases of natural hybrids were found between the Asian *O. sativa* and *O. glumaepatula* in Cuba (Chu & Oka, 1970; Ghesquiere, 1986), and thus hybridization and introgression cannot be ruled out as reasons for the position of accession 100184. Further sampling, in conjunction with morphological investigation, would be necessary to clarify its taxonomic identity.

Origin and radiation of the A-genome species in *Oryza*

A wide pantropical distribution of the A-genome species raises a long-standing and interesting question regarding when and how these species diverged and became distributed throughout all the continents (Nayar, 1973; Chang, 1985; Second, 1991). Previous authors have attempted to date the various events in the evolution of the genus *Oryza* but their estimations have differed greatly (Chang, 1976; Second, 1991). For example, the divergence time between the African *O. longistaminata* and Asian *O. rufipogon* was estimated from 5 to 7 million years ago (mya) (Barbier *et al.*, 1991; Second, 1991) to more than 100 mya (Chang, 1976, 1985).

To obtain insights into the origin and biogeographic history of the A-genome species in a phylogenetic perspective, we employed a molecular clock approach to estimate the divergence time of this lineage. As the analysis is based on the assumption that all the lineages evolve homogeneously, we first evaluated whether substitution rates are homogeneous within the A-genome and between the A- and B-genome species using Tajima's relative rate test (Tajima, 1993). The results showed that the substitution rate variations are not significant for *Adh1*-i3, *Os1283* and *Os17357*, whereas significant heterogeneity ($P < 0.05$) is observed in *Os9971*. Therefore, the molecular-clock approach discussed later is applied to the *Adh1*-i3, *Os1283*, and *Os17357* data sets.

One important aspect of divergence time estimation is clock calibration. Because the sequence divergence rates of the introns used in this study are unknown, as an alternative we estimated absolute divergence time by applying substitution rates derived from *Adh1* nuclear genes. The fossil-calibrated synonymous rate of *Adh1* divergence is 7.0×10^{-9} substitutions per site per year for grasses (Gaut *et al.*, 1996). The average divergence rate for *Adh1*-i3 intron sequences is calculated to be approximately 1.29 times as fast as that of *Adh1* synonymous sites, so the substitution rate of *Adh1* intron is estimated to be 9.0×10^{-9} substitution per site per year. Using this estimate of absolute rate and the sequence divergence estimate of *Adh1*-i3, we calculated approximate divergence times for various events within the A-genome group. The mean sequence divergence among the three main lineages (*O. meridionalis*, *O. longistaminata* and the remaining species) is 3.66%, which translates into a divergence of 2.0 mya. Based on the same method, the Asian cultivated rice (*O. sativa*) diverged from the African cultivated rice (*O. glaberrima*) about 0.7 mya, while two subspecies of *O. sativa* (ssp. *indica* and ssp. *japonica*) separated approximately 0.4 mya. Date calibrations on the basis of *Os17357* and *Os1283* molecular clocks provide similar estimations for the radiation of this group (data not shown).

If 2 mya could be the divergence time for the A-genome group, it seems possible that the A-genome lineage radiated in the Mid-Pleistocene and extant A-genome species diversified so recently that they might only achieve their global distributions by long-distance, transoceanic dispersals rather than by the vicariance resulting from the fracture and drift of the Gondwana supercontinent suggested by Chang (1976, 1985). Long-distance dispersal events were also suggested for other recently diverged lineages of the genus *Gossypium* (Wendel & Albert, 1992; Seelanan *et al.*, 1997). By contrast, the divergence of the *indica* and *japonica* subspecies dated to 0.4 mya, long before the domestication of the two crops (Oka, 1988) and is consistent with the polyphyletic hypothesis of independent domestication of *indica* and *japonica* subspecies.

Although there are a number of limitations to the use of clocks based on sequence data, including substitution rate heterogeneity among lineages, uncertainties of clock calibration and other potential sources for estimation errors (Seelanan *et al.*, 1997; Gaut, 2002), the divergence events for the A-genome group estimated in the present study provide a rough timeframe for understanding of the evolutionary tempos and biogeographic history of this important plant group.

Utility of the intron sequences of nuclear genes

At the lower taxonomic level, noncoding sequences are sources of choices as phylogenetically informative characters. Of various noncoding regions, nuclear introns are the best candidates because they evolve fast and thus provide sufficient resolution and because they are flanked by conserved coding sequences from which primers can be designed to amplify the introns in a broader array of taxa (Doyle *et al.*, 1996). To date, there have been many examples where nuclear introns have been successfully used for phylogenetic reconstruction at lower taxonomic levels (Doyle *et al.*, 1996; Bailey & Doyle, 1999; Howarth & Baum, 2002; Oh & Potter, 2003; Grob *et al.*, 2004). This study shows that the intron sequences of nuclear genes provide sufficient information for resolving relationships of the A-genome species.

Table 6 Comparisons of sequence divergence and phylogenetic information of *mat*K, *trn*L-*trn*F, nrITS and intron data sets

	Aligned length (bp)	GC (%)	Percentage of sequence divergence (range)	Number of variable sites (%)	Number of informative sites (%)
ITS	622	70.3	2.15 (0–5.97)	44 (7.1%)	11 (1.8%)
*mat*K	1558	34.2	0.54 (0–1.47)	34 (2.2%)	8 (0.5%)
*trn*L-*trn*F	891	33.7	0.6 (0–2.36)	24 (2.7%)	3 (0.3%)
Adh1-i3	933	35.4	6.51 (0–15.58)	174 (18.6%)	73 (7.8%)
Os1283	762	38.6	4.33 (0–13.21)	100 (13.1%)	36 (4.7%)
Os17357	776	34.7	4.01 (0–10.27)	93 (12.0%)	28 (3.6%)
Os9971	757	39.6	8.05 (0–25.33)	155 (20.5%)	42 (5.5%)

The species compared include *japonica* (Nipponbare), *Oryza rufipogon* (105942), *Oryza nivara* (103824), *Oryza glumaepatula* (105561), *Oryza glaberrima* (103474), *Oryza longistaminata* (105061), *Oryza barthii* (104140), *Oryza meridionalis* (105306), *Oryza punctata* (105984) and *Oryza australiensis* (105263). Average of pairwise sequence divergences was estimated using the Jukes–Cantor distance.

In order to evaluate the phylogenetic utility of these four nuclear introns in the A-genome species, we contrasted the introns with several commonly used DNA regions (i.e. nuclear ribosomal ITS, chloroplast *mat*K and *trn*L-*trn*F intergenic spacer). We randomly selected one accession from each species to sequence ITS, *mat*K and *trn*L-*trn*F (GenBank Accession Numbers AY749361–AY749374) because sequence variation within species was low and initial phylogenetic analysis revealed that sequences from the same species clustered together (Fig. 3). The aligned length, number of parsimony informative characters, and average pairwise sequence divergence for each of these fragments are listed in Table 6. For these same taxa, percentages of informative characters in introns were all more than two times higher (7.82%, 4.72%, 3.61% and 5.55%) than those of cpDNA (0.34–0.51%) and ITS (1.77%). It is obvious that the four introns provided a higher proportion of phylogenetic information than that of ITS and cpDNA.

It is suggested that different genes have different information content and a single nuclear marker could not diverge at a rate appropriate for resolving phylogenetic issues (Bremer *et al.*, 1999). Also, growing evidence has showed that reliance on a single data set might often result in misleading phylogenetic inferences because the gene tree does not necessarily reflect the species tree owing to many technical reasons and biological processes (Wendel & Doyle, 1998). This has led to the recognition that sequences from multiple, preferably unlinked loci to track organismal history independently. In the present study, introns of the nuclear genes on different chromosomes are used as unlinked and independent characters to elucidate the phylogenetic relationships of the A-genome species. The results demonstrate clearly that these introns provide well-supported gene genealogies. Therefore, fast-evolving nuclear gene introns are likely to become important markers for phylogenetic reconstruction at lower taxonomic levels such as the A-genome group and many other closely related groups in general.

Acknowledgements

We thank Xiao-Li Shi for her help in the genome database searches, Ying Bao, Ya-Long Guo, Lin-Bin Zhang and Fu-Hong An for their technical assistance and Tong-ming Yin and Yan-ru Zeng for helpful suggestions. We are also grateful to two anonymous reviewers for their critical comments and suggestions, and to the International Rice Research Institute (Los Banos, Philippines) for providing seed samples and Hai-Rong Wu for providing *Oryza sativa* ssp. *indica* (ZH249) leaves. This research was supported by the National Natural Science Foundation of China (30025005, 30121003), the Chinese Academy of Sciences (kscxz-sw-101 A) and Program for Key International S & T Cooperation project of P. R. China (20011CB711103).

References

Aggarwal RK, Brar DS, Nandi S, Huang N, Khush GS. 1999. Phylogenetic relationships among *Oryza* species revealed by AFLP markers. *Theoretical and Applied of Genetics* **98**: 1320–1328.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Bailey CD, Doyle JJ. 1999. Potential phylogenetic utility of the low-copy nuclear gene. *Pistillata* in dicotyledonous plants: comparison to nrDNA ITS and trnL intron in *Sphaerocardamum* and other Brassicaceae. *Molecular Phylogenetics and Evolution* **13**: 20–30.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82**: 247–277.

Barbier P, Morishima H, Ishihama A. 1991. Phylogenetic relationships of annual and perennial wild rice probing by direct DNA sequencing. *Theoretical and Applied of Genetics* **81**: 693–702.

Bautista NS, Solis R, Kamijima O, Ishii T. 2001. RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice. *Genes and Genetic Systems* **76**: 71–79.

Bremer B, Jansen RK, Oxelman B, Backlund M, Lantz H, Kim KJ. 1999. More characters or more taxa for a robust phylogeny: case study from the coffee family (Rubiaceae). *Systematic Biology* **48**: 413–435.

Chang TT. 1976. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rice. *Euphytica* **25**: 435–441.

Chang TT. 1985. Crop history and genetic conservation: rice-a case study. *Iowa State Journal of Research* **59**: 425–455.

Cheng CY, Motohashi R, Ohtsubo E. 2003. Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. *Molecular Biology and Evolution* **20**: 67–75.

Cheng CY, Tsuchimoto S, Ohtsubo H, Ohtsubo E. 2002. Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. *Genes and Genetic Systems* **77**: 323–334.

Chu YE, Oka HI. 1970. Introgression across isolating barriers in wild and cultivated *Oryza* species. *Evolution* **24**: 344–355.

Cunningham CW. 1997. Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution* **14**: 733–740.

Doyle JJ, Kanazin V, Shoemaker RC. 1996. Phylogenetic utility of histone H3 intron sequences in the perennial relatives of soybean (*Glycine*: Leguminosae). *Molecular Phylogenetics and Evolution* **6**: 438–447.

Farris JS, Kallersjo M, Kluge AG, Bult C. 1995. Testing significance of incongruence. *Cladistics* **10**: 315–319.

Felsenstein J. 1985. Confidence limits on phylogeny: an approach using the bootstrap. *Evolution* **39**: 783–791.

Gaut BS. 2002. Evolutionary dynamics of grass genomes. *New Phytologist* **154**: 15–28.

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*. *Proceedings of the National Academy of Sciences, USA* **93**: 10274–10279.

Ge S, Sang T, Lu BR, Hong DY. 1999a. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proceedings of the Natural Academy of Sciences, USA* **96**: 14400–14405.

Ge S, Oliveira GCX, Schaal BA, Gao LZ, Hong DY. 1999b. RAPD variation within and between natural populations of wild rice (*Oryza rufipogon*) from China and Brazil. *Heredity* **82**: 638–644.

Ge S, Sang T, Lu BR, Hong DY. 2001. Phylogeny of the genus *Oryza* as revealed by molecular approaches. In: Khush GS, Brar DS, Hardy B, eds. *Rice genetics IV. Proceedings of the fourth international rice genetics symposium.* Los Banos, The Philippines: IRRI, 89–105.

Ghesquiere A. 1986. Evolution of *Oryza longistaminata*. In: *Rice Genetics.* Manila, The Philippines: IRRI, 15–25.

Grob GBJ, Gravendeel B, Eurlings MCM. 2004. Potential phylogenetic utility of the nuclear *FLORICAULA/LEAFY* second intron: comparison with three chloroplast DNA regions in *Amorphophallus* (Araceae). *Molecular Phylogenetics and Evolution* **30**: 13–23.

Hasegawa M, Kishino H, Yano T. 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160–174.

Hass BL, Pires JC, Porter R, Phillips RL, Jackson SA. 2003. Comparative genetics at the gene and chromosome levels between rice (*Oryza sativa*) and wildrice (*Zizania palustris*). *Theoretical and Applied of Genetics* **107**: 773–782.

Howarth DG, Baum DA. 2002. Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. *Molecular Phylogenetics and Evolution* **23**: 525–528.

Huelsenbeck JP, Ronquist FR. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754–755.

Ishii T, Nakano T, Maeda H, Kamijima O. 1996. Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis. *Genes and Genetic Systems* **71**: 195–201.

Ishii T, Xu Y, McCouch S. 2001. Nuclear- and chloroplast-microsatellite variation in A-genome species of rice. *Genome* **44**: 658–666.

Iwamoto M, Nagashima H, Nagamine T, Higo H, Higo K. 1999. *P-SINE1*-like intron of the *CatA* catalase homologs and phylogenetic relationships among AA genome *Oryza* and related species. *Theoretical and Applied of Genetics* **98**: 853–861.

Juliano AB, Naredo MEB, Jackson MT. 1998. Taxonomic status of *Oryza glumaepatula* Steud. I. Comparative morphological studies of New World diploids and Asian AA-genome species. *Genetic Resources and Crop Evolution* **45**: 197–203.

Kanazawa A, Akimoto M, Morishima H, Shimamoto Y. 2000. Inter- and intra-specific distribution of *Stowaway* transposable elements in AA-genome species in rice. *Theoretical and Applied of Genetics* **101**: 327–335.

Kelchner SA. 2000. The evolution of noncoding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* **87**: 482–498.

Kellogg EA, Appels R, Mason-Gamer RJ. 1996. When genes tell different stories: the diploid genera of Triticeae (Gramineae). *Systematic Botany* **21**: 321–347.

Kimura M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences, USA* **78**: 454–458.

Kumar S, Tamura K, Jakobsen IB, Nei M. 2001. *MEGA2.1 molecular e volutionary genetics analysis software.* Tempe, AZ, USA: Arizona State University

Lu BR, Naredo MBE, Juliano AB, Jackson MT. 2000. Preliminary studies on taxonomy and biosystematics of the AA-genome. *Oryza* species (Poaceae). In: Jacobs SWL, Everett J, eds. *Grasses, systematics and evolution.* Melbourne, Australia: CSIRO, 51–58.

Lu BR, Zheng KL, Qian HR, Zhuang JY. 2002. Genetic differentiation of wild relatives of rice as referred by the RFLP analysis. *Theoretical and Applied of Genetics* **106**: 101–106.

Martin C, Juliano A, Newbury HJ, Lu BR, Jackson MT, Ford-Lloyd BV. 1997. The use of RAPD markers to facilitate the identification of *Oryza* species within a germplasm collection. *Genetic Resources and Crop Evolution* **44**: 175–183.

Mochizuki K, Ohtsubo H, Hirano H, Sano Y, Ohtsubo E. 1993. Classification and relationships of rice accessions with AA genome by identification of transposable elements at nine loci. *Japanese Journal of Genetics* **68**: 205–217.

Morishima H. 2001. Evolution and domestication of rice. In: Khush GS, Brar DS, Hardy B, eds. *Rice genetics IV. Proceedings of the fourth international rice genetics symposium.* Los Banos, The Philippines: IRRI, 63–78.

Morishima H, Sano Y, Oka HI. 1992. Evolutionary studies in cultivated rice and its wild relatives. *Oxford Surveys in Evolutionary Biology* **8**: 135–184.

Naredo EB, Juliano AB, Lu BR, Jackson MT. 1998. Taxonomic status of *Oryza glumaepatula* Steud. II. Hybridization between New World diploids and AA genome species from Asia and Australia. *Genetic Resources and Crop Evolution* **45**: 205–214.

Nayar NM. 1973. Origin and cytogenetics of rice. *Advances in Genetics Incorporating Molecular Genetic Medicine* **17**: 153–292.

Oh SH, Potter D. 2003. Phylogenetic utility of the second intron of *LEAFY* in *Neillia* and *Stephanandra* (Rosaceae) and implications for the origin of *Stephanandra*. *Molecular Phylogenetics and Evolution* **29**: 203–215.

Oka HI. 1988. *Origin of cultivated rice.* Tokyo, Japan: Scientific Societies Press/Academic Press.

Oka HI, Morishima H. 1982. Phylogenetic differentiation of cultivated rice. 23. Potentiality of wild progenitors to evolve the *indica* and *japonica* types of rice cultivars. *Euphytica* **31**: 41–50.

Park KC, Kim NH, Kim NS. 2003. Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. *Theoretical and Applied of Genetics* **107**: 203–209.

Petersen G, Seberg O. 2002. Molecular evolution and phylogenetic application of *DMC1*. *Molecular Phylogenetics and Evolution* **22**: 43–50.

Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Ren FG, Lu BR, Li SQ, Huang JY, Zhu YG. 2003. A comparative study of genetic relationships among the AA-genome *Oryza* species using

RAPD and SSR markers. *Theoretical and Applied Genetics* **108**: 113–120.

Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* **37**: 121–147.

Sano Y, Morishima H, Oka HI. 1980. Intermediate perennial-annual populations of *Oryza perennis* found in Thailand and their evolutionary significance. *Botany Magazine (Tokyo)* **93**: 291–305.

Second G. 1982. Origin of the genetic diversity of cultivated rice: study of the polymorphism scored at 40 isozyme loci. *Japanese Journal of Genetics* **57**: 25–57.

Second G. 1991. Molecular markers in rice systematics and the evaluation of genetic resources. In: Bajaj YPS, ed. *Biotechnology in agriculture and forestry*. Berlin Heidelberg, Germany: Springer-Verlag, 14: 468–494.

Seelanan T, Schnabel A, Wendel JF. 1997. Congruence and consensus in the cotton tribe (Malvaceae). *Systematic Botany* **22**: 259–290.

Soltis DE, Soltis PS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis DE, Soltis PS, Doyle JJ, eds. *Molecular systematics of plants. II. DNA sequencing*. Dordrecht, the Netherlands: Kluwer, 1–42.

Soltis DE, Soltis PS. 2000. Contributions of plant molecular systematics to studies of molecular evolution. *Plant Molecular Biology* **42**: 45–75.

Swofford DL. 2001. *PAUP: phylogenetic analysis using parsimony (*and other methods), version 4.0b10.* Sunderland, MA, USA: Sinauer Associates.

Tajima F. 1993. Simple methods for testing molecular clock hypothesis. *Genetics* **135**: 599–607.

Thompson JDTJ, Gibson Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**: 4876–4882.

Vargas P, Baldwin BG, Constance L. 1998. Nuclear ribosomal DNA evidence for a western North American origin of Hawaiian and South American species of *Sanicula* (Apiaceae). *Proceedings of the National Academy of Sciences, USA* **95**: 235–240.

Vaughan DA. 1989. *The genus* Oryza *L. Current status of taxonomy*. Manila, The Philippines: IRRI.

Wall DP. 2002. Use of the nuclear gene *glyceraldehyde 3-phosphate dehydrogenase* for phylogeny reconstruction of recently diverged lineages in *Mitthyridium* (Musci: Calymperaceae). *Molecular Phylogenetics and Evolution* **25**: 10–26.

Wang ZY, Second G, Tanksley SD. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theoretical and Applied of Genetics* **83**: 565–581.

Wendel JF, Albert VA. 1992. Phylogenetics of the cotton genus (*Gossypium* L.): character-state weighted parsimony analysis of chloroplast DNA restriction site data and its systematic and biogeographic implications. *Systematic Botany* **17**: 115–143.

Wendel JF, Doyle JJ. 1998. Phylogenetic incongruence: window into genomes history and molecular evolution. In: Soltis DE, Soltis PS, Doyle JJ, eds. *Molecular systematics of plants. II. DNA sequencing*. Dordrecht, the Netherlands: Kluwer, 265–296.

Wessler SR, Bureau TE, White SE. 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Current Opinion in Genetics and Development* **5**: 814–821.

Xia X, Xie Z. 2001. DAMBE: Data analysis in molecular biology and evolution. *Journal of Heredity* **92**: 371–373.

Yamanaka S, Nakamura I, Nakai H, Sato YI. 2003. Dual origin of the cultivated rice based on molecular markers of newly collected annual and perennial accessions of wild rice species, *Oryza nivara* and *O. rufipogon*. *Genetic Resources and Crop Evolution* **50**: 529–538.

About *New Phytologist*

- *New Phytologist* is owned by a non-profit-making **charitable trust** dedicated to the promotion of plant science, facilitating projects from symposia to open access for our Tansley reviews. Complete information is available at **www.newphytologist.org**.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as-ready' via *OnlineEarly* – the 2004 average submission to decision time was just 30 days. Online-only colour is **free**, and essential print colour costs will be met if necessary. We also provide 25 offprints as well as a PDF for each article.
- For online summaries and ToC alerts, go to the website and click on 'Journal online'. You can take out a **personal subscription** to the journal for a fraction of the institutional price. Rates start at £109 in Europe/$202 in the USA & Canada for the online edition (click on 'Subscribe' at the website).
- If you have any questions, do get in touch with Central Office (**newphytol@lancaster.ac.uk**; tel +44 1524 592918) or, for a local contact in North America, the US Office (**newphytol@ornl.gov**; tel +1 865 576 5261).