Theor Appl Genet (2010) 121:769–787
DOI 10.1007/s00122-010-1348-z

ORIGINAL PAPER

Structure, allelic diversity and selection of *Asr* genes, candidate for drought tolerance, in *Oryza sativa* L. and wild relatives

Romain Philippe · Brigitte Courtois · Kenneth L. McNally · Pierre Mournet · Redouane El-Malki · Marie Christine Le Paslier · Denis Fabre · Claire Billot · Dominique Brunel · Jean-Christophe Glaszmann · Dominique This

Received: 29 January 2010 / Accepted: 19 April 2010 / Published online: 8 May 2010

Abstract *Asr* (ABA, stress, ripening) genes represent a small gene family potentially involved in drought tolerance in several plant species. To analyze their interest for rice breeding for water-limited environments, this gene family was characterized further. Genomic organization of the gene family reveals six members located on four different chromosomes and with the same exon–intron structure. The maintenance of six members of the *Asr* gene family, which are the result of combination between tandem duplication and whole genome duplication, and their differential regulation under water stress, involves probably some sub-functionalization. The polymorphism of four members was studied in a worldwide collection of 204 accessions of *Oryza sativa* L. and 14 accessions of wild relatives (*O. rufipogon* and *O. nivara*). The nucleotide diversity of the *Asr* genes was globally low, but contrasted for the different genes, leading

Communicated by M. Wissuwa.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1348-z) contains supplementary material, which is available to authorized users.

R. Philippe · B. Courtois · P. Mournet · D. Fabre · C. Billot · J.-C. Glaszmann · D. This
CIRAD, UMR Développement et Amélioration des Plantes, TA-A 96/03, 34398 Montpellier, France

B. Courtois
e-mail: brigitte.courtois@cirad.fr

P. Mournet
e-mail: pierre.mournet@cirad.fr

D. Fabre
e-mail: denis.fabre@cirad.fr

C. Billot
e-mail: claire.billot@cirad.fr

J.-C. Glaszmann
e-mail: jean-christophe.glaszmann@cirad.fr

D. This
e-mail: dominique.this@supagro.inra.fr

K. L. McNally
International Rice Research Institute, DAPO Box 47777, 1301 Metro Manila, Philippines
e-mail: K.McNally@CGIAR.ORG

P. Mournet · R. El-Malki · M. C. Le Paslier · D. Brunel
INRA, UR 1279 Etude du Polymorphisme des Génomes Végétaux, CEA Institut de Génomique, Centre National de Génotypage, 2, rue Gaston Crémieux, CP5724, 91057 Evry, France

R. El-Malki
e-mail: r_elmalki@hotmail.com

M. C. Le Paslier
e-mail: lepaslier@cng.fr

D. Brunel
e-mail: brunel@versailles.inra.fr

D. This
Montpellier SupAgro, UMR Développement et Amélioration des Plantes, 2, place P. Viala, 34060 Montpellier, France

R. Philippe (✉)
Institut National de Recherche Agronomique, UMR INRA-UBP 1095, Génétique, Diversité et Ecophysiologie des Céréales, 234, avenue de Brézet, 63100 Clermont-Ferrand, France
e-mail: rphilippe@clermont.inra.fr

to different shapes of haplotype networks. Statistical tests for neutrality were used and compared to their distribution in a set of 111 reference genes spread across the genome, derived from another published study. *Asr3* diversity exhibited a pattern concordant with a balancing selection at the species level and with a directional selection in the tropical japonica sub-group. This study provides a thorough description of the organization of the *Asr* family, and the nucleotide and haplotype diversity of four *Asr* in *Oryza sativa* species. *Asr3* stood out as the best potential candidate. The polymorphism detected here represents a first step towards an association study between genetic polymorphisms of this gene family and variation in drought tolerance traits.

Introduction

Drought is a major agronomic problem, resulting in reduction in yields of crops exposed to chronic or sporadic periods of water deficit. Rice is the most important food crop in the world with a total production of 652 Mt in 2007 (FAOSTAT 2008) and feeds over half of the world population. However, about 15% of Asian rice area (23 million ha) experiences frequent yield losses due to drought (Widawsky and O'Toole 1990). The problem is particularly severe in Eastern India, with more than 10 million ha of drought-prone fields, where annual losses due to drought average US$250 million (Pandey et al. 2000). Owing to global warming, crops, and notably rice, will be more and more exposed to chronic or sporadic periods of drought. Therefore, an important objective of current research is to enhance drought tolerance in rice. One way may be to identify genes controlling performance under water-limited conditions and favourable alleles at these genes.

Plant response to stress conditions occurs through a number of changes at physiological and developmental levels, brought about by altering the expression of stress-inducible genes. ABA (abscisic acid) is produced under environmental stress and is important for the tolerance of plants to drought, salinity and cold (Bray 1993; Skriver and Mundy 1990). ABA leads to the accumulation of various gene products during periods of water deficit. One of the genes induced by ABA was characterized in tomato and was called *Asr* (abscisic acid, stress and ripening) (Iusem et al. 1993). Since then, four *Asr* genes have been identified in tomato (Frankel et al. 2006). In parallel, *Asr* homologs were cloned from a wide range of other plant species including potato (Schneider et al. 1997), maize (Meyer 1995), pomelo (Canel et al. 1995), loblolly pine (Chang et al. 1996), lily (Wang et al. 2005), rice (Vaidyanathan et al. 1999) and grape (Cakir et al. 2003). Surprisingly, these genes are not present in *Arabidopsis*. In most species, *Asr* genes belong to a small gene family characterized by the presence of an ABA/WDS domain (Pfam 02496). The number of plant *Asr* available in plant genome databases varies from one in grape (http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/), four in *Brachypodium distachyon* (http://www.brachypodium.org/), to up to seven in *Sorghum bicolor* (http://phyto3.phytozome.net/). Not much is known about the function of ASR proteins. In most studies, only one member of the family was studied. Because orthology was unclear (Frankel et al. 2006), attributing a function to each member of the family was difficult. Targeted experiments have revealed that members of the *Asr* gene family are induced by abscisic acid (ABA), various abiotic stresses including water stress and during the process of fruit ripening (Carrari et al. 2004). Therefore, the corresponding proteins were classified as hydrophilins, a wide protein group involved in adaptation to water deficits (Battaglia et al. 2008). Yeast one-hybrid experiments revealed that grape ASR protein bind to the promoter of a hexose transporter gene (Cakir et al. 2003). The above data were in close agreement with the proposed role of ASR product as a transcription factor involved in sugar metabolism (Carrari et al. 2004; Kalifa et al. 2004). Furthermore, evidences suggested a dual role for an ASR from *Lilium longiflorum* (LLA23) uncovering an independent function as a protective molecule against water loss (Yang et al. 2005). Tomato ASR1 proteins were localized to both the cytosol, as unstructured monomers, and to the nucleus, as structured DNA-bound homodimers (Kalifa et al. 2004; Konrad and Bar-Zvi 2008). The unstructured form of tomato ASR1 proteins was shown to present a chaperone-like activity and could stabilize a number of proteins against denaturation caused by heat and freeze–thaw cycles.

Several genetic experiments have implicated *Asr* in the response to water deficit. An increase in foliar senescence under drought conditions was shown in transgenic *Asr1* over-expressing lines in maize, whereas antisense lines showed the opposite effect (Jeanneau et al. 2002). Moreover, *Asr1* colocalized with a QTL for xylem sap ABA content, a QTL for leaf senescence and a QTL for anthesis–silking interval responsive to mild water deprivation in maize (de Vienne et al. 1999). Accelerated rates of amino acid substitutions were demonstrated in the water stress-induced *Asr2* gene of tomato species inhabiting dry environments, suggesting some positive selection during adaptation to arid conditions (Frankel et al. 2003). Very recently, the study of the extent of nucleotide diversity in *Asr2* in two *Solanaceae* species provided some evidence of non-neutral evolution, potentially associated with unique climatic features or demographic events (Giombini et al. 2009).

To our knowledge, no prior diversity study has involved several members of *Asr* gene family and their co-evolution

in rice. Multigenic families are very common in plant genomes. The proportion of genes belonging to a gene family was estimated to be 53% in rice and 68% in *Arabidopsis* (Lin et al. 2008). Three mechanisms can occur to induce extension of gene families: whole genome duplication (WGD), segmental duplication and tandem duplication. In rice, 18 duplicated segments covering 65.7% of the genome was identified (Yu et al. 2005). The authors concluded that a WGD occurred before the origin of grasses, 53 million years ago, for all but one of these 18 segment pairs, followed by a segmental duplication 21 million years ago affecting a segment on chromosomes 11 and 12. They also showed massive ongoing individual gene duplication. Some authors suggested that duplicate gene evolution is a balance between maintaining at least one functional copy of a critical gene and diversification of the different copies for evolutionary innovation (Chapman et al. 2006). Therefore, gene diversity within a gene family, and in our case the *Asr* gene family, may be a source of adaptation to fluctuating and/or stressfull environments.

Rice is a good model for studying the evolution of the *Asr* gene family because of its huge genetic and ecological diversification. A wide range of genetic variation exists within the main cultivated species, *Oryza sativa*. The first classification was based on the agro-morphological traits, as well as intervarietal hybrid fertility, and separated two major groups, indica and japonica types (Kato et al. 1928). Based on isozyme markers, six groups were later identified that can be related in a general sense to previous classification systems (Glaszmann 1987). Group I correspond to indica types (mainly lowland varieties), group II to aus (short-duration varieties from South Asia) and boro types, group III to aswina types (deepwater varieties), group IV to rayada types (floating varieties), group V to sadri and basmati types (aromatic varieties)and group VI to japonica types. The indica and japonica groups are numerically the most important and are close to the aus/boro and sadri/basmati groups, respectively, which are of intermediate importance. The aswina and rayada groups represent only very few accessions. Recently, DNA markers have confirmed the genetic structure of four major groups (indica, aus/boro, sadri/basmati and japonica) and subdivided the japonica group into temperate and tropical components (Caicedo et al. 2007; Garris et al. 2005). The indica and japonica groups are probably the result of two independent domestications, in eastern India and southern China, from local ancestral populations of *Oryza rufipogon* while the origin of the secondary groups is still under debate (Kovach et al. 2007; Londo et al. 2006).

Rice is also a good model for studying the evolution of the *Asr* gene family because it is a very plastic crop which is grown in a broad range of ecosystems with differing hydrology, from irrigated to fully rainfed, where the root systems encounter anaerobic to aerobic conditions, respectively. Rice varieties present some degree of specialization to an ecosystem. However, this specialization is not an absolute in the sense that a change of ecosystem is not lethal to the plants but reduces the varietal performance. A relationship between ecosystems as defined by hydrology (anaerobic or aerobic) and genetic organization is observed. Traditionally, indica, temperate japonica, boro, and sadri/basmati are mostly grown under anaerobic conditions during nearly the whole cycle of the crop (irrigated, or rainfed lowland ecosystems). Aus and tropical japonica are mostly grown under aerobic conditions in the rainfed upland ecosystems. Nevertheless, exceptions are encountered, notably among improved varieties. Comparison between sets of accessions better adapted to aerobic or anaerobic hydrological situations seems to be a good model to analyze adaptation to water deficit conditions. One can hypothesize that within a genetic group that includes sub-groups with different adaptations (upland aus versus irrigated boro for the aus/boro group, and upland tropical versus irrigated temperate for the japonica group), the opposition among the sub-groups can help identify specific alleles of adaptation.

Nucleotide diversity studied on a reference collection of rice should allow using population genetic tests of departure from the neutral equilibrium model to identify the diverse selective modes that shaped the evolution of the *Asr* gene family during differential adaptation of the varieties to aerobic and anaerobic conditions and such a study conducted in *Arabidopsis thaliana* and *Arabidopsis halleri* showed different selective modes between paralogs and orthologs in the chalcone synthase gene family due to local adaptation (Wang et al. 2007). Reviews of neutrality tests highlight the need for several tests and the necessity to consider population structure and domestication history (Walsh 2008; Wright and Gaut 2005). These tools were used in several published works to look more specifically at the molecular basis of plant adaptation (reviewed in Ehrenreich and Purugganan 2006).

In this work, we characterized the *Asr* gene family in rice, including its expression pattern, and examined sequence intraspecific polymorphism and interspecific divergence in the *Asr* genes. The study addresses questions about the molecular evolution of orthologs between a cultivated species and its wild relative, and that of paralogs within *O. sativa.* More specifically, we try to determine whether the divergence of *Asr* paralogs is governed by different evolutionary forces and whether the patterns of nucleotide variation of *Asr* genes are determined by local adaptation to hydrological environments.

Methods

Plant material

The GCP (Generation Challenge Program) rice collection that represents 3,000 *Oryza* spp. accessions had been genotyped with 48 SSR (simple sequence repeat) markers (data available at http://gcpcr.grinfo.net/). Based on these data, 265 accessions of *O. sativa* which represented the diversity of the collection were chosen. SSR data on this collection were used to assign individuals to sub-populations, using a model-based clustering method with the software STRUCTURE (Pritchard et al. 2000). Based on the membership percentage above 60 in the four detected sub-populations, 113 accessions were assigned to the indica group, 30 to the aus/boro group, 31 to the sadri/basmati group and 91 to the japonica group. Passport data on culture conditions allowed splitting the japonica group into temperate japonica and tropical japonica and the aus/boro group into aus and boro. CIRAD breeders were consulted to validate this classification.

Four accessions of *Oryza nivara* (annual form) and 16 accessions of *Oryza rufipogon* (perennial form) were added. Some researchers distinguish between these two forms due to their variation in life history habit (Li et al. 2006; Yamanaka et al. 2003). Others consider them to be ecotypes of a single species because they exhibit continuous variation in nature for their life history habit, coupled with the fact that there is no major reproductive barrier or significant genetic differentiation between them (Lu et al. 2002; Oka 1988), as demonstrated by the abundance of shared sequence polymorphisms, the lack of fixed differences, and very low F_{ST} values (Lu et al. 2002; Zhu and Ge 2005; Zhu et al. 2007). As such, we will treat these two species as a single gene pool referred to as wild rice.

Sequence search, domain detection and protein alignment

A reference *Asr* cDNA (acc. no. AF039573) cloned in rice (Vaidyanathan et al. 1999) and its predicted protein was used to do BLAST searches (BlastN, TBlastN and BlastP) in the OryGenes database (http://orygenesdb.cirad.fr/). Sequences with an *e* value equal to or lower than 10^{-4} were retained. InterProScan program (Quevillon et al. 2005) was used to characterize the specific domain of the *Asr* gene family (Pf02496). Six *Asr* genes were identified and visualized on the rice pseudo-molecules using the genome browser of the OryGenes database (Droc et al. 2006). The multiple sequence alignment of the ASR family proteins was performed using the MUSCLE software (Edgar 2004).

DNA extraction, primer design, DNA amplification and sequencing

DNA extraction was performed using the CTAB method. Primer pairs for *Asr1*, *Asr2*, *Asr3*, *Asr5* and *Asr6* were designed from the 5′UTR and the 3′UTR of *Oryza sativa* L. var. Nipponbare sequences extracted from OryGenesDB using Oligo v6.71 (Rychlik and Rhoads 1989). Primer sequences are presented in Online Resource 1.

DNA amplification and sequencing were performed by the INRA EPGV group in CNG (CEA Institut de Génomique/Centre National de Génotypage, Evry, France) in 384-well microplates. After PCR with a high fidelity enzyme (Platinium Taq HiFi by Invitrogen), the products were purified with Exo I enzyme (New England Biolabs) and SAP enzyme (USB Corporation), then sequenced in both directions using the BigDye v3.1 sequencing kit (Applied Biosystems). Sequencing reactions products were purified by ethanol precipitation and loaded onto a 3730*xl* 96-capillary sequencer (Applied Biosystems). SNP data from this article have been deposited with the NCBI/DbSNP Data Libraries are presented in Online Resource 2.

Patterns of nucleotide diversity

Sequence quality control, alignment and nucleotide polymorphism detection were performed using Codon Code Aligner v1.6.3 (http://www.codoncode.com/index.htm) with minimum Phred scores set to 20. This criterion of quality sequence and the primers positioning allowed us to align complete gene sequences from 5′untranslated region (UTR) to 3′UTR. Each polymorphic site was manually checked. No heterozygote was detected in domesticated rice, but a few were detected in wild rice. In each heterozygote accession, only one site was heterozygous and that particular situation made cloning of PCR products, prior to sequencing, unnecessary because there were only two possible haplotypes, which have thus been considered. Basic parameters including the number of single nucleotide polymorphisms (SNPs), synonymous (S) and nonsynonymous (NS) substitutions and minimum recombination events (Rm) were calculated using DnaSP v4.52 (Rozas and Rozas 1999). Levels of genetic diversity within domesticated rice, wild rice and each varietal group were quantified with measures of nucleotide diversity, θ_W based on the number of segregating sites (Watterson 1975) and π based on the average number of nucleotide differences per site between sequences (Nei 1987). These parameters were computed with DnaSP, on sequences without insertions/deletions events (INDELs), at three different levels: (1) entire sequence, (2) silent sites (including synonymous sites and noncoding regions), (3) nonsynonymous sites. To take into

consideration the heterozygosity of some wild accessions, each haplotype was considered as a unique sequence.

The number of haplotypes, the number of singletons and the haplotype diversity (Hd) were calculated with DnaSP v4.52. Haplotype networks, representing unique alleles separated by mutational steps, were constructed using the program NETWORK v4.5 (http://www.fluxus-technology.com/sharenet.htm) using the median-joining method (Bandelt et al. 1999), with an equal weight for all sites and an ε-parameter of 0 for all *Asr* out of *Asr2* ($\varepsilon = 10$). Both SNP and INDELs were considered.

Tests for selection were performed to estimate whether the considered genes followed the model of neutral evolution or not. Tajima's *D* (Tajima 1989), Fu and Li's *D** (Fu and Li 1993) and HKA (Hudson et al. 1987) tests were carried out for the whole sequences using DnaSP.

We compared our diversity indices and the results of selection tests with a reference at the genome level (Caicedo et al. 2007). The data-set was composed of 111 randomly chosen STS (sequence-tag sites), sequenced in a population of 73 accessions of *Oryza sativa*. Over half of these accessions were common with our set. We calculated π, θ_W and Tajima's *D* for each STS with DnaSP v4.52 and drew the reference distributions of these parameters in the rice genome.

To analyze the relationships between genes, linkage disequilibrium (LD) between gene haplotypes was estimated using the r^2 coefficient computed by the program TASSEL v2.01 (Bradbury et al. 2007). Only haplotypes with a frequency over 0.05 were used. LD was computed first for the common set of domesticated rice accessions, then separately for the indica, the aus/boro and the tropical japonica components of the set; the temperate japonica and the sadri/basmati groups were not submitted to the same computation because of their monomorphism.

Expression analysis

Six accessions from the collection were used to study expression of the *Asr* gene family: IR64 and Apo (indica), Azucena (tropical japonica), M202 (temperate japonica), DomZard (sadri) and FR13A (boro). After germination, seedlings were transplanted in 1 l pots and grown in a greenhouse under controlled environment (21°C night time, 28°C daytime, relative humidity at 80%, and a minimum luminosity of 600 µmol m² s over a 14 h-day length). For each accession, three biological replications were tested under two conditions: well-watered control plants and water-stressed plants with a mean FTSW (fraction of transpirable soil water) of 0.4. Total mRNA was extracted from mature leaves (carbon source) and young non-photosynthetic leaves (carbon sink) separately. Water stress, mRNA extraction and RT–PCR were implemented as described in a previous work (Luquet et al. 2008).

Primers for qRT–PCR were designed using VectorNTI (http://invitrogen.com) and Oligo v6.71, and synthesized by Operon Biotechnologies, GmbH (Cologne, Germany). The control gene model was *actin 1* (Os05g36290), based on published data in rice showing stable expression of this gene in leaves (Caldana et al. 2007; Kim et al. 2003). Primer sequences are shown in the Online Resource 3.

PCR reactions were conducted in a LightCycler-480 instrument (Roche Molecular Biochemicals). The expression ratio between stressed and non-stressed conditions was assessed by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001), using the mean of ΔCt for the control plants.

Results

Structure and genomic distribution of the Asr gene family

A genomic scale search allowed the identification of six *Asr* genes in Nipponbare, the *japonica* rice with a sequenced genotype. The predicted gene structures in Nipponbare are presented in Fig. 1. The *Asr* gene structure consists of two small exons (between 135 and 492 bp in Nipponbare) separated by one intron (between 84 and 440 bp in Nipponbare). ESTs for all of these genes were available in TIGR rice gene index database (http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=rice) and confirmed the gene

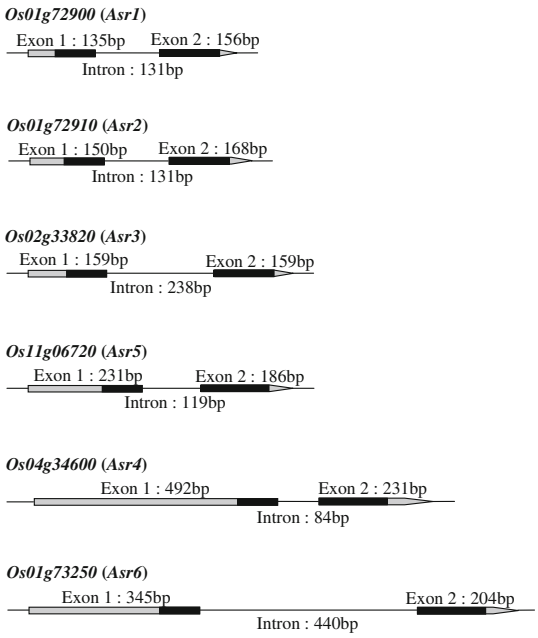

Fig. 1 Exon–intron structure of the *Asr* genes family in Nipponbare. Coding and non-coding regions are indicated by *boxes* and *black lines*, respectively. The well-conserved domain ABA/WDS is *highlighted* in *black* in the *boxes*

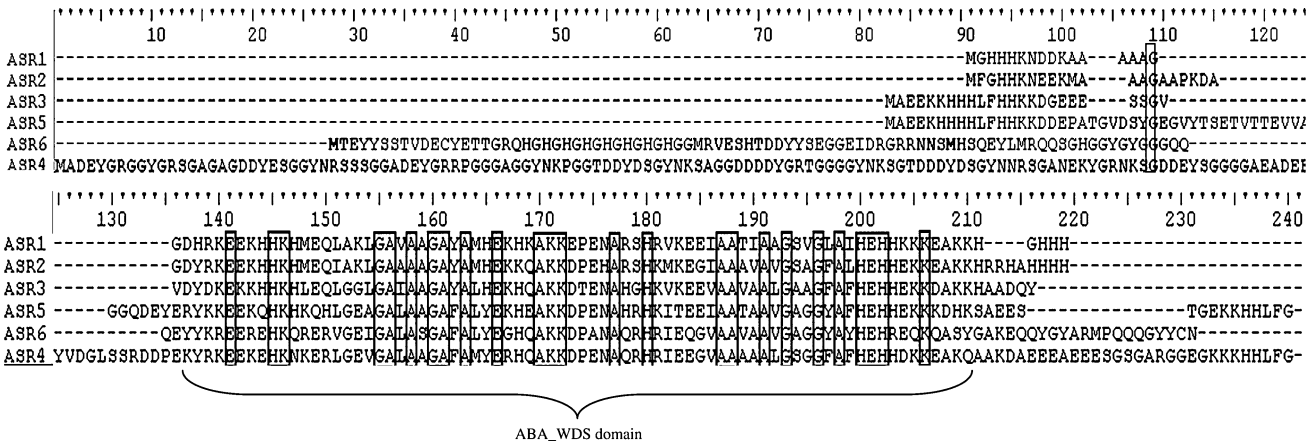

Fig. 2 Alignment of the *Oryza sativa* ASR protein family. Alignment was computed with MUSCLE software (Edgar 2004). *Boxes* show completely conserved sites

Table 1 Protein identity between the six members of the ASR family

	ASR1	ASR2	ASR3	ASR4	ASR5	ASR6
ASR1	–	73.7%	58.9%	48.4%	52.6%	34.3%
ASR2		–	60.0%	51.6%	50.5%	37.4%
ASR3			–	52.6%	64.2%	42.4%
ASR4				–	55.8%	51.5%
ASR5					–	58.6%
ASR6						–

Identity was calculated on the common region defined by a MUSCLE alignment (amino acids 91–102, 107–109 and 136–211)

expression and the predicted structure. The predicted proteins varied between 96 and 229 amino acids (aa), and included a family-specific domain ABA_WDS of 80 aa (Pfam family domain Pf02496), with 25 aa totally conserved between all members of the gene family (Fig. 2). The proteins could be divided into two classes based on their length: the short ASR proteins (ASR1, ASR2, ASR3 and ASR5) between 96 and 138 aa and the long ones (ASR4 and ASR6) between 182 and 229 aa. The ASR4 protein revealed an unshared motif of 144 aa that included four imperfect 35 aa repeats. The ASR6 protein contained an unshared motif of 86 aa, very different from the ASR4 motif, with eight repeats of two histidine and glycine. ASR proteins carry five distinctive motifs (Battaglia et al. 2008). Three of them compose the ABA/WDS domain and diverge between ASR family members. ASR3 and ASR5 shared a fourth motif while ASR4 and ASR5 had potentially a fifth one. The N-terminal regions of ASR4 and ASR5 contained some structural similarity to a consensus domain involved in Zn DNA-binding activity (Battaglia et al. 2008). Table 1 shows the pairwise percentage identity of the rice ASR proteins, based on the alignment without INDEL. ASR1 and ASR2 were the most similar proteins with 73.7% aa identity. ASR3 was closest to ASR5 with 64.2% aa identity. They both contained a C-terminal domain with a long histidine stretch. The two long ASR showed more similarity, with 58.6% aa identity, than with the short ASR (between 34.3 and 55.8% aa identity) in the common region.

The six rice *Asr* family members are located on chromosomes 1 (*Asr1*, *Asr2* and *Asr6*), 2 (*Asr3*), 4 (*Asr4*) and 11 (*Asr5*) (see Online Resource 4). *Asr1* and *Asr2* were separated by only 850 bp whereas *Asr6* was separated from this cluster by approximately 177 kb (see Online Resource 4). To reconcile this organization with a pattern of duplication, we looked for colinearity around the *Asr* loci. Ten coding sequences directly surrounding rice *Asr* genes (5 on each side) were extracted from OrygenesDB, translated to proteins and concatenated. Only two hits were found by pairwise BlastP: the first one between two expressed uncharacterized sequences (Os02g33830 and Os04g34620 with 82% identity and *e* value = 2e-170) near *Asr3* and *Asr4*, and the second one between two syntaxins (Os01g73230 and Os11g06740 with 69% identity and *e* value = 2e-79) near *Asr5* and *Asr6* (see Online Resource 5). This suggests that the segments around the two couples of *Asr* are the direct result of duplication events.

Expression and potential role of the Asr gene family during water deficit

Not much is known on the respective function of the members of the *Asr* gene family in rice. We obtained preliminary information on the expression and role of each member during water stress by assessing in parallel the co-localization of the different loci with known QTLs controlling drought resistance traits, the occurrence of corresponding sequences in EST databases and the differential expression of the genes between stressed and control conditions.

Table 2 Number of EST of the *Asr* genes available in the TIGR rice gene index database

EST was assigned in different classes according to different types of EST libraries, as assessed on December 2009 (http://compbio.dfci.harvard.edu/tgi/)

Gene	Leaves	Roots	Panicles	Others	Drought-stressed leaves	Drought-stressed roots	Total
Asr1	2	0	0	1	3	0	7
Asr2	10	0	0	9	0	0	19
Asr3	38	13	10	35	0	1	97
Asr4	122	4	11	37	5	0	179
Asr5	434	96	100	481	4	0	1,115
Asr6	48	3	1	56	3	0	111

Table 3 Differential expression of the *Asr* observed in source and sink leaves of six rice accessions

	Source leaves						Sink leaves					
	Asr1	*Asr2*	*Asr3*	*Asr4*	*Asr5*	*Asr6*	*Asr1*	*Asr2*	*Asr3*	*Asr4*	*Asr5*	*Asr6*
IR64 (indica)	(⇑) 2.6	⇑ 5.0	⇓ 10.0	(⇓) 2.0	⇔	(⇑) 2.1	⇑ 2.3	∅	(⇑) 2.3	⇓ 4.0	⇓ 7.1	∅
APO (indica)	⇑ 2.1	⇑ 4.9	⇓ 5.2	⇓ 1.4	⇔	(⇑) 4.4	⇑ 3.7	⇑ 7.6	(⇑) 4.0	⇓ 5.0	(⇓) 2.9	⇑ 3.4
FR13A (aus/boro)	⇑ 2.0	⇔	⇓ 7.7	⇑ 3.4	⇑ 3.3	⇓ 4.6	⇑ 2.1	⇔	⇑ 4.8	⇓ 2.3	⇓ 2.6	⇑ 3.0
Dom-Zard (basmati)	⇑ 1.7	⇔	⇓ 2.9	⇔	(⇓) 2.3	⇔	(⇑) 2.2	∅	⇔	⇓ 5.9	⇓ 2.8	⇔
Azucena (trop. Jap)	⇑ 2.2	⇑ 3.2	⇓ 4.8	⇓ 1.9	(⇓) 2.6	⇔	⇑ 2.4	∅	⇔	⇔	⇓ 2.2	∅
M202 (temp. jap)	⇑ 1.9	⇑ 4.6	⇑ 3.2	⇔	⇓ 3.0	⇑ 4.1	⇑ 2.1	∅	⇑ 1.8	⇔	⇓ 8.3	∅

The differentials were calculated between gene expressions at 0.4 FTSW and gene expressions of a well watered control

⇑: upregulated with P value < 0.05; ⇓: downregulated with P value < 0.05; () : 0.05 < P < 0.1; ⇔: no significant effect; ∅: no expression in control and stressed plants

Numbers below arrows indicate the ratio of up or down regulation of each gene at 0.4 FTSW compared to well watered control. Data are averaged on three biological replicates and based on $2^{-\Delta\Delta Ct}$ methodology, actin being used as control

We used a drought QTL compilation available in the rice module of TropgeneDB (http://tropgenedb.cirad.fr/html/rice_QTL.html) to compare the position of known QTLs with the *Asr* loci (see Online Resource 4). QTLs for deep root number, leaf rolling and leaf drying scores were located in the vicinity of the *Asr1-Asr2-Asr6* chromosomal region. A QTL for root thickness contained *Asr5* in its confidence interval; a QTL for shoot biomass after a water stress period was centered on the *Asr3* locus, and a cluster of QTLs for panicle number, and biomass under water stress and osmotic adjustment colocalized with *Asr4*.

The TIGR rice gene index database included a large number of ESTs derived from several rice cDNA libraries, presenting at least 98% nucleotide sequence identity with the different *Asr* genes (Table 2). Only seven ESTs shared homology with *Asr1*, three of them being derived from leaves submitted to drought stress. Nineteen ESTs corresponding to *Asr2* were identified mainly in leaf tissues, but not specifically under water deficit. A larger number of ESTs was identified for *Asr3*, *Asr4*, *Asr5* and *Asr6* (97, 179, 1115 and 111, respectively) in different conditions, including water deficit conditions (1, 5, 4 and 3, respectively). All genes were expressed in leaves, and all, except *Asr1* and *Asr2,* in roots.

Asr gene expression in mature and young leaves was assessed by Q-RT-PCR in stressed and well-watered plants of six rice accessions from indica, japonica, basmati and boro varietal groups (Table 3) with three biological replicates per accession, and with *actin* being used as the internal reference gene. No significant difference was shown between the two treatments on actin *C*t averaged for each tissue (Student's test $\alpha = 0.05$). Ratio of expression between stressed and well-watered plants showed a general increase in expression of *Asr1*, *Asr2* and *Asr6* in mature (source) leaves, but with some varietal differences, whereas the expression of *Asr3*, *Asr4* and *Asr5* in the same tissue showed a general decrease in expression. In young non-photosynthetic (sink) leaves, *Asr1* and *Asr3* expression generally increased, while *Asr4* and *Asr5* expression generally decreased. No gene expression was detected in sink leaves under well watered or stress conditions for *Asr2* in four out of six accessions and for *Asr6* in three out of six accessions. This experiment suggested a regulation of *Asr* genes under drought stress in leaves, highly depending on the developmental stage or autotrophic status of the leaf material, and some genetic variation in the extent, or even direction in a few cases, of gene regulation under artificial water deficit.

Table 4 Estimates of nucleotide diversity at four *Asr* genes in *O. sativa* and wild relatives

Gene	Species	*n*	Total sites						Silent sites				Non-synonymous Sites			
			L	*S*	Indel	*H*	π	θ	*L*	*S*	π	θ	*L*	*S*	π	θ
Asr1	*O. sativa*	204	513	5	0	5	1.40	1.65	283.7	4	2.30	2.39	229.3	1	0.29	0.74
	Wild	14	483	10	2	9	3.47	5.32	256.3	7	4.74	7.02	226.7	3	2.04	3.40
Asr2	*O. sativa*	204	580	6	2	9	1.04	1.76	328.6	4	1.62	2.07	251.4	2	0.30	1.35
	Wild	14	575	18	3	11	7.67	8.04	323.4	14	11.05	11.12	251.6	4	3.32	4.09
Asr3	*O. sativa*	204	742	22	11	11	9.02	5.03	494.0	22	13.55	7.56	248.0	0	0.00	0.00
	Wild	14	745	28	8	16	13.20	9.66	494.6	27	19.37	14.04	250.4	1	1.01	1.03
Asr5	*O. sativa*	204	582	4	0	5	0.89	1.17	256.3	2	1.94	1.32	325.7	2	0.06	1.04
	Wild	14	581	4	0	3	0.83	1.77	255.3	4	1.90	4.03	325.7	0	0.00	0.00

n number of accessions compared, *L* length of the sequence without indel, *S* number of polymorphic sites, *indel* number of insertion/deletion, *H* number of haplotypes including indel, π (10^{-3}) Tajima's estimate per kb, θ (10^{-3}) Watterson's estimate per kb

Polymorphism of the short Asr genes in *O. sativa* and wild rice

A core set of 204 accessions of *O. sativa*, three accessions of *O. nivara* and 11 accessions of *O. rufipogon* had sequences of quality meeting the chosen criterion (Phred score of 20) with no missing data for *Asr1*, *Asr2*, *Asr3* and *Asr5*, the four short *Asr*. This set was used for further data analyses (see Online Resource 6). The rates of good quality sequences were very low for *Asr4* and intermediate for *Asr6* and they were further eliminated from the analyses.

The nucleotide diversity of the *Asr* genes in *Oryza sativa* was globally low but contrasted for the different genes (Table 4). *Asr5* presented the lowest diversity with two NS and two S polymorphisms ($\pi = 0.89 \times 10^{-3}$ and $\theta_W = 1.17 \times 10^{-3}$ in total sites). *Asr1* and *Asr2* had an intermediate diversity with one and two NS mutations, respectively, and four S mutations each ($\pi = 1.40 \times 10^{-3}$ and 1.04×10^{-3} and $\theta_W = 1.65 \times 10^{-3}$ and 1.76×10^{-3} in total sites, respectively). *Asr3* showed a much higher diversity than the other *Asr* with 22 S polymorphisms but without NS polymorphism ($\pi = 9.01 \times 10^{-3}$ and $\theta_W = 5.01 \times 10^{-3}$ in total sites). The observed difference between π and θ_W in *Asr3* was due to the absence of low frequency SNP alleles, with all SNP alleles having a frequency higher than 7% and nearly half having a frequency higher than 44%. An additional source of diversity in *Asr3* came from the presence in the intron of a MITE element belonging to the stowaway family. This transposable element occurred in nearly half of the collection. The MITE distribution varied across population types: most of the indica accessions carried the MITE while the japonica and sadri/basmati accessions did not. Half of the aus/boro accessions possessed it. All accessions, which did not possess the MITE, contained a 6 bp sequence that might correspond to a trace of an ancient occurrence of the MITE element.

We compared the values observed for π and θ_W for all sites in the *Asr* genes to the π and θ_W distributions of a reference set (Fig. 3) (Caicedo et al. 2007). *Asr1*, *Asr2* and *Asr5* presented a level of diversity corresponding to the peak of the distribution for both indices. For *Asr3*, in contrast, π was higher than the mean value (2.29×10^{-3}) and outside the 95% threshold of the distribution and θ_W was also high and close to the 90% threshold. The same analysis run only for silent sites gave similar results (data not shown).

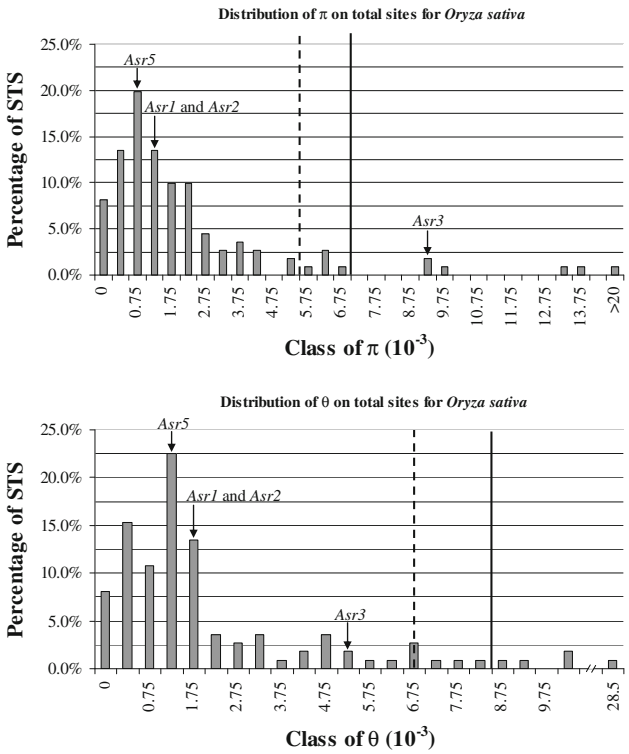

Fig. 3 Comparison of *Asr* genes to *Oryza sativa* general nucleotide diversity. π and θ were computed for 111 STS (sequence-tag sites). The *left part* of the *bold line* and the *dashed line* represent 95 and 90% of the distribution, respectively. *Asr1*, *Asr2*, *Asr3* and *Asr5* π and θ are indicated by *arrows* on the corresponding classes

Table 5 Test of departure from neutrality computed for four *Asr*

Gene	Group	*n*	*S*	*D*	*D**	K_a/K_s
Asr1	*Oryza sativa*	204	5	−0.294 n.s.	0.979 n.s.	0.256
	Wild rice	14	10	−1.117 n.s.	0.879 n.s.	0.193
	indica	80	5	0.413 n.s.	0.023 n.s.	0.256
	aus/boro	27	3	−0.508 n.s.	0.963 n.s.	
	sadri/basmati	25	0			
	japonica	71	4	−1.280 n.s.	−0.186 n.s.	
Asr2	*Oryza sativa*	204	6	−0.820 n.s.	0.024 n.s.	0.253
	Wild rice	14	18	−0.847 n.s.	−0.537 n.s.	0.252
	indica	80	4	0.497 n.s.	0.960 n.s.	
	aus/boro	27	1	−0.338 n.s.	0.607 n.s.	
	sadri/basmati	25	1	−1.158 n.s.	−1.620 n.s.	
	japonica	71	4	−1.218 n.s.	−2.499*	
Asr3	*Oryza sativa*	204	22	2.150*	1.807**	–
	Wild rice	14	28	1.333 n.s.	1.702**	0.086
	indica	80	13	−0.005 n.s.	1.513*	
	aus/boro	27	19	1.779 n.s.	1.603**	
	sadri/basmati	25	13	−1.447 n.s.	1.490*	
	japonica	71	13	−0.365 n.s.	1.514*	
Asr5	*Oryza sativa*	204	4	−0.424 n.s.	−3.000*	0.543
	Wild rice	14	4	−1.384 n.s.	0.066 n.s.	–
	indica	80	2	−0.159 n.s.	−1.017 n.s.	
	aus/boro	27	2	0.277 n.s.	−0.702 n.s.	
	sadri/basmati	25	0			
	japonica	71	2	−1.006 n.s.	−0.987 n.s.	

The different tests were computed in the *O. sativa* and wild rice (in bold) and in different *O. sativa* subgroups

n number of accessions compared, *S* number of polymorphic sites, *D* Tajima's *D* test, *D** Fu and Li's *D** test, K_a number of substitution at non-synonymous site per non-synonymous sites, K_s number of substitution at synonymous sites per synonymous sites, *n.s.* not significant

* $P < 0.05$; ** $P < 0.01$ for *D* and *D** Tajima's tests and $P < 0.02$ for Fu and Li's tests

In wild rice, the diversity range among *Asr* was parallel to that observed for *O. sativa*, with *Asr5* registering the lowest diversity, followed by *Asr1*, *Asr2*, and, lastly, *Asr3* due to the large number of silent segregation sites (Table 4). *Asr3* was also the gene with the highest number of SNPs shared by the wild and the cultivated groups (21 out of 29, i.e. 72%) while, for the other *Asr*, the proportion did not overtake 25%. In addition to the high number of polymorphic sites in *Asr3*, five wild rice accessions contained the MITE insertion while nine did not. No correlation between the geographical origin of these accessions and the MITE occurrence was found.

We observed, as expected, a lower level of diversity in domesticated than in wild rice. This loss was first evident through the reduced number of haplotypes in cultivated rice as compared to that of the wild rice even though the large majority of the haplotypes from cultivated rice were shared by the wild rice. The same bottleneck could also be observed for all *Asr* when comparing the ratio of θ_W of *O. sativa* to that of the wild rice. This ratio was 66% for *Asr5*, 52% for *Asr3*, but only 31 and 22% for *Asr1* and *Asr2*, respectively, which indicated a severe reduction in diversity as compared to the average of 57% observed previously (Caicedo et al. 2007). When using the ratio of π, the trends were the same, except for *Asr5* for which the diversity was similar in *O. sativa* and wild rice.

Neutrality tests on the small Asr gene diversity

Tajima's *D* (Tajima 1989), Fu and Li's *D** (Fu and Li 1993) methods, K_a/K_s ratio and HKA test (Hudson et al. 1987) were used to test the null hypothesis that the variations identified in the *Asr* regions were selectively neutral under the Wright–Fisher model with a constant population size. These tests were conducted separately for *O. sativa*, the wild rice, and the varietal groups (Table 5), then for the aus and boro sub-groups and the tropical and temperate japonica sub-groups (see Online Resource 7).

Provided that no demographic event occurred, selection on a genomic region leads to a switch on the distribution of the mutations frequency, which can be detected by Tajima's and Fu and Li's tests. At the species level, Tajima's *D* and Fu and Li's *D** were significantly positive for the *O. sativa* set and only *D** was significantly positive in wild rice for *Asr3*. *D** was significantly negative in *O. sativa* for *Asr5* (Table 5). At the varietal group (isozymic group) level, a significant negative *D** was detected in japonica group for *Asr2* and a significant

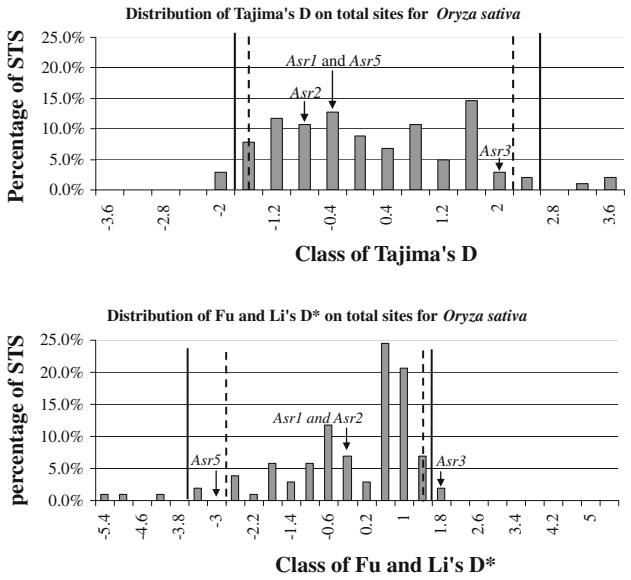

Fig. 4 Comparison of *Asr* genes to *Oryza sativa* general indices of neutrality tests. Tajima's *D* and Fu and Li's *D** were computed for 111 STS (sequence-tag sites). 95% of the distribution (respectively 90%) is represented between the two *dashed lines* (respectively *bold lines*). *Asr1*, *Asr2*, *Asr3* and *Asr5* Tajima's *D* and Fu and Li's *D** are indicated by *arrows* on the corresponding classes

positive *D** in all groups for *Asr3*. At the sub-group level, a significant negative *D** was detected for *Asr1* both in the temperate and tropical japonica and for *Asr2* in the temperate japonica (see Online Resource 7). For *Asr3*, significant positive *D** values were also detected in the boro and temperate japonica sub-groups at 2 and 5% levels, respectively. The tropical japonica sub-group also showed highly significant, but negative *D* and *D** in this gene.

Demographic events, such as expansion or reduction of population size or bottleneck, could have a similar effect on this type of test as selection events. However, demographic events affect the entire genome while selection targets only a genomic region. We used the same dataset (Caicedo et al. 2007) as for assessing the level of diversity (π and θ_W) as neutral reference set, and drew the distribution of Tajima's *D* and Fu and Li's *D** on the 111 randomly distributed STS on the rice genome on the total sites (Fig. 4). The four *Asr* had a Tajima's *D* inside the 95% of the distribution with *Asr3* at the positive extremity of the distribution. The same computations for the silent sites gave similar results (data not shown). For *Asr3,* Fu and Li's *D** was significantly higher (threshold of 95%), while it was lower for *Asr5* (threshold of 90%).

Another way to test a departure from neutrality is to compare the substitution rate of synonymous and non-synonymous mutations which are not affected by demographic events. Under a neutral evolution, these two rates are equal and $K_a/K_s = 1$. In domesticated rice, the K_a/K_s ratio was clearly lower than one for *Asr1*, *Asr2* and *Asr5* while it was not possible to compute the ratio for *Asr3* because of the absence of NS mutation (Table 5). In wild rice, similar low ratios were observed for *Asr1* and *Asr2* while a ratio close to zero was found for *Asr3* and no NS mutation was detected for *Asr5*. All these ratios or absence of NS mutations show some degree of constraint on the evolution of the ASR proteins. The K_a/K_s ratio of domesticated and wild rice was similar for *Asr1*, *Asr2* and *Asr3* showing no change in constraints due to the domestication. For *Asr5*, however, no NS mutation was observed in wild rice, but two were identified in domesticated rice. This suggests some relaxation of constraints on this gene since domestication.

Comparisons between wild and domesticated rice can also be performed through an HKA test. The neutral theory predicts a positive correlation between levels of polymorphism within species and divergence between species. The HKA test was used to determine if levels of nucleotide variation within and between species at two or more loci conformed to this expectation. No significant difference between polymorphism and divergence was detected by this test for the *Asr* gene family (data not shown).

Haplotype analysis

Figure 5 shows the haplotype networks of the short *Asr* genes. We chose to analyze the haplotype organization taking into account the species genetic structure because it gave a much clearer view of the distribution than did geographic patterns. The different varietal groups, sub-groups and wild rice are therefore represented by different colors on the figure. To assess the relationships between haplotypes and hydrological adaptation, we compared the haplotypes detected in the aus sub-group (aerobic) with those detected in the boro sub-group (anaerobic). We did the same for the tropical japonica sub-group (aerobic) versus the temperate one (anaerobic). The within-group comparisons avoided the effect of the strong species structure. For these comparisons, the few accessions of the two varietal groups with unknown sub-group adaptation were removed from the analysis.

The haplotype network of *Asr1*, composed of five haplotypes, showed a major haplotype (H1) occurring in 84.6% of the accessions of *O. sativa*. These accessions belonged to all varietal groups. Four minor haplotypes (H2–H5), derived from H1 by two or three substitutions, were observed almost exclusively in indica accessions. The wild rice accessions shared the same haplotypes as the cultivated accessions, except for H4, and contained five additional haplotypes. Pairwise F_{ST} between varietal groups confirmed the limited effect of rice structure on the haplotypic organization of this gene because of the presence of a major widespread haplotype (see Online Resource 8). Significant,

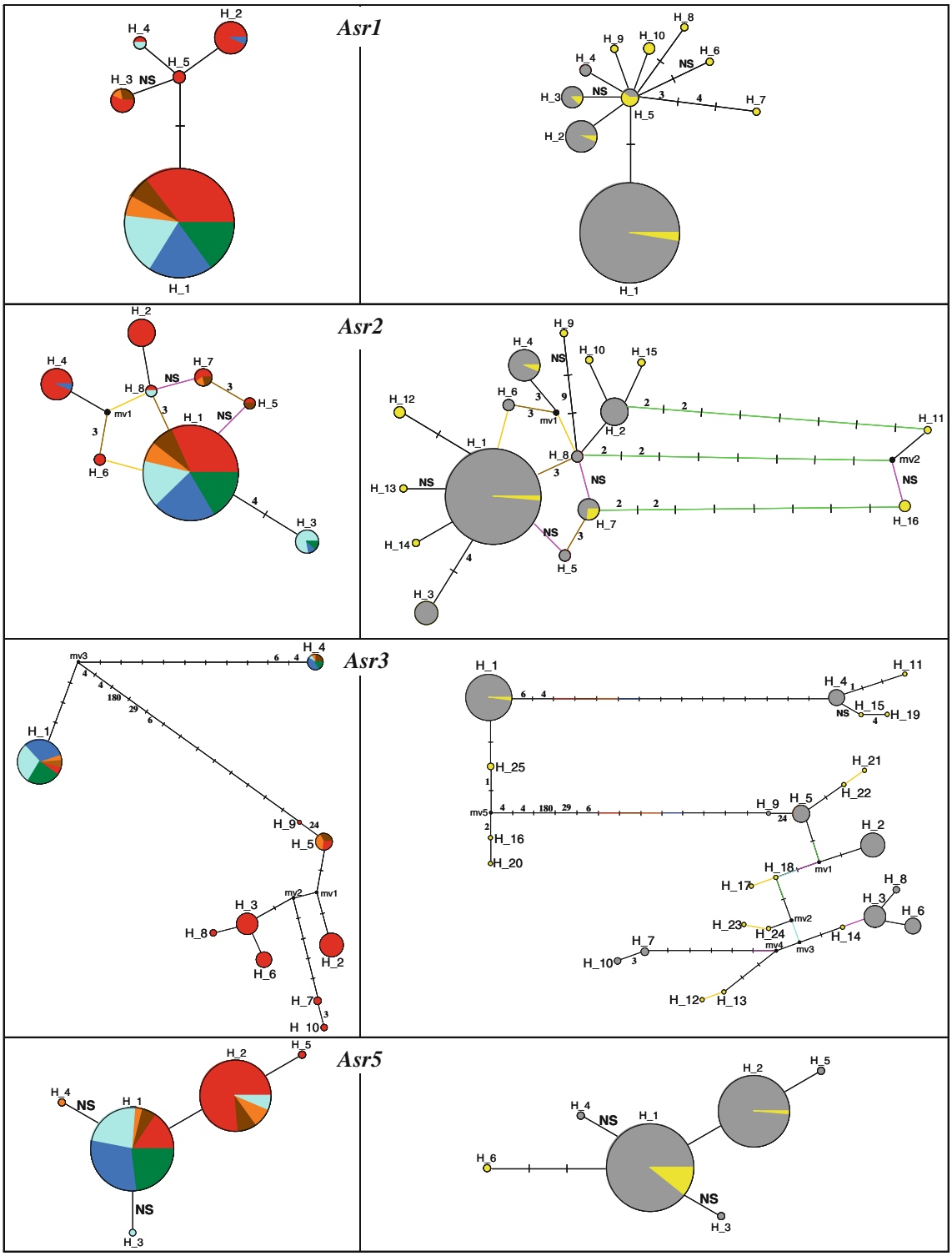

Fig. 5 Haplotype networks of *Asr1*, *Asr2*, *Asr3* and *Asr5*. Each node represents a haplotype, its size being proportional to its frequency. The nodes called mv are non-identified haplotypes created by Networks v4.5 to explain identified haplotypes. A segment alone corresponds to a substitution; a segment with a *number* refers to an Indel and its size. *Colored segments* correspond to identical mutations in different branches. *Part left* represents haplotype networks of the 197 *O. sativa* accessions: *red* indica, *orange* aus, *brown* boro, *light blue* tropical japonica, *dark blue* temperate japonica and *green* sadri/basmati. *Part right* represents together the 197 *O. sativa* accessions colored in *gray* and 14 wild rice accessions colored in *yellow*

but low F_{ST}, ranging from 0.101 to 0.127, were identified only between indica and sadri/basmati or the two japonica sub-groups on the one hand and between boro and the two japonica sub-groups on the other. No significant F_{ST} were detected between the aus and boro sub-groups or between the tropical and temperate japonica sub-groups.

The haplotype network of *Asr2* was similar to that of *Asr1*, with a slightly higher differentiation. It was composed of eight haplotypes and, as for *Asr1*, it showed a major haplotype H1 that represented 75.6% of accessions across all varietal groups. Seven minor haplotypes (H2–H8) were derived from H1 by one to three mutations. Six of them (H2 and H4–H8) were composed almost exclusively of indica accessions while H3 was composed mainly of tropical japonicas. H7 and H4 could be derived directly or indirectly from a recombination event. The composition of haplotype H1 of *Asr1* corresponded to the composition of H1, H2 and H3 of *Asr2*, H2 of *Asr1* to H4 of *Asr2*, H3 to H5 and H7, H4 to H8 with just a few exceptions. The wild rice accessions carried H1, H4, H7 or nine specific haplotypes that were located mostly in positions external to the network. The pairwise F_{ST} between varietal groups showed a stronger effect of rice structure on the haplotypic organization of *Asr2* than on that of *Asr1* (see Online Resource 8). Significant but low F_{ST}, comprised between 0.090 and 0.168, were identified between the indica and all the other groups, between the sadri/basmati and the tropical japonica groups, and between the temperate and tropical japonica sub-groups. However, no significant F_{ST} was detected between the aus and boro sub-groups.

Asr5 was characterized by five haplotypes in *O. sativa*. The two major haplotypes H1 and H2 had a frequency of 56.5% and 42.0%, respectively, in the collection. They were separated by only one substitution in the intron. H1 was mostly composed of japonica and sadri/basmati accessions (53.2% japonica, 22.9% sadri/basmati, 15.6% indica, and 8.3% aus/boro) while H2 was mostly composed of indica and aus/boro accessions (76.6% indica, 17.3% aus/boro and 6.1% tropical japonica). The three other haplotypes, H3–H5, were singletons. Most Wild rice accessions were included in H1 except one accession that had H2 and another one that had a specific haplotype. The pairwise F_{ST} between varietal groups reflected this organization (see Online Resource 8). Pairwise F_{ST} between two groups—indica, aus and boro on the one hand, and sadri/basmati, tropical japonica and temperate japonica on the other hand—were significant and high (between 0.240 and 0.781). A low, but significant F_{ST} was registered between tropical and temperate japonica (0.158). No significant F_{ST} was detected between aus and boro sub-groups.

We performed a detailed analysis of *Asr3* haplotype structure given its high level of variation. A total of 55 polymorphisms spanned along the ca. 1-kb sequence, forming 25 distinct haplotypes of which 19 are rare (Fig. 6). Cultivars displayed three groups of haplotypes (Fig. 5). The first group was composed of H2, H3 and H5–H10 in *O. sativa*, and carried the MITE insertion and at least 16 additional mutations. This group was made of indica types; of particular interest was one of the alleles in position 319 which formed H6 at the edge of the haplotype network and was found in accessions from the Indian subcontinent and Madagascar, whereas it next neighbor (H3) was composed predominantly of accessions from Southeast Asia. The two other groups were represented by H1 and H4 in *O. sativa*, and did not carry the MITE insertion. The former group (H1) was the largest group and had representatives of all cultivar types, albeit with fewer indicas. However, H1 was intermediate in the sense that it had interspersed markers typical of both other groups, and also had specific markers in positions 335, 659, 681 and 696. The later group (H4) included representatives of all rice types except indica, with a bias towards cold tolerant forms (temperate japonicas and boro types). Adding wild forms made the network more complex but the same classification in three groups was found. Wild rice accessions shared H1 or 15 other specific haplotypes, with nine of them close to the MITE insertion group (adding six polymorphisms compared to cultivars), four close to the H1 haplotype (two more polymorphisms), and three close to the H4 haplotype (three more SNPs). A remarkable feature in the MITE insertion group was the appearance of two SNPs in the upstream 5′ region, one of which (position 43) differentiated all wild types from all cultivars. An alternation between alleles typical of H1 or H4 groups along the gene sequence was observed, except in the 3′ edge, which was purely made of H4 group alleles. The variation among wild forms did not relate to geographic origins. In contrast to the situation encountered for *Asr1* and *Asr2*, the high values of pairwise F_{ST} between varietal groups confirmed the strong effect of rice structure on the haplotypic organization of this gene (see Online Resource 8). The F_{ST} differentiated the japonica and sadri/basmati groups from the indica and aus/boro groups (pairwise F_{ST} between 0.399 and 0.623). The F_{ST} between sadri/basmati and the japonica sub-groups were not significantly different whereas the F_{ST} between the aus and boro sub-groups and the indica group were significantly different (between 0.228 and 0.257) but twofold lower than their respective F_{ST} with the sadri/basmati and japonica groups. No significant F_{ST} was detected between the aus and boro sub-groups or the tropical and temperate japonica sub-groups.

Linkage disequilibrium (LD) between Asr genes

To assess a possible co-evolution between genes, pairwise LD between the haplotypes of the four genes were estimated first at the species level. Pairwise LD was generally very low

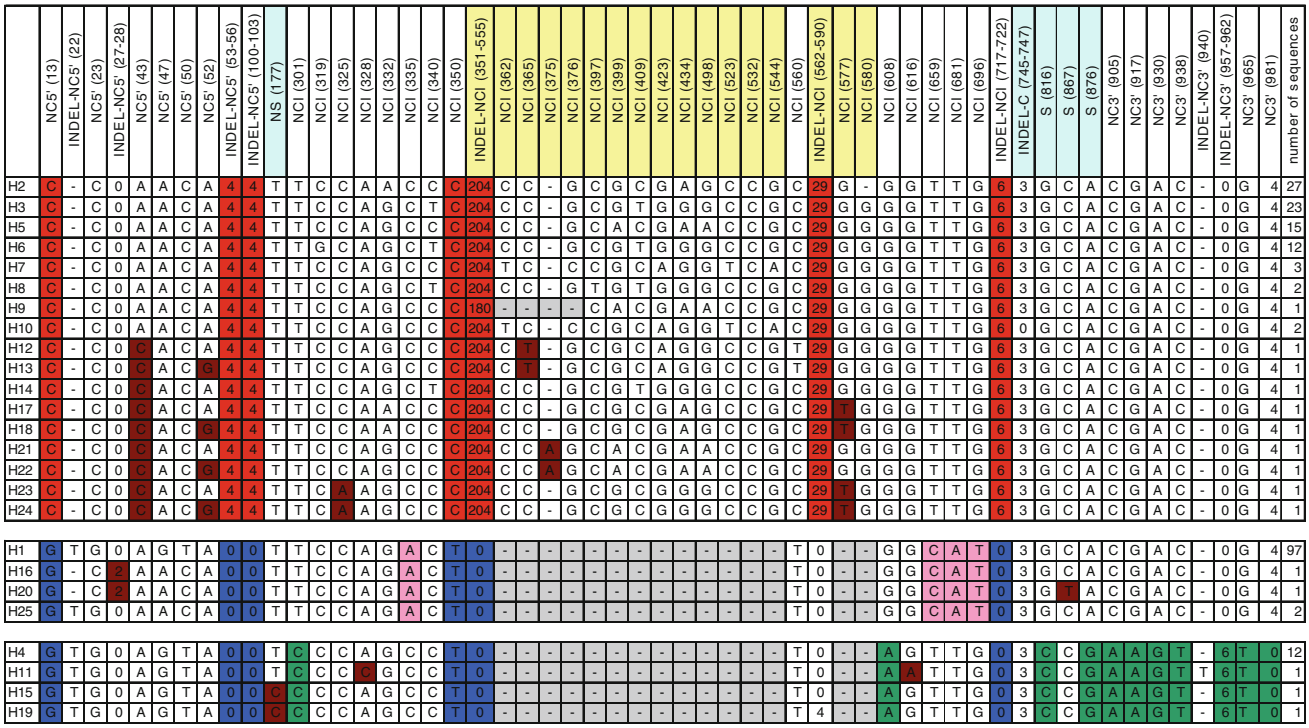

Fig. 6 Haplotypes sequences of the *Asr*3 gene in the rice collection with their respective occurrence. *Light blue* mutation in the coding region, *yellow* insertion/deletion and substitutions in the INDEL, *dark blue* specific mutations of two groups of haplotypes, *pink*, *green* and *red* mutations specific of one group, *brown* mutations specific of wild rice, *NC5′* mutations in the 5′UTR, *INDEL* insertion or deletion, *NCI* mutation in the intron, *NS* non-synonymous mutation in the coding region, *S* synonymous mutation in the coding region, *NC3′* mutation in the 3′UTR

(r^2 below 0.06), except between *Asr3* and *Asr5* (moderate $r^2 = 0.20$) and between *Asr1* and *Asr2* (high $r^2 = 0.50$), which was expected because of their physical proximity (below 1-kb distance) and their parallel haplotypic organization. However, such a level of LD can be explained not only by genome organization but also by population structure. To reduce the effect of population structure, the same analysis was conducted within the groups that showed enough polymorphism in the four genes. The sadri/basmati group and the temperate japonica sub-groups contained only one haplotype per gene with a frequency above 5% for three of the four *Asr* genes and LD could not be computed. LD between *Asr1* and *Asr2* stayed at a high level in the indica ($r^2 = 0.50$) and in the aus/boro groups ($r^2 = 0.30$), but decreased in the tropical japonica sub-group ($r^2 = 0.14$). The LD between all other gene pairs decreased to below 0.10 except for the LD between *Asr2* and *Asr5* for the aus/boro group ($r^2 = 0.16$). These results indicate little concerted evolution between genes besides that possibly due to physical proximity or population structure.

Discussion

Rice *Asr* genes represent an example of small gene family with a simple structure and potential role in drought tolerance. This paper is the first attempt to characterize its diversity in rice.

Asr gene family extension and subfunctionalization

Gene families can arise through tandem amplification, resulting in a clustered occurrence, or through either segmental duplication of chromosomal regions or WGD, resulting in a scattered occurrence of family members (Schauser et al. 2005). Recently, 18 distinct pairs of duplicated segments that cover 65.7% of the rice genome have been identified (Yu et al. 2005). The duplication of 17 of these pairs can date back to a common time before the divergence of grasses, 55–70 Mya ago (Paterson et al. 2004; Yu et al. 2005). The authors concluded this to be a result of a WGD. Although no *Asr* is located in these 17 pairs of duplicated segments, nevertheless some of them may result from this event because of their locations on different chromosomes. Massive gene losses and chromosome rearrangement, following the large-scale genome duplications, have occurred in rice (Yu et al. 2005). This may explain why more than 30% of the genome was not identified as duplicated segments. However, although the stringent criteria used in that study did not identify duplicated segments in *Asr* regions, traces of synteny are still visible between the *Asr3* and *Asr4* areas and between the

Asr5 and *Asr6* areas. *Asr3* might be a paralog of *Asr4* resulting from WGD and *Asr5* a paralog of *Asr6*. *Asr1* and *Asr2*, orientated in opposite directions 1 kb apart, may have more likely resulted from an illegitimate recombination in which inversion occurs relative to the original duplication (Katju and Lynch 2003). Although located on chromosome 1 like *Asr1* and *Asr2*, *Asr6* is 177 kbp away from *Asr1* and *Asr2*. Multiple hypotheses can be proposed for its origin, but it is not possible to decide on the best with the present data. It may correspond to the ancestral gene or be a result of replicative processes such as illegitimate recombination (Holt et al. 2002) or insertion(s) after duplication (Achaz et al. 2000, 2001).

ASR1 and ASR2 proteins are the closest ASR in rice with 73.7% of amino acid identity. Gene duplication might lead to an accumulation of mutations in one copy (Ohno 1970). This may result in a pseudogenisation or the appearance of a novel function. In any case, it would result in a high divergence between the two genes. *Asr1* and *Asr2*, however, cannot be considered as highly divergent genes, notably when considering the similarity between their haplotypic organizations. Recent duplication can be ruled out to explain both the high identity and the similarity of the intron lengths in these two genes. Sorghum and *Brachypodium dystachion* orthologs to *Asr1* and *Asr2*, determined by phylogenetic analysis (data not shown) and physically positioned by blast on the available pseudomolecules of the two species, are also clustered in inverted tandem repeats, suggesting that the duplication had already occurred in the last common ancestor of the grasses dated to 55–70 Mya.

Gene conversion is known to cause concerted evolution of gene family members through reciprocal exchange of small sequences between paralogs. It could explain the higher similarity between these two genes than between the others *Asr* genes and such phenomenon occurs frequently among gene family members in rice (average frequency of 1.4 conversions per megabase) and all the more among genes that are physically close on a chromosome (Xu et al. 2008). No conversion was detected between *Asr1* and *Asr2* (data not shown), but the genes are small, and therefore, the detection power is not very high.

The maintenance of this cluster, as well as other *Asr* genes and more generally the gene family, is an interesting evolutionary question. The DDC (duplication–degeneration–complementation) model (Force et al. 1999) may explain maintenance of paralogs. In one hypothesis of this model, after duplication, the promoter is affected by mutations, but not the coding region, leading to different but complementary expression patterns. *Asr1* and *Asr2* seem to have the same expression pattern in source leaves but a different expression pattern in sink leaves, at least in our preliminary experiment. Additional expression experiments in other organs such as roots and at other development stages could help in answering the question of whether these two genes are functional complements. Sequence information on the promoter regions of *Asr1* and *Asr2* (part of it probably included within the 850 bp region between both genes) may also provide some clues. The other hypothesis of the DDC model proposes that mutations in functional domains lead to a functional specialization for each gene and complementation between genes. All duplicated and specialized genes are required to keep the ancestral function. The divergence among the five distinctive motifs (Battaglia et al. 2008) and the different combinations of motifs encountered in the ASR family could be a result of such specialization.

Studying functional diversification of *Asr* gene family under drought stress is still uneasy, because it would require a standardized protocol and it should take into account immediate and long-term expression responses to stress. In our preliminary expression analysis, some *Asr* genes were expressed differentially in source and sink leaves. The most expressed one, *Asr5*, showed a downregulation in source leaves in three varieties (Dom-Zard, Azucena and M202), an upregulation in variety (FR13A) and no expression change in two varieties (IR64 and Apo), at a mean FTSW of 0.4, highlighting potential intrinsic difference in sensibility to drought stress (Salekdeh et al. 2002). In contrast, Rabbani et al. (2003) reported that *Asr5* was constitutively expressed and upregulated by drought stress and ABA application in source leaves of rice (var. Nipponbare). Takasaki et al. (2008) confirmed the *Asr5* upregulation by ABA and GA at the protein level, in source leaves of two varieties, Nipponbare and T65. However, time scale of expression and varieties were different between the different studies. A more precised kinetics and an extension to a larger genotypic diversity would be required now to address the issue of sub-functionalization among the *Asr* gene family.

Asr paralogs have experienced different evolutionary patterns in *O. sativa* and related wild species

Small *Asr* represent a set of genes with similar exon–intron structure and a range of diversity patterns going from a clear differentiation between the two main varietal group (indica and japonica) to nearly no differentiation between them.

The organization of *Asr3*, where a MITE insertion separates the indica and aus/boro groups on the one hand, and the japonica and sadri/basmati groups on the other hand, and large F_{ST} values, is congruent with the hypothesis of two independent domestications from two geographically isolated wild pools that is now widely accepted (Kovach et al. 2007). Isozymic markers (Glaszmann 1987) and more recently DNA markers (Garris et al. 2005) had already

shown that the aus/boro group was closest to the indica group and the sadri/basmati closest to the japonica group, as observed in this case. The organization is less marked for *Asr5* with its two complementary haplotypes, one dominantly japonica and the other dominantly indica, since a large proportion of the accessions of one group carried the haplotype of the other group as well. A similar bipolar but not exclusive organization has been found in both a neutral and a functional nuclear genes (Londo et al. 2006).

The differentiation is even lower in *Asr2* and the least in *Asr1*. Both genes are characterized by a major haplotype which represents 75.6 and 84.6%, respectively, of the collection and which groups indica with japonica accessions. Two hypotheses can be proposed to account for this lack of divergence, considering the fact that none of the genes exhibit a clear pattern of selective constraint. The first hypothesizes that the major haplotype was selected during the domestication of one of the *O. sativa* groups and was subsequently introgressed into all rice genotypes (Second and Ghesquière 1995). An example of such phenomenon was described for pericarp color (Sweeney et al. 2007). The authors demonstrated that the white pericarp allele originated in a japonica ancestor and was introgressed into the indica compartment to the extent of being present in 98% of the *O. sativa* varieties they surveyed. *Asr1*, *Asr2* and to a lower extent, *Asr5* could be examples of such genetic convergence, the most frequent allele not having yet reached full fixation. Enlarging the haplotype structure around the *Asr1* and *Asr2* and *Asr5* loci in diverse populations of *O. rufipogon* and *O. sativa* may help in assessing the value of this hypothesis and, if valid, determine the origin of the ancestor allele (Kovach et al. 2007).

The second hypothesis to explain the existence of a major haplotype common to both indica and japonica groups assumes that this haplotype arose and was selected before the differentiation of indica and japonica from a common *O. rufipogon* ancestor that was dated back to circa 400,000 years (Vitte et al. 2004; Zhu and Ge 2005). This is the conclusion of a study concerning the fixation in both the indica and japonica groups of the same non-shattering allele at locus *Sh4* that is known to account for about 70% of the trait variability (Lin et al. 2007). Other authors go further and, with the same evidences, strongly favor the hypothesis of a unique domestication, since the hypothesis of multiple origins of domesticated rice supposes a very intense set of hybridizations and varietal movements permitting a fast introduction of the favorable alleles across the Himalayan geographic barrier as well as across the fertility barrier, conditions they judge unlikely to be fulfilled (Vaughan et al. 2008). The equal timing of domestication in the indica and japonica groups is also advanced as an argument in favor of a unique event (Caicedo et al. 2007).

Unfortunately, our data do not bring strong evidence toward one hypothesis or the other. Because of the structure of *Asr3*, our data would favor the hypothesis of a double domestication, but, due to the structure of *Asr1* and *Asr2*, the possibility of fixation of some chromosomal segments before domestication cannot be discarded.

However, the examples of fixation of white pericarp and non-shattering alleles in domesticated rice, both concern genes that have been subject to strong human selection pressure since domestication and such intense selection pressure leaves clear traces as was shown for the waxy locus around which a strong selective sweep was detected (Olsen et al. 2006). No trace of selection, in our data, was detected at *Asr1*, *Asr2* and *Asr5*, neither at the species level nor in the varietal groups by Tajima's *D*. The more sensitive Fu and Li's *D** tests indicated significant departure from neutrality only for the japonica component. This suggests that the lack of strong divergence between varietal groups in *Asr1* and *Asr2* and the intermediate situation of *Asr5* may be due to genetic drift. In contrast, the diversity of *Asr3* seems to be partly driven by a balancing selection at the species level and by directional selection in tropical japonica sub-group, a feature of this gene of particular interest.

Asr3, the most divergent gene in the Asr gene family in rice

In domesticated rice such as in wild rice, the estimators of mutation parameters were much higher for *Asr3* than those of the other short *Asr* and than the average of the rice genome. The MITE insertion may have played a role in this diversification. Altogether, the case of *Asr3*, which stands out in terms of diversity parameters, retrieves information of general value for rice diversity analysis. First, the distribution of diversity among wild forms shows no trace of geographic structure. The cultivars display essentially three groups of alleles. This pattern is remarkable given the usual bipolarity of patterns in rice in relation to the indica-japonica differentiation. However, the MITE insertion clearly distinguishes japonica and indica groups, except for a few accessions for which introgressions at this region cannot be excluded. Rice wild forms contrast with cultivars on the basis of both the alleles that disappeared through domestication and the alleles which appear specific to the cultivated species. The former category is expected given the bottleneck and the selection associated with domestication; yet the case of SNP 43 is remarkable because it differentiates all the nine wild forms from all the 86 cultivated accessions of the indica-specific haplotype group. This SNP is located in the upstream 5′ region, which may harbor regulatory elements. Cultivar-specific alleles are more questionable given the limited representation of wild types in our study. Yet, some haplotypes at the edges of the haplotype networks

represent possible, recent radiations within the crop; here, H2 as well as H7 and H10 are dispersed throughout Asia, but H6 is localized in India and Madagascar and may have arisen from a recent mutation at site 319 in the intron. The frequency of such cases of potential crop neo-diversity, also highlighted in sorghum (Figueiredo et al. 2008), is yet unknown, and their interpretation requires survey of more genes in rice.

Asr3 was also the gene with the strongest signal of selection. The Tajima's *D* and Fu and Li's *D** tests were positive and significant at the species level and, for the Fu and Li's *D** test, for all varietal groups and sub-groups, except for tropical japonica for which *D* and *D** were strongly negative. Positive *D* or *D** indicates an excess of intermediate frequency mutations in *O. sativa*. Three hypotheses could explain this pattern: a balancing selection, a population expansion after a bottleneck or the structure of the species. A demographic explanation should have a genome-wide effect, but no such pattern was found in the other *Asr*. Moreover, *Asr3* neutrality tests were extreme compared to the distribution of Caicedo's data. *Asr3* evolution seems to be generally shaped by a balancing selection. In an inbred such as rice, one form of the balancing selection, heterozygote advantage, is ineffective. Possible other mechanisms include frequency-dependent selection and fluctuating environmental conditions (Stahl et al. 1999). The tropical japonica group, however, is characterized by negative *D* and *D** that indicates an excess of rare mutations potentially indicative of an ancient bottleneck, a recent expansion or ancient directional selection. These two statistics are the lowest as compared to those observed for the tropical japonica subgroup in the other *Asr* and for a reference set of 111 genes (Caicedo et al. 2007). If demographic parameters have played a role in shaping the diversity of *Asr3*, they alone cannot explain these extreme values. Directional selection on *Asr3* in tropical japonica, linked to the transition to aerobic conditions and/or related to a direct or indirect selection on biomass under water stress (suggested by QTL co-localization), seems to be the best explanation of the diversity pattern.

Complex links between structural diversity of the short Asr and local adaptation to hydrology environment

If *Asr* structural evolution is important for drought adaptation in rice, we may expect haplotypic differences between upland and lowland accessions for the Asr genes for which selection in the two environments following domestication favoured specific alleles. Upland accessions, more adapted to aerobic conditions, as well as lowland accessions, more adapted to anaerobic conditions, are present in the aus/boro and japonica varietal groups, enabling an assessment of the haplotypic differences independently of the species structure. Within-group differentiation measure by F_{ST} between sub-groups and departure from neutrality can also be used to strengthen the conclusions.

The haplotype networks, however, show that accessions from the aus and boro sub-groups generally share the same haplotype. Based on the *D**, no differentiation is observed in the aus/boro group nor is there any trace of selection in any of the genes, except for *Asr3* in the boro sub-group.

For the temperate and tropical japonica sub-groups, the pattern is less clear-cut, possibly as a result of the larger size of the sample. One haplotype of *Asr2*, H3, seems dominantly composed of upland japonicas. H4 in *Asr3* is mostly composed of irrigated accessions belonging to the temperate japonica sub-group or to other groups. In both cases, however, several accessions of opposite adaptation blur the pattern. Exceptions in the association between H3 and H4 and the adaptation to upland or lowland conditions might be partly explained by the fact that the adaptation of the varieties to a specific hydrology is not very strict. Varieties are often shifted by farmers from one ecosystem to another in areas where there is a continuum from upland to lowland fields along the topo sequence. This situation is frequent in the plateaux of Eastern India where aus accessions are grown (Courtois and Lafitte 1999), but it is less common in the upland conditions of south-east Asia.

Low, but significant F_{ST} were nevertheless observed between tropical japonica and temperate japonica in *Asr2* and *Asr5*. Selection footprints were also found in both sub-groups, which is not surprising for the temperate japonica that were hypothesized to derive from the tropical japonica through a strong bottleneck linked with adaptation to higher latitude (Garris et al. 2005). This adaptation, however, supposed not only changes in water regime, but also adaptation to lower temperature and longer day length.

Several explanations can be attributed to the apparent lack of water-regime specificity of the haplotypes for the members of the gene family other than *Asr3*.

First, because of sub-functionalization, the genes of this family may now play different roles. Some genes of the family might be specifically induced by a water stress, others might be induced by other abiotic stresses and some could have a pleiotropic effect on different functions. The effect of selection for adaptation to a hydrology is therefore seen only on the most specialized genes which may be the case of *Asr3*.

Another explanation is that the *Asr* genes alone do not significantly contribute individually to drought tolerance in rice. Drought tolerance is known to be a multifactorial trait, resulting from the complex interaction of several mechanisms (Fukai and Cooper 1995; Ludlow and Muchow 1990). The secondary factors involved, such as root morphology or osmotic adjustment, are themselves controlled by many genes. It is, therefore, unlikely that a unique gene

family, even a large one, can by itself explain a major part of adaptation to different water regimes and does not disqualify some of, or all *Asr* genes from contributing to drought resistance in rice. To precisely define involvement of *Asr* genes in local adaptation to a hydrological environment, additional experimentation is needed. Association mapping could be used within the japonica varietal group of *Oryza sativa*. Precise phenotyping would be required for this purpose, together with targeted genotyping involving broad chromosomal segments around the *Asr* genes.

Conclusions

The *Asr* gene family plays potentially a role in drought tolerance in several plant species. Our study provides a thorough description of the organization of this family, and the worldwide nucleotide and haplotype diversity of four *Asr* in *Oryza sativa* species. Globally low, but contrasted nucleotide diversity was found in the different genes having experienced different evolutionary patterns in rice. No simple link between *Asr* haplotypes and adaptation to different hydrological environment has been detected. Implication of the *Asr* gene family in drought resistance in rice will have to be confirmed in future association study relating stress tolerance and genetic polymorphisms in this collection. However, the apparent balancing selection at the species level and directional selection in the tropical japonica subgroup in *Asr3* is a feature of particular interest.

Acknowledgments This work was funded by the Generation Challenge Program project number 2006-02. Thanks to Carine Charron and Nabila Yahiaoui for their critical reviews.

References

Achaz G, Coissac E, Viari A, Netter P (2000) Analysis of intrachromosomal duplications in yeast *Saccharomyces cerevisiae*: a possible model for their origin. Mol Biol Evol 17:1268–1275

Achaz G, Netter P, Coissac E (2001) Study of intrachromosomal duplications among the eukaryote genomes. Mol Biol Evol 18:2280–2288

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Battaglia M, Olvera-Carrillo Y, Garciarrubio A, Campos F, Covarrubias AA (2008) The enigmatic LEA proteins and other hydrophilins. Plant Physiol 148:6–24

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Bray EA (1993) Molecular responses to water deficit. Plant Physiol 103:1035–1040

Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, Bustamante CD, Purugganan MD (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. Plos Genet 3:1745–1756

Cakir B, Agasse A, Gaillard C, Saumonneau A, Delrot S, Atanassova R (2003) A grape ASR protein involved in sugar and abscisic acid signaling. Plant Cell 15:2165–2180

Caldana C, Scheible WR, Mueller-Roeber B, Ruzicic S (2007) A quantitative RT-PCR platform for high-throughput expression profiling of 2500 rice transcription factors. Plant Methods 3

Canel C, Baileyserres JN, Roose ML (1995) Pomelo fruit transcript homologous to ripening-induced genes. Plant Physiol 108:1323–1324

Carrari F, Fernie AR, Iusem ND (2004) Heard it through the grapevine? ABA and sugar cross-talk: the ASR story. Trends Plant Sci 9:57–59

Chang SJ, Puryear JD, Dias M, Funkhouser EA, Newton RJ, Cairney J (1996) Gene expression under water deficit in loblolly pine (*Pinus taeda*): isolation and characterization of cDNA clones. Physiol Plant 97:139–148

Chapman BA, Bowers JE, Feltus FA, Paterson AH (2006) Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. Proc Natl Acad Sci USA 103:2730–2735

Courtois B, Lafitte R (1999) Improving rice for drought-prone upland environments. In: O'Toole J, Ito O, Hardy B (eds) Genetic improvement of rice for water-limited environments. IRRI, Los Banos, pp 35–56

de Vienne D, Leonardi A, Damerval C, Zivy M (1999) Genetics of proteome variation for QTL characterization: application to drought-stress responses in maize. J Exp Bot 50:303–309

Droc G, Ruiz M, Larmande P, Pereira A, Piffanelli P, Morel JB, Dievart A, Courtois B, Guiderdoni E, Perin C (2006) OryGenesDB: a database for rice reverse genetics. Nucleic Acids Res 34:D736–D740

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:1–19

Ehrenreich IM, Purugganan MD (2006) The molecular genetic basis of plant adaptation. Am J Bot 93:953–962

FAOSTAT (2008) Food and Agriculture Organization of the United Nations. http://faostat.fao.org/

Figueiredo LFD, Calatayud C, Dupuits C, Billot C, Rami JF, Brunel D, Perrier X, Courtois B, Deu M, Glaszmann JC (2008) Phylogeographic evidence of crop neodiversity in sorghum. Genetics 179:997–1008

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Frankel N, Hasson E, Iusem ND, Rossi MS (2003) Adaptive evolution of the water stress-induced gene Asr2 in Lycopersicon species dwelling in arid habitats. Mol Biol Evol 20:1955–1962

Frankel N, Carrari F, Hasson E, Iusem ND (2006) Evolutionary history of the Asr gene family. Gene 378:74–83

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Fukai S, Cooper M (1995) Development of drought-resistant cultivars using physio-morphological traits in rice. Field Crops Res 40:67–86

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Giombini MI, Frankel N, Iusem ND, Hasson E (2009) Nucleotide polymorphism in the drought responsive gene Asr2 in wild populations of tomato. Genetica 136:13–25

Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. Theor Appl Genet 30:781–792

Holt SJ, Cress WA, Van Staden J (2002) Evidence for dynamic alteration in histone gene clusters of *Caenorhabditis elegans*: a topoisomerase II connection? Genet Res 79:11–22

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Iusem ND, Bartholomew DM, Hitz WD, Scolnik PA (1993) Tomato (Lycopersicon-Esculentum) transcript induced by water-deficit and ripening. Plant Physiol 102:1353–1354

Jeanneau M, Gerentes D, Foueillassar X, Zivy M, Vidal J, Toppan A, Perez P (2002) Improvement of drought tolerance in maize: towards the functional validation of the Zm-Asr1 gene and increase of water use efficiency by over-expressing C4-PEPC. Biochimie 84:1127–1135

Kalifa Y, Gilad A, Konrad Z, Zaccai M, Scolnik PA, Bar-Zvi D (2004) The water- and salt-stress-regulated Asr1 (abscisic acid stress ripening) gene encodes a zinc-dependent DNA-binding protein. Biochem J 381:373–378

Katju V, Lynch M (2003) The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. Genetics 165:1793–1803

Kato S, Kosaka H, Hara S (1928) On the affinity of rice varieties as shown by the fertility of rice plants. Centr Agric Kyushu Imp Univ 2:241–276

Kim BR, Nam HY, Kim SU, Kim SI, Chang YJ (2003) Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice. Biotechnol Lett 25:1869–1872

Konrad Z, Bar-Zvi D (2008) Synergism between the chaperone-like activity of the stress regulated ASR1 protein and the osmolyte glycine-betaine. Planta 227:1213–1219

Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. Trends Genet 23:578–587

Li CB, Zhou AL, Sang T (2006) Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. New Phytol 170:185–193

Lin ZW, Griffith ME, Li XR, Zhu ZF, Tan LB, Fu YC, Zhang WX, Wang XK, Xie DX, Sun CQ (2007) Origin of seed shattering in rice (*Oryza sativa* L.). Planta 226:11–20

Lin HN, Ouyang S, Egan A, Nobuta K, Haas BJ, Zhu W, Gu X, Silva JC, Meyers BC, Buell CR (2008) Characterization of paralogous protein families in rice. BMC Plant Biol 8

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods 25:402–408

Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc Natl Acad Sci USA 103:9578–9583

Lu BR, Zheng KL, Qian HR, Zhuang JY (2002) Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theor Appl Genet 106:101–106

Ludlow MM, Muchow RC (1990) A critical evaluation of traits for improving crop yields in water-limited environments. Adv Agron 43:107–153

Luquet D, Clement-Vidal A, Fabre D, This D, Sonderegger N, Dingkuhn M (2008) Orchestration of transpiration, growth and carbohydrate dynamics in rice during a dry-down cycle. Funct Plant Biol 35:689–704

Meyer P (1995) Controlling and understanding transgene expression. Transgenics 13:332–337

Nei M (1987) Molecular evolutionary genetics, New York

Ohno S (1970) Evolution by gene duplication. Springer, New York

Oka HI (1988) Origin of cultivated rice. Scientific Societies Press, Japan

Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD (2006) Selection under domestication: evidence for a sweep in the rice Waxy genomic region. Genetics 173:975–983

Pandey S, Behura R, Villano R, Naik D (2000) Economic cost of drought and farmers' coping mechanisms: a study of rainfed rice in eastern India. In: Series IDP (ed), pp 1–35

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R (2005) InterProScan: protein domains identifier. Nucleic Acids Res 33:W116–W120

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA get-blot analyses. Plant Physiol 133:1755–1767

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Rychlik W, Rhoads RE (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. Nucleic Acids Res 17:8543–8551

Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B, Bennett J (2002) A proteomic approach to analyzing drought- and salt-responsiveness in rice. Field Crops Res 76:199–219

Schauser L, Wieloch W, Stougaard J (2005) Evolution of NIN-like proteins in Arabidopsis, rice, and *Lotus japonicus*. J Mol Evol 60:229–237

Schneider A, Salamini F, Gebhardt C (1997) Expression pattern and promoter activity of the cold-regulated gene *ci21A* of potato. Plant Physiol 113:335–345

Second G, Ghesquière A (1995) Cartographie des introgressions réciproques entre les sous-espèces indica et japonica de riz cultivé. In: INRA (ed) Techniques et utilisation des marqueurs moléculaires, Paris, pp 83–93

Skriver K, Mundy H (1990) Gene expression in response to abscisic acid and osmotic stress. Plant Cell 2:503–512

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature 400:667–671

Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, McCouch SR (2007) Global dissemination of a single mutation conferring white pericarp in rice. Plos Genet 3:1418–1424

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Takasaki H, Mahmood T, Matsuoka M, Matsumoto H, Komatsu S (2008) Identification and characterization of a gibberellin-regulated protein, which is ASR5, in the basal region of rice leaf sheaths. Mol Genet Genomics 279:359–370

Vaidyanathan R, Kuruvilla S, Thomas G (1999) Characterization and expression pattern of an abscisic acid and osmotic stress responsive gene from rice. Plant Sci 140:21–30

Vaughan DA, Lu BR, Tomooka N (2008) The evolving story of rice evolution. Plant Sci 174:394–408

Vitte C, Ishii T, Lamy F, Brar D, Panaud O (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Mol Genet Genomics 272:504–511

Walsh B (2008) Using molecular markers for detecting domestication, improvement, and adaptation genes. Euphytica 161:1–17

Wang HJ, Hsu CM, Jauh GY, Wang CS (2005) A lily pollen ASR protein localizes to both cytoplasm and nuclei requiring a nuclear localization signal. Physiol Plant 123:314–320

Wang WK, Schaal BA, Chiou YM, Murakami N, Ge XJ, Huang CC, Chiang TY (2007) Diverse selective modes among orthologs/paralogs of the chalcone synthase (Chs) gene family of *Arabidopsis thaliana* and its relative A-halleri ssp. gemmifera. Mol Phylogenet Evol 44:503–520

Watterson GA (1975) Number of segregating sites in genetic models without recombination. Theor Popul Biol 7:256–276

Widawsky DA, O'Toole JC (1990) Prioritizing the rice biotechnology research agenda for eastern India, New York

Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol 22:506–519

Xu SQ, Clark T, Zheng HK, Vang S, Li RQ, Wong GKS, Wang J, Zheng XG (2008) Gene conversion in the rice genome. BMC Genomics 9

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 strains of wild rice species, *Oryza nivara* and *O. rufipogon*. Genet Resour Crop Evol 50:529–538

Yang CY, Chen YC, Jauh GY, Wang CS (2005) A lily ASR protein involves abscisic acid signaling and confers drought and salt resistance in Arabidopsis. Plant Physiol 139:836–846

Yu J, Wang J, Lin W, Li SG, Li H, Zhou J et al (2005) The genomes of *Oryza sativa*: a history of duplications. Plos Biol 3:266–281

Zhu Q, Ge S (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytol 167:632–632

Zhu QH, Zheng XM, Luo JC, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol 24:875–888