Developing rice with high yield under phosphorus deficiency: Pup1 sequence to application

Joong Hyoun Chin1, Rico Gamuyao1, Cheryl Dalid1, Masdiar Bustamam2, Joko Prasetiyono3, Sugiono Moeljopawiro2, Matthias Wissuwa3, Sigrid Heuer1*

1International Rice Research Institute (IRRI)
Plant Breeding, Genetics, and Biotechnology division (PBGB)
DAPO Box 7777
Metro Manila, Philippines
E-mail: s.heuer@cgiar.org
Phone: 0063 (2) 5805600- 2761

2Institute for Agricultural Biotechnology & Genetic Resources Research and Development
ICABIOGRAD
Jalan Tentara Pelajar 3A
Bogor, Indonesia
E-mail: sugionom@indo.net.id

3Japan International Research Center for Agricultural Sciences (JIRCAS)
Crop Production & Environment Division
Ibaraki 305-8686
Tsukuba, Japan
E-mail: wissuwa@affrc.go.jp

Keywords: rice, phosphorus, molecular markers, QTL, Pup1
The major quantitative trait locus (QTL) *Phosphorus uptake 1 (Pup1)* confers tolerance of phosphorus (P) deficiency in soil and is currently one of the most promising QTLs for the development of tolerant rice varieties. To facilitate targeted introgression of *Pup1* into intolerant varieties, the gene models predicted in the *Pup1* region in the donor variety Kasalath were used to develop gene-based molecular markers that are evenly distributed over the fine-mapped 278-kb QTL region. To validate the gene models and optimize the markers, gene expression analyses and partial allelic sequencing were conducted. The markers were tested in more than eighty diverse rice accessions revealing three main groups with different *Pup1*-allele constitution. Accessions with tolerant (group I) and intolerant (group III) *Pup1* alleles were distinguished from genotypes with Kasalath alleles at some of the analyzed loci (partial *Pup1*; group II). A germplasm survey additionally confirmed earlier data showing that *Pup1* is largely absent from irrigated rice varieties but conserved in varieties and breeding lines adapted to drought-prone environments. A core set of *Pup1* markers has been defined and sequence polymorphisms suitable for SNP marker development for high-throughput genotyping were identified. Following a marker-assisted backcrossing approach, *Pup1* was introgressed into two irrigated rice varieties and three Indonesian upland varieties. First phenotypic evaluations of the introgression lines suggest that *Pup1* is effective in different genetic backgrounds and environments, and that it has the potential to significantly enhance grain yield in rainfed and irrigated environments under P-deficient field conditions.
INTRODUCTION

*Phosphorus uptake 1 (Pup1)* is a major quantitative trait locus (QTL) located on rice chromosome 12 that is associated with tolerance of phosphorus (P) deficiency in soil (Wissuwa et al., 1998; Wissuwa et al., 2002). Kasalath, the *Pup1* donor variety, was initially identified in a screening of thirty diverse rice genotypes in a P-deficient soil in Japan under rainfed conditions. Subsequently, phenotypic data derived from Nipponbare contrasting near-isogenic lines (NILs) with and without the QTL showed that *Pup1* increases P uptake (Wissuwa and Ae 2001a; Wissuwa et al., 2002) and confers a significant yield advantage (2-4-fold higher grain weight plant⁻¹) in pot experiments using different P-deficient soil types and environments (Chin et al., 2010).

Since the identification of the major QTL *Pup1* almost a decade ago, considerable efforts have been made to understand the underlying tolerance mechanism (Wissuwa and Ae, 2001a; Wissuwa, 2005). However, detailed analyses of common plant P-responses (e.g., organic acid exudation, root hair growth) are difficult in soil and could not be detected under the experimental conditions applied (Wissuwa, 2005). Likewise, gene expression analyses (Agilent gene chips) of root samples did not reveal specific upregulation of P-transporters or other P uptake-related genes in *Pup1* NILs (Pariasca-Tanaka et al., 2009). In the original QTL mapping as well as in a parallel study, an intermediate-effect QTL was identified on the short arm of chromosome 6 (Wissuwa et al., 1998; Ni et al., 1998). Many P-responsive genes colocalize with this QTL region in the Nipponbare reference genome, including the transcription-factor gene *OsPTF1*. As was shown by Yi et al. (2005), overexpression of *OsPTF1* in Nipponbare increased P uptake and vegetative growth in soil and in P-deficient culture solution. However, no obvious P uptake-related gene is located in the *Pup1* region on chromosome 12 although this *Pup1* had a larger effect on tolerance (Heuer et al., 2009).

To identify the genes associated with *Pup1*, the genomic region had been sequenced in the donor variety Kasalath, revealing a complex locus with large insertion-deletions (INDELs) and overall low conservation compared with the corresponding genomic regions in the Nipponbare and 93-11 reference genomes. Of the sixty-eight preliminary gene models, sixteen genes code for transposable elements, more than forty show partial sequence similarity to transposons, and other genes cannot currently be annotated with certainty. The remaining candidate genes encode a putative fatty acid oxygenase, a dirigent-like protein, an aspartic proteinase, hypothetical proteins, and a putative protein kinase (Heuer et al., 2009; this paper). In parallel to the identification of the *Pup1* genes, molecular markers have been developed to initiate marker-assisted breeding. A first set of gene-based markers was developed based on the preliminary *Pup1* gene models and published by Chin et al (2010). The germplasm survey conducted with these markers revealed that *Pup1* is largely absent from modern irrigated rice varieties but that it is highly conserved in drought-tolerant IRRI breeding lines and upland varieties. This suggested that breeders are unknowingly selecting for *Pup1* in breeding.
programs that target drought-prone environments (Chin et al., 2010).

The impact of *Pup1* and other QTLs that enhance yield in P-deficient soil and/or under drought stress (Bernier et al., 2009; Venuprasad et al., 2009) is potentially very high since about fifty percent of the rainfed rice in Asia is grown on problem soils (Haefele and Hijman, 2007). Both P-deficiency and drought are widespread problems, particularly in soils with acidic pH and high concentrations of cations for example, aluminum and iron, which complexate Pi (Kochian et al., 2005; Ismail et al., 2007; Xue et al., 2007). Due to the poor productivity of soils as well as the frequent occurrence of drought and floods in the geographical regions concerned, the local incidence of poverty is overproportionally high. In addition, poor farmers often have no access to or means to purchase fertilizer and other agro-chemicals, which further aggravates the situation. In the future, increasing energy costs and limited P resources (Cordell et al., 2009; Van Kauwenbergh, 2010) will lead to rising fertilizer prices, which will increase the costs of food production and consequently food prices and availability. The recent food crisis had severe and long-lasting effects on resource-poor farmers and consumers in Asia and Africa, which often spend a large proportion (25-40%) of their income on rice alone (Dawe et al., 2010). Efforts have since been intensified to develop rice that is tolerant of major abiotic (and biotic) stresses and first rice varieties tolerant of submergence and drought have been released in Asia (Manzanilla et al., 2011; Bernier et al., 2009; Singh and Singh, 2010). In addition, a global multi-disciplinary network on rice breeding has been put in place to enhance the efficiency of rice breeding and increase the likelihood of impact (IRRI, 2010).

The importance of rice varieties with enhanced yield in P-deficient soils and/or enhanced P-fertilizer use efficiency has now been recognized and requires large-scale breeding efforts. However, phenotypic screenings in problem soils are often constrained by the occurrence of companion stresses (e.g., Al toxicity) that restrict root growth and impair phenotyping for P-efficiency. For those soils, varieties with multiple tolerance need to be developed. On the other hand, favorable soils without major stresses are often saturated with P due to continuous P-fertilizer application and endogenous P concentrations of the soil need to be reduced before genotypic differences can be determined. This requires several years of continuous cropping without application of P-fertilizer. For some abiotic stresses (e.g., salinity and Al-toxicity) reliable culture solution-based phenotyping systems are available that can be used for large-scale screenings. However, it has repeatedly been shown that the phenotypic differences between contrasting *Pup1* NILs cannot be observed in culture solution and soil-based screenings are therefore inevitable. The availability of molecular markers that can at least partially replace and/or complement phenotypic evaluations in the field is therefore of particular value for the development of P-efficient rice varieties. A first generation of gene-based *Pup1* molecular markers for the development of Pup1-introgression lines and an assessment of the *Pup1* locus in diverse rice genotypes has recently been published (Chin et al., 2010). However, the germplasm survey conducted showed that individual Kasalath *Pup1* alleles are present (“partial *Pup1*”) in many
of the analyzed rice accessions. This is likely due to the large number of transposable elements within
the Pup1 region (Heuer et al., 2009) causing genomic re-arrangements or to unspecific amplification
of markers in certain genotypes. Because of the genetic instability of the Pup1 locus, many of the
developed markers were not informative when tested in a wider range of rice accessions and the most
reliable markers were located in a large Kasalath-specific INDEL and were therefore dominant (Chin
et al., 2010). Dominant markers are of some limitation in breeding programs since they do not
differentiate homozygous and heterozygous progeny plants and are also not suitable for the
development of SNP markers, which are increasingly used for genotyping in high-throughput
breeding programs (McNally et al., 2009; McCouch et al., 2010; Zhao et al., 2010).

Here, we report on the development of an extended and improved set of Pup1-specific molecular
markers based on gene models that have now been verified through gene expression and allelic
sequencing data. The markers differentiate between three main groups of genotypes with different
Pup1 haplotypes and were applied for the development of two irrigated and three Indonesian upland
Pup1-introgression lines by marker-assisted backcrossing (MABC). First field-based phenotypic data
are presented indicating that Pup1 has the potential to enhance yield in different genetic backgrounds
and P-deficient environments.

RESULTS
The Kasalath genomic Pup1 sequence (GenBank accession AB458444.1) and the preliminary gene
models published by Heuer et al. (2009) were initially used as a basis for the development of an
improved and extended set of molecular markers. The Kasalath Pup1 locus is structurally defined by
the presence of a large (~90 kb) INDEL that is absent from the Nipponbare reference genome and an
earlier set of Pup1 markers targeting this region was published by Chin et al. (2010). Since all of these
markers were clustered in the INDEL and were therefore dominant, additional genes that are at least
partially conserved in Nipponbare were targeted in this study to develop co-dominant markers and to
identify SNPs suitable for high-throughput marker applications.

Pup1 gene models
The annotation of Pup1 genes in the Kasalath genomic region is challenged by the presence of many
transposable elements and truncated genes (Heuer et al., 2009). Based on in silico analyses, a set of
eight priority candidate genes that are non-transposon-related protein-coding genes has now been
selected for in-depth analyses (Table 1). For the development of molecular markers, these genes were
prioritized since coding genes are more likely to be conserved and stable than transposable elements
across diverse rice accessions. Detailed marker information on target genes can additionally provide
supportive information on the Pup1 major gene and is therefore complementary to the ongoing
cloning efforts (Gamuyao et al, IRRI, unpublished). An overview of the relative position of the genes
and current gene models is given in Fig 1a-e.

Based on the below-described gene expression analyses in conjunction with detailed in silico sequence analyses, two Pup1-gene models have been revised. The new model for the dirigent-like gene (OsPupK20-2) has an altered C-terminal intron-exon structure (Fig 1c) and is highly similar to LOC_Os12g26380 (3.5E-100). The new model for the Pup1 protein-kinase gene (OsPupK46-2) has a revised coding region and predicts an intron-less gene (Fig 1e) most similar to LOC_Os01g49580 (1E-93). Both revised models are supported by cDNA sequence data (data not shown; accession numbers will be provided). All other Pup1 gene models targeted for marker design in this study are published under GenBank accession AB458444.1 (Heuer et al., 2009).

To assess whether the short-listed Pup1 candidate genes are expressed and responsive to phosphorus, RT-PCR analyses were conducted using root and shoot samples derived from Nipponbare and contrasting Pup1 NILs grown in P-deficient soil with and without application of P-fertilizer (Fig 2). The data show that OsPupK04 and OsPupK05 are both ubiquitously expressed. However, higher expression of OsPupK05 in the Pup1 NILs was detected at a lower PCR cycle number (Fig 2, panels 1-3). The gene model of OsPupK05 is a hypothetical protein located in reverse direction within a putative intron of OsPupK04, a putative fatty acid oxygenase (Fig 1b). Expression of the dirigent-like gene OsPupK20 (Fig 1c) is restricted to roots with an apparent higher transcript abundance under P-deficient conditions (Fig 2, panel 4). Likewise, the hypothetical protein OsPupK29 (Fig 1d; see Heuer et al., 2009 for details) is specifically expressed in roots with a higher transcript abundance under P-deficient conditions, especially in the Pup1 NILs (Fig 2, panels 5-6). The Pup1 protein-kinase gene OsPupK46-2 is expressed under −P and +P conditions in roots and in shoots (Fig 2, panel 7). Since this gene is absent from the Nipponbare genome, no expression was detected in Nipponbare and sister NILs without the tolerant Pup1 locus.

For the genes OsPupK01, OsPupK53, and OsPupK67, several primer pairs were designed targeting different regions of the predicted gene models but expression was never observed in any of the root and shoot RNA samples analyzed (data not shown).

**Gene-based molecular markers**

As mentioned above, the hypothetical gene OsPupK05 is located within a predicted intron of OsPupK04 (Fig 1b) based on a sequence comparison with the most similar Nipponbare gene coding for a fatty acid oxygenase (alpha-DOX2; LOC_Os12g26290). The genomic sequence of the Nipponbare gene is >10 kb and several alternative splice products have been identified (Rice Genome Browser v.6.1 at Rice Genome Annotation Project). More detailed analyses are needed to clarify whether OsPupK05 is an independent gene or it represents a splice variant of OsPupK04. Several markers specifically targeting OsPupK04 were designed but failed to reveal reliable data (data not shown). A marker (K5) located within OsPupK05 was therefore selected to target this locus and
accordingly represents both genes (Fig 1a, b). This marker amplifies a region with five INDEL mutations (Fig 4; Table 2).

For the dirigent-like gene, two markers (K20-1 and K20-2) were designed that target different regions of the gene (Fig 1a, c; Fig 4; Table 2). To validate sequence polymorphisms between the Kasalath gene and the most similar Nipponbare gene (Os12g26380), K20-2 PCR amplicons of fourteen rice accessions were sequenced. The data revealed four SNPs and a Kasalath-specific 3-bp deletion (Fig 3a; only polymorphic region is shown). The detected polymorphisms are highly conserved and only three different alleles can be distinguished. Based on this information, the marker K20-1 was designed targeting the INDEL mutation. The resolution of this marker can be enhanced by subsequent digestion of the PCR product with MseI (K20-1<sub>MseI</sub>), which recognized the TTAA motif located 2 bp upstream of the INDEL in the Kasalath allele (Fig 3a; Fig 4). A non-synonymous SNP (at 620 bp) creates a mutation in a Bsp1286I restriction site (GDGCHC; D=A/G/T, H=A/C/T) and causes a change in the amino acid sequence (L<sub>177</sub>: Kasalath; F<sub>177</sub>: Nipponbare). Digestion of PCR amplicons with Bsp1286I (K20-2<sub>Bsp</sub>) generates three DNA fragments specific for the Kasalath allele and two DNA fragments in non-Kasalath alleles (Fig 3a; Fig 4).

The <i>Pup1</i> gene OsPupK29 codes for a hypothetical protein with high but partial sequence similarity to two expressed-protein genes (LOC_Os12g26390 and LOC_Os12g26410) in the Nipponbare reference genome (Fig 1d; Heuer et al., 2009). Based on a comparative sequence analysis of the three genes, three markers were developed that target INDELS located in the first exon (K29-1), intron (K29-2), and second exon (K29-3), respectively (Fig 1a, d; Fig 4; Table 2). The markers K29-1 and K29-3 were the most reliable (see below) and PCR amplicons were sequenced in representative rice varieties revealing high conservation of the targeted polymorphism (Fig 3b, 3c; only polymorphic regions are shown).

For the <i>Pup1</i> protein-kinase gene, two markers were designed amplifying a region including the conserved kinase domain (marker K46-1) and the 3′UTR (marker K46-2), respectively (Fig 1e). Both markers are dominant since this gene is located in the large ~90-kb Kasalath-specific INDEL region (Fig 1a) and therefore are absent from the Nipponbare reference genome and other intolerant rice accessions (Fig 4; see below). The <i>Pup1</i> gene OsPupK67 is nearly identical to an aspartic proteinase gene (Os12g26470) and only four SNPs are present that distinguish the two sequences (data not shown). Since we were unable to detect expression of this gene (see above), this gene was excluded from further analyses in this study.

The gene model OsPupK01 is located in the <i>Pup1</i> region partially conserved in Nipponbare and is highly similar (96.0%) to an intergenic region between the Nipponbare genes Os12g26260 and Os12g26270 annotated as Gypsy-type transposon and hypothetical protein, respectively. RT-PCR analyses of roots and shoots showed that this putative gene is not expressed (data not shown) and it is therefore likely that OsPupK01 is a pseudogene or incorrectly predicted. However, a molecular
marker (K1) for this gene was included in the study because of its physical location at the 5′-border of the *Pup1* locus (Fig 1a). This marker amplifies a DNA fragment flanking a 3-bp insertion (GTC) specific to the Kasalath-gene model (Fig 4; Table 2). The remaining markers target genes coding for putative transposable elements (*OsPupK41* to *OsPupK44*), hypothetical proteins (*OsPupK45* and *OsPupK48*), and an unknown gene (*OsPupK52*) located in the Kasalath-specific INDEL region (Chin et al., 2010; Fig 1a; Fig 4; Table 2).

**Determination of *Pup1* haplotypes in breeding materials**

To standardize PCR conditions and evaluate the markers, genomic DNA from Kasalath, Nipponbare, and five rice varieties that had been selected for *Pup1* breeding was analyzed. The data show that the developed markers clearly differentiate between Kasalath and Nipponbare (non-Kasalath) alleles (Fig 4). With the dominant markers K46-2 and K52, unspecific PCR amplicons are sometimes observed in Nipponbare and other genotypes that can be distinguished from the specific amplicons due to their larger size.

The marker data revealed that IR64 and IR74, the two irrigated varieties selected as *Pup1*-recipient parents, did not possess Kasalath alleles at any of the analyzed loci, with the exception of marker K5. As mentioned above, the gene models targeted by this marker are currently under validation. The data obtained for the three Indonesian varieties selected as *Pup1*-recipient parents suggest that they possess different haplotypes with at least partial presence of Kasalath *Pup1* alleles. Although Situ Bagendit is most similar to IR64 and IR74 (only the K5 Kasalath allele), the other two Indonesian varieties possess additional Kasalath alleles, including the *Pup1* protein kinase (Fig 4). In both varieties, the markers targeting *OsPupK29* revealed inconsistent results, suggesting that Batur and Dodokan may have a different allele of this gene (Fig 4).

**Germplasm survey with *Pup1* markers**

In order to determine the *Pup1* haplotype in a wider range of genotypes, eighty-one rice varieties and breeding lines were genotyped. The number of accessions that possess Kasalath alleles at all analyzed loci was small and was, apart from Kasalath, observed in only five genotypes. Among these were three varieties (Dular, IAC25, and IAC47) that are known to be highly tolerant of P-deficiency (Wissuwa and Ae, 2001b). Likewise, only a few varieties (Akihikari and Nipponbare from Japan) lacked Kasalath alleles at all loci.

More detailed analyses showed that several markers are not associated and are not suitable to differentiate between *Pup1* haplotypes within the analyzed genotypes. This was especially the case for the markers K1 (14.8% Kasalath allele frequency) and K5 (90.3% Kasalath allele frequency) (Fig 5), though K1 showed some specificity for the analyzed *aus*-type varieties (Kasalath, Dular, AUS257, AUS196) and traditional upland-adapted genotypes (e.g., IAC47, IAC25, Vary Lava 701, Ashoka...
A subsequent analysis of all \textit{Pup1} markers with GGT 2.0 (Van Berloo, 2008) revealed very low $r^2$ values for K1 and K5, suggesting that they are not associated with the other \textit{Pup1} markers (Fig 6). The highest $r^2$ values were identified for markers targeting the hypothetical protein \textit{OsPupK29} (markers K29-1 and K29-3), two transposable elements (markers K42 and K45), and the \textit{Pup1} protein kinase gene (marker K46-1) (Fig 6). With the exception of K29-1 and K29-3, all markers are located within the Kasalath-specific INDEL region (Fig 1a; Chin et al., 2010).

It is noteworthy that the intron-specific marker K29-2, in contrast to markers K29-1 and K29-3 that target exons of the same gene, is not associated with the other \textit{Pup1} markers ($r^2=0.34$; Fig 6). As mentioned above, this Kasalath gene corresponds to two genes in Nipponbare that might have evolved by a duplication event and several transposon integrations (Fig 1d; Heuer et al., 2009). The low $r^2$ value of marker K29-2 therefore suggests that some regions of the \textit{OsPupK29} gene are unstable and are still evolving in the genotypes analyzed (Fig 5; Supplementary Table 1).

Likewise, data derived with the two markers targeting the \textit{Pup1} protein-kinase gene are not always consistent. Whereas marker K46-2 indicates presence of the gene in fifty-seven genotypes, the data derived with marker K46-1 suggest that the gene is present in only thirty-eight genotypes (Fig 5; Fig 6; Supplementary Table 1). Sequencing of the K46-2 PCR amplicon subsequently confirmed the specificity of this marker and showed a high conservation between the four analyzed sequences (Fig 3d). However, a low $r^2$ value ($r^2=0.36$; Fig 6) indicates that marker K46-2 is less informative than K46-1 and the latter is therefore recommended for breeding applications and allelic surveys.

Based on this assessment, a core set of the six most informative \textit{Pup1} markers was defined (K29-1, K29-3, K41, K43, K45, and K46-1; Fig 6). Using these core markers, a UPGMA cluster analysis was conducted revealing three main groups of genotypes with a different \textit{Pup1} allele constitution (Fig 5; Supplementary Table 1). Group I includes genotypes with all or most Kasalath \textit{Pup1} alleles. In contrast, Kasalath alleles for the core markers are absent in genotypes within group II and III. Interestingly, group II is mainly defined by the presence of Kasalath alleles for the markers K29-2 and K46-2 (see above), whereas genotypes in group III do not possess any Kasalath allele, with the exception of K5.

In agreement with earlier data (Chin et al., 2010), all genotypes within group I are adapted to rainfed drought-prone environments, whereas the majority (63%) of genotypes within group II and group III are adapted to lowland/irrigated conditions (Fig 5; Supplementary Table 1). The \textit{japonica}-type accessions that were included in this study are represented in all three groups, suggesting that \textit{Pup1} is not an \textit{indica/aus} type-specific locus.

**Development of \textit{Pup1} varieties by marker-assisted backcrossing**

The \textit{Pup1} marker survey described above has revealed the presence or partial presence of the Kasalath \textit{Pup1} locus in the majority of the varieties and breeding lines analyzed. For marker-assisted breeding
applications it is therefore critically important to first determine the Pup1 haplotype in the parental lines. The core marker data indeed indicate that two of the three Indonesian varieties selected for breeding indeed possess some (Batur) or all (Dodokan) Kasalath alleles (Fig 4; Fig 8). In contrast, Kasalath alleles were not detected with the core markers IR64 and IR74 and both belong to group III (Fig 4; Fig 5).

For the development of Indonesian Pup1 varieties, Kasalath and the Pup1 near-isogenic line C443 (Wissuwa et al., 2002) were used as donor parent. Starting at the F₁ generation, progeny plants were genotyped with a set of best suitable Pup1 foreground markers and backcrossed to the recipient parent. After a repeated genotyping, selected BC₁F₂ plants were backcrossed again to the respective recipient parent to restore all desirable traits of the upland varieties. A background screening of the three different BC₂F₁ breeding populations was conducted with, depending on the population, forty-seven to sixty-one polymorphic SSR markers distributed across the twelve chromosomes. The graphical genotype of the populations is shown in Supplementary Figure 1. In the generation analyzed, 16–19% of remaining donor alleles were detected with 1–9% heterozygous loci. Analyses of advanced generations is now ongoing.

For the development of irrigated Pup1 varieties, the near isogenic line NIL14-4 (Chin et al., 2010) was used as the Pup1 donor. Progenies were backcrossed twice to IR64 and IR74, respectively, and genotyped with at least three Pup1 foreground markers (K20-1<sub>Mse</sub>, K46-1, K29-1; data not shown). In the BC₂F₁ generation, twenty-nine out of one hundred thirteen plants with the Pup1 introgression were selected for background genotyping using the Illumina Beadexpress™ SNP chipset that contained 326 markers (M Thomson, IRRI unpublished). The SNP data revealed that, on average, 88.3% of the recipient genome was recovered. Ten plants with the least number of background introgressions were selected and the presence of the tolerant Pup1 locus was re-confirmed in BC₂F₃ progenies (data not shown). From these plants, contrasting sister lines with and without the tolerant Pup1 locus (Pup1+/Pup1−) were selected and again genotyped. The data showed that background introgressions were further reduced and that 89–93% of the recipient genome was restored (IR74-Pup1+: 93%; IR74-Pup1−: 93.4%; IR64-Pup1+: 92.6%; IR64-Pup1−: 89.4%).

**Phenotyping of Pup1 breeding lines under field conditions**

The IR64-Pup1 and IR74-Pup1 breeding lines have been subjected to a very stringent genotypic selection (see above) and only nearly identical and non-segregating BC₂F₃ sister lines (Pup1+/Pup1−) from each cross were selected for an initial field evaluation. Plants were grown at the IRRI experimental station (Philippines) in 2010 in a P-deficient irrigated plot and a control plot under P-fertilized conditions (equivalent to 60 kg P₂O₅ ha⁻¹). In the unfertilized P-deficient field plot, IR74-Pup1+ had significantly higher grain yield (+73.1%) and was taller (20.5%) than the IR74-Pup1− sister line and the IR74 parent (Fig 7a). Some degree of superiority remained under P-fertilized...
conditions, in which IR74-Pup1+ developed a 15.1% higher grain weight plant⁻¹ and was 14.2% taller than the controls. The number of panicles was increased in IR74-Pup1+ under both P conditions but the data were not statistically significant. A higher grain yield (33.0%) was also observed in IR64-Pup1+. However, this cannot be attributed to Pup1 since a significantly higher (44.7%) grain yield was also observed in the sister line without Pup1 (IR64-Pup1–). This is likely due to the presence of remaining background introgressions and analyses are now ongoing. However, IR64-Pup1+ plants were significantly taller than both checks (IR64 and IR64-Pup1–) under +P and –P conditions (Fig 7 a). Sister lines (BC₂F₄) of both crosses have now been selected based on genotypic and phenotypic data (Fig 7b) and seeds are being multiplied for further field evaluations in collaboration with rice breeders in Asia and Africa. In parallel, an additional backcross has been conducted to further remove the remaining background introgressions and to provide additional Pup1 breeding lines.

The Indonesian Pup1 breeding lines (BC₂F₄) were assessed under field conditions in Sukabumi (Indonesia, West Java) in 2010 under rainfed conditions in a field with an intermediate high endogenous P content (8.3 ppm available P) and in a P-fertilized control plot (39.3 kg P ha⁻¹). The phenotypic data of the BC₂F₄ generation analyzed showed that most Situ Bagendit-Pup1 lines outperformed the original parent under both P-fertilized and P– conditions (Fig 8a, left). On average, the best lines developed a 22% (P+) and 14% (P–) higher grain yield compared with Situ Bangendit. In the Batur-Pup1 population, the largest effect was observed under P-fertilized conditions and the best lines developed a two fold higher grain yield than Batur (Fig 8, middle). In contrast, only one Dodokan-Pup1 line showed a better performance compared with Dodokan (Fig 8, right). This finding is in agreement with the genotypic data (Fig 4) showing that the tolerant Pup1 locus is naturally present in Dodokan, whereas it is absent in Situ-Bangendit, in which we have observed the largest improvement in grain yield. In Batur, some Kasalath Pup1 alleles were detected, including the protein-kinase gene (Fig 4). Interestingly, improved performance of the Batur breeding lines was mainly detected under P+ conditions (Fig 8), suggesting that Batur naturally possesses some tolerance of P-deficiency, possibly related to the presence of the Pup1 protein kinase.

DISCUSSION

Pup1 core markers

The main objective of this study was to develop an ideal set of molecular markers that facilitates targeted introgression of the Pup1 major QTL into diverse intolerant genetic backgrounds and for validation under field conditions. For breeders, only QTLs that express their beneficial effect in different genetic backgrounds and environments are of interest and can be widely applied for the development of tolerant rice varieties.

The gene-based Pup1 markers that are now available (Chin et al., 2010; this paper) have been
extensively tested and provide sufficient details on the different Pup1 haplotypes present in the diverse rice genotypes analyzed. Among the tested markers, a core set of six markers is best associated with Pup1. The finding that this core region is flanked by two markers located within targeted genes (OsPupK29 and OsPupK46) is interesting and requires more detailed analyses. As outlined above, OsPupK29 appears to be an unstable gene and its location at the 5′-border of the Pup1-specific INDEL region suggests that it might have contributed to the evolution of Pup1. The 3′-border of the core-marker set is located within the INDEL region and is defined by K46-2, one of the two markers targeting the protein-kinase gene. The K46-2 reverse primer is located within the predicted 3′-UTR and, since untranslated regions are generally less conserved, it was expected that this marker is more specific than K46-1, which is located within the coding region. The finding that, in most genotypes within group II, an amplicon was derived only with marker K46-2 but not K46-1 suggests that the coding region is mutated or truncated in some genotypes and that the K46-1 primers do not bind to the mutated gene. These data demonstrate the complexity of functional-marker design and suggest that multiple markers should be used to assess major genes in breeding programs.

Recently, rapid progress has been made with the development of SNP markers that facilitate high-throughput genotyping of breeding populations (Kim et al., 2009; McCouch et al., 2010). This technology is a very useful tool for background genotyping of mapping and breeding populations and is already being applied for whole-genome association studies (McNally et al., 2009). The development of SNP chips with a collection of high-value genes for foreground selection is under way (IRRI, unpublished) and will enable breeders to routinely screen their populations for important tolerance genes and other agronomically useful genes. In order to include Pup1 in this panel, it will be necessary to develop SNP markers for the Pup1 core genes. Based on allelic sequencing data, SNPs have been identified for three genes (OsPupK20, OsPupK29, OsPupK67) that can be targeted for this purpose. However, the allelic sequence data derived from K20-2 and K29-2 amplicons suggest that the INDEL mutations targeted for marker design in this study might be more informative than the SNPs across genotypes (Fig 3a, b). Likewise, the genes located in the Pup1-specific INDEL cannot be analyzed by SNP markers since SNPs are based on sequence polymorphisms between genes present in both targeted genotypes. A dominant marker system for the Kasalath-specific genes, most importantly for OsPupK46, is therefore inevitable. However, our sequence data show that multiple alleles exist within genotypes that possess this gene, which might be explored for SNP marker development. Interestingly, the highest number of SNPs was detected in Dular, the aus-type variety that showed the highest tolerance in the original screening from which Pup1 was identified (Wissuwa and Ae 2001a). However, whether these polymorphisms are of any significance with respect to gene function and tolerance of P-deficiency remains to be determined.

Pup1 candidate genes
Based on the above-described marker data and the RT-PCR expression analysis, the dirigent-like gene OsPupK20, the hypothetical protein OsPupK29, and the protein-kinase gene OsPupK46 have been short-listed as priority candidate genes. Since none of these genes code for a known structural P uptake-related gene, the mode of action of Pup1 is still unclear. However, the expression data suggest that OsPupK20 and OsPupK29 have a root-specific function. Dirigent proteins catalyze the polymerization of lignin monomers, a reaction that involves free radicals, and the dirigent gene family has been shown to play a role in plant defense response (Wu et al., 2009; Bhuiyan et al., 2009). Dirigent proteins also catalyze the synthesis of lignans, a diverse group of polyphenolic substances derived from phenylalanine that may play a role in plant pathogen defense and are important anti-cancer substances in human diets (Davin and Lewis, 2000; Davin and Lewis, 2005; Yoo et al., 2010).

Up-regulation of dirigent genes under P-deficiency in roots (At1g64160) and leaves (At2g21100) has been reported in Arabidopsis (Misson et al., 2005), which is in agreement with our data. Furthermore, a dirigent-like gene has also been identified as a candidate gene for heat tolerance from the aus-type variety N22 (Jagadish et al., 2010), suggesting that this gene family may have a more general function in responses to abiotic stresses.

The hypothetical protein OsPupK29 is of interest because of its root-specific expression and putative role in Pup1 evolution (see above). The two Nipponbare genes that show high partial sequence similarity are according to MPSS data (http://mpss.udel.edu/rice/mpss_index.php), expressed at a low level in crown meristems (Os12g26410) and in young, drought-stressed roots (Os12g26390). The function of these genes is currently unknown and it will be interesting to conduct comparative studies between the Kasalath and Nipponbare genes.

The presence of a novel protein-kinase gene suggests that Pup1 might confer tolerance via a regulatory pathway. OsPupK46 is predicted as a Ser/Thr kinase and as such shows homology to many members of this gene family in rice. In Arabidopsis, an Affymetrix microarray analysis has shown up-regulation of fifteen Ser/Thr kinase genes under P-deficient conditions (Misson et al., 2005). Of these, At1g67000, which was specifically up-regulated in roots, was most similar to the Pup1 kinase (8e-93). However, a TAIR BLASTp search revealed a higher sequence similarity with SNC4 (At1g66980; 9e-99) and PR5K (At5g38280; 6e-97), two defense-related receptor kinases (Bi et al., 2010; Wang et al., 1996). All of the above-mentioned genes possess an N-terminal receptor/membrane domain that is absent from the Kasalath gene and it is therefore unlikely that OsPupK46 is an orthologous gene. In yeast, the PHO regulon has been described in detail (for a review see Oshima, 1997). The regulon involves a CDK-dependent kinase (PHO85) that regulates the phosphorylation of the PHO4 transcription factor in a P-dependent manner. Under low-P conditions, PHO4 enters into the nucleus in its under-phosphorylated form, where it activates transcription of the acid phosphatase gene PHO5 (Schneider et al., 1994; O’Neill et al., 1996). In contrast to the other genes mentioned above, PHO85 has no N-terminal transmembrane domain and is located in the cytoplasm. Whether the Pup1 kinase
has a function similar to PHO85 remains to be analyzed.

**Pup1 breeding**

The marker survey confirmed earlier data (Chin et al., 2010) that Pup1 is conserved in most rice varieties developed for rainfed stress-prone environments, suggesting that breeders are unknowingly selecting for this QTL. The molecular markers developed now provide breeders with a tool to ensure that Pup1 is present in their breeding materials. Among the genotypes with the tolerant Pup1 locus are several varieties that are known for their superior behavior under P-deficient conditions (Dular, IAC47, IAC25). Dular had shown a better performance under stress than Kasalath in the original screening (Wissuwa and Ae, 2001) and it will be important to assess whether allelic differences in the Pup1 kinase gene are related to the higher tolerance and to map any additional QTLs that might be present in Dular to further enhance tolerance.

For the development of the Pup1 breeding lines, we have followed a marker-assisted backcrossing approach that was first applied for the development of submergence-tolerant rice varieties (e.g., Swarna-Sub1 and IR64-Sub1). These Sub1 varieties are now being widely adopted in submergence-prone regions in Asia and Africa (Xu et al., 2006; Septiningsih et al., 2008; Venuprasad et al., 2009). The same approach is also being applied for the development of drought- and salinity-tolerant varieties (Bernier et al., 2007; Bernier et al., 2009; Ismail et al., 2007; Thomson et al., 2010; Kim et al., 2009). The principal idea of MABC is to use widely grown and locally well-adapted rice varieties from target countries as recipient parents for the introgression of tolerance QTLs. The desirable plant type of the local variety is subsequently restored by repeated backcrossing and selfing in conjunction with genotypic selection using foreground, background, and QTL-flanking (recombinant) markers (Collard and Mackill, 2008). Tolerant varieties developed by this approach are more likely to be adopted by farmers since eating quality and other traits remain unchanged.

The development of the Pup1 varieties is still work in progress and additional backcrosses and marker selection are ongoing to further remove remaining background introgressions. However, the initial field data of the five breeding populations suggest that Pup1 has the potential to significantly improve plant performance. In Indonesia, the largest beneficial effect of Pup1 on grain yield (up to a 2-fold increase) was obtained in the genetic background of Situ Bagendit and Batur. In Batur, this was mainly observed under P-fertilized conditions. It will be interesting to analyze whether the lack of improvement under P-deficient conditions is related to the Pup1 protein kinase, which is naturally present in Batur, implying that this variety might already be tolerant of P-deficiency. In agreement with this, Dodokan-Pup1 breeding lines did not show improved performance in the field experiment since the tolerant Pup1 locus is already present in Dodokan.

For the indica breeding lines, a significant, Pup1-specific increase in grain yield was observed in the genetic background of IR74 under irrigated P-deficient and P-fertilized conditions. Based on earlier
data derived from pot experiments with Nipponbare-Pup1 NILs that suggested that Pup1 does not have a large effect under flooded conditions (Chin et al., 2010), the observed yield advantage in IR74-Pup1 is very encouraging. A beneficial effect of Pup1 under P-deficient lowland/irrigated conditions in at least some genetic backgrounds would greatly enhance the potential impact of this QTL. For the identification and evaluation of major QTLs, a robust and reliable phenotyping system is a prerequisite. Since Pup1 cannot be phenotyped in liquid culture solutions, in contrast to e.g., salinity and aluminum toxicity (Chin et al., 2010; Thomson et al., 2010; Chen et al. 2006), phenotyping has to be conducted in soil, which is inconvenient and a problem for physiological and molecular analyses. In addition, many P-deficient soils are constrained by other stresses (e.g., Al-toxicity, salinity, acidity, nematodes; Ismail et al., 2007; Kumar et al., 2007; Wissuwa et al., 2002) that restrict root growth and interfere with Pup1 phenotyping. In the future it will therefore be important to combine tolerance of P-deficiency with tolerance of other predominant stresses. Major QTLs for tolerance of drought and salinity have already been identified and efforts are now under way to identify and validate QTLs for tolerance of aluminum toxicity (Bernier et al., 2009; Venuprasad et al., 2009; Wu et al., 2000; Nguyen et al., 2003; Kochian et al., 2005).

Despite the large number of published QTLs, still only a few are actively used by breeders (for a review see Jena and Mackill, 2008; Xu and Crouch, 2008; Collins et al., 2008). This is partly due to the small effects of published QTLs, but mainly due to the absence of data showing that a given QTL is effective in different genetic backgrounds and environments. The recent success in the development of submergence-tolerant rice varieties has demonstrated the advantages of MABC and the potential impact of large-effect QTLs on food security for poor farmers. The data presented here are therefore important milestones toward the application of another large-effect rice QTL with potentially high impact in farmers’ fields.

MATERIALS AND METHODS

Plant materials for reverse transcription-polymerase chain reaction (RT-PCR)

Near-isogenic lines (NILs) with the tolerant Pup1 locus, NIL24-4 and NIL14-4, a NIL-sister line without the tolerant Pup1 locus (NIL14-4), and Nipponbare control plants were grown in P-deficient Philippine soil and the equivalent of 90-60-40 kg ha⁻¹ N-P₂O₅-KCl fertilizer was applied to the pots. No P-fertilizer was applied for the –P treatment. Plants were grown in a greenhouse with an extended (16 h) light period for about 6 weeks to prevent early flowering of the photosensitive genotypes. Root and shoot samples were collected 49 days after sowing (DAS), frozen in liquid nitrogen, and stored at -80°C until extraction of total RNA using Trizol according to the instructions from the manufacturer (Invitrogen). RNA samples were treated with RNase-free DNase (Promega) to remove any contaminating genomic DNA. cDNA synthesis at 55°C for 1 h was performed in a 20-µl reaction
containing 5 μg DNase-treated total RNA, 2.5 μM oligo dT, 0.5 mM dNTP mix, 0.01 M DTT, 1x first-strand buffer, and 200 U of Superscript II RT (Invitrogen). PCR was carried out in a total volume of 20 μl containing 0.5 μl cDNA, gene-specific primers (0.2 μM each; sequences are provided in Table 1); 1x PCR buffer, 0.5 mM dNTP mix, and 1 U Taq DNA polymerase (Beijing SBS Genetech Co. Ltd.). The PCR cycle settings were 94°C for 5 min, followed by 33 or 40 cycles of 94°C for 15 sec, 55°C for 30 sec, 30°C for 30 sec, and a final extension at 72°C for 10 min. The cytosolic glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene was used as a control for successful amplification and absence of genomic DNA (for: 5´−GCAGGAACCCTGAGGAGATC−3´, rev: 5´−TTCCCCCTCCAGTCTTGCT−3´). RT-PCR products were separated by agarose electrophoresis and stained with SYBR Safe (Invitrogen).

Gene-specific marker development and Pup1 haplotyping
To develop specific molecular markers, Pup1 gene models and nucleotide sequences (AB458444.1) were adapted from Heuer et al. (2009) except for OsPupK20 and OsPupK46-1. For these genes, primer design was based on revised gene models (accession numbers will be provided). For all Kasalath gene models, nucleotide and amino acid sequences were subjected to BLASTn and BLASTp analyses at NCBI (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=4530) and MSU Rice Genome Annotation Project v.6.1 (RGAP, rice.plantbiology.msu.edu/analyses_search_blast.shtml) to identify the most similar sequences in Nipponbare and other reference genomes. For Kasalath genes with similarity to Nipponbare genes in the syntenic region on chr. 12, co-dominant markers were developed based on sequence polymorphism (INDEL, SNP) between alleles. Primers targeting specific regions in the genes were designed using Primer3 software (http://frodo.wi.mit.edu/primer3) or manually. For co-dominant markers with small size differences between alleles, allele-specific restriction sites were identified to develop cleaved amplified polymorphic site (CAPS) markers using NEB cutter V.2 software (http://tools.neb.com/NEBcutter2). For genes that were specific to Kasalath, dominant markers were designed generally targeting predicted exon regions. Specificity of all primers was re-confirmed by BLASTn at NCBI and GRAMENE (www.gramene.org/Multi/blastview) and representative PCR amplicons were sequenced at Macrogen Inc. (Korea).
Genomic DNA was extracted from leaf samples as described by Palotta et al. (2000) from plants grown under control conditions. PCR was carried out in a total volume of 20 μl containing 40 ng genomic DNA, gene-specific primers (0.2 μM each; sequences are provided in Table 2), 0.5 mM dNTP mix, and 1.5 U Taq DNA polymerase with the provided PCR buffer (Beijing SBS Genetech Co. Ltd.). The PCR cycle settings were 94°C for 5 min, followed by 35 cycles of 94°C for 30 sec, 55°C or 58°C for 45 sec, 30°C for 90 sec, and a final extension at 72°C for 10 min. Size differentiation of marker amplicons was carried out in 3.5% agarose gels or 8% polyacrylamide gels.
In total, eighty one diverse rice accessions from the rice germplasm collection and IRRI breeding lines were included in the study (Supplementary Table 1). The Pup1 near-isogenic line NIL14-4 (Wissuwa, 2005) and the Indonesian upland varieties Dodokan, Situ Bagendit, and Batur were additionally included.

Development of Pup1 breeding lines
For the development of Pup1 breeding materials, NILC443 and Kasalath were used as the Pup1 donor for crosses involving Dodokan, Batur, and Situ Bagendit. Pup1 NIL14-4 was used as the donor for the IR64-Pup1 and IR74-Pup1 populations. For backcrossing, recipient varieties were used as recurrent parents. Crosses were conducted according to standard protocols with plants grown under control conditions in a greenhouse.

Phenotyping of Pup1 breeding lines under field conditions
The IR64-Pup1 and IR74-Pup1 BC2F3 breeding lines and parents were phenotyped at the IRRI experimental station (Philippines) in the dry season of 2010 in an irrigated plot with full fertilizer treatment (equivalent of elemental N: 120 kg ha⁻¹, K₂O: 40 kg ha⁻¹, P₂O₅: 60 kg ha⁻¹). A duplicate setup was grown in parallel in P-deficient soil without application of P-fertilizer. The equivalent of 20 kg ha⁻¹ of zinc sulfate was applied to each plot. Plants were grown in rows with 12-15 plants per line with 20 cm x 20 cm spacing. Phenotypic data from three plants located in the centre of the rows were collected from three replicated plots.

Situ Bagendit, Batur, and Dodokan BC2F4 breeding lines were grown in 2010 in Sukabumi, West Java (Indonesia), under rainfed conditions in a field with a medium endogenous P level (8.3 ppm available P). Urea (250 kg ha⁻¹) and KCl (100 kg ha⁻¹) were applied to all plots. P-fertilizer (SP18: 500 kg ha⁻¹, equivalent to 39.3 kg P ha⁻¹) was applied only to +P control plots. Breeding lines were grown in three replicates and three plants from each replicate were phenotyped.

Software used for sequence and marker analyses
Comparative sequence analyses were conducted using MAFFT (http://mafft.cbrc.jp/alignment/software). Based on the Pup1 marker data, a dendrogram was developed using UPGMA (Unweighted Pair Group Method with Arithmetic Mean) cluster analysis (Sokal and Michener, 1958). The genetic distance between the analyzed rice genotypes was calculated based on the method published by Nei (1973) using Powermarker v3.25 (Liu and Muse, 2005). To calculate the r² value for the individual Pup1 markers, the software GGT 2.0 was used (Van Berloo, 2008).

Supplemental Data
The following materials are available in the online version of this article:
Supplemental Table S1. Plant materials used for Pup1 genotyping

Supplemental Figure S1. Graphical genotype of Indonesian Pup1 introgression lines

ACKNOWLEDGMENTS
We would like to thank Anthony Cruz and Ranie Ramos for their technical assistance and devotion to the project. Special thanks go to Stephan Haefele and Serafin Amarante at IRRI for the support of the project, as well as to Juan Pariasca-Tanaka at JIRCAS. This project was funded by Generation Challenge Program (GCP) grants G3008.04 and G4008.41.

LITERATURE CITED


In S Pandey et al., eds, Rice in the global economy: strategic research and policy issues for food security, IRRI, Los Baños pp 379–407


McCouch SRM, Zhao K, Wright M, Tung CW, Ebana K, Thomson MJ, Reynolds A, Wang D,


1 Hered 99: 232–236
6 Wissuwa M, Ae N (2001a) Further characterization of two QTLs that increase phosphorus uptake of rice (Oryza sativa L.) under phosphorus deficiency. Plant Soil 237: 275–286
7 Wissuwa M, Ae N (2001b) Genotypic variation for tolerance to phosphorus deficiency in rice and the potential for its exploitation in rice improvement. Plant Breed 120: 43–48
FIGURE LEGENDS

Figure 1. Pup1 gene models and position of markers.

The genomic Pup1 region of the tolerant donor Kasalath (a) is defined by the markers T5-4 and Ba76H14_7154 (Heuer et al., 2009; accession number AB458444.1) and characterized by a large insertion/deletion (INDEL, gray line) with adjacent regions partially conserved in Nipponbare (black lines). The position and name of the developed co-dominant and dominant markers are indicated. Markers K4 and K67 were not polymorphic. The targeted gene models and marker position within the models are illustrated in b-e. The putative gene OsPupK05-1 is located within an intron of gene OsPupK04-1, which is highly similar to the Nipponbare gene Os12g26290 (b). For the dirigent-like gene model OsPupK20-2 (c), two markers were developed and PCR products can be digested with Bsp1286I (open triangles) and MseI (gray triangle). The hypothetical protein gene OsPupK29-1 is partially similar to two genes in Nipponbare and the markers target exon 1, exon 2, and an intron, respectively (d). For the protein kinase OsPupK46-2, the markers amplify a region flanking the protein-kinase domain (K46-1) and the 3’ region (K46-2), respectively (e). Black boxes in the gene models indicate exons and lines indicate introns and untranslated regions. Conserved protein domains are highlighted in gray.

Figure 2. RT-PCR gene expression analysis.

The Pup1 genes indicated were analyzed by RT-PCR using RNA samples derived from roots and shoots of plants grown in P-deficient soil with (+P) and without (−P) P-fertilizer application were analyzed. Two different tolerant Nipponbare-Pup1 near-isogenic lines with the Kasalath Pup1 locus (NIL24-4 = 1+; NIL14-4 = 2+). Nipponbare (NB) and a NIL14-4 sister line without the tolerant Pup1 locus (NIL14-6 = −”) were used as intolerant controls. cDNA samples were analyzed by PCR using the cycle number indicated on the right. GAPDH was included as a control. H2O: water control.

Figure 3. Sequence comparisons of marker amplicons.

PCR amplicons of five markers were sequenced in the rice varieties indicated to determine the specificity of the primers and validate sequence polymorphisms. For OsPupK20-2, a non-synonymous SNP at +620 bp was identified that changed the amino acid sequence in exon 2
(T: Nipponbare, C: Kasalath) (a, left) and created a Bsp1286I site in Kasalath-type alleles but not in non-Kasalath (N-type) alleles. In the adjacent intron, one SNP (+660 bp) and one INDEL (+665−671 bp) were identified. The SNP creates an MseI restriction site (a, right). For OsPupK29-1, a polymorphic region with one INDEL and four SNPs is amplified by the marker K29-1 (b). Another INDEL in OsPupK29-1 is targeted by the marker K29-3 (c). PCR products of the dominant marker K46-2 (stop codon is highlighted in gray) were sequenced in three varieties that possess the OsPupK46 gene (d). Only polymorphic regions of the sequence regions are shown.

**Figure 4. PCR amplicons of Pup1 gene-specific markers.**
Seven representative rice varieties were genotyped with nine co-dominant (top panels) and eleven dominant Pup1-specific markers (bottom panels). The markers target a total of twelve Kasalath Pup1 gene models. PCR products K20-1 and K20-2 were digested with MseI (K20-1Mse) and Bsp1286I (K20-2Bsp), respectively, to enhance resolution. K: Kasalath; N: Nipponbare; B: Batur; D: Dodokan; SB: Situ Bagendit; H2O: water control. The size of the DNA fragments in Kasalath and Nipponbare (K/N) and unspecific amplicons that sometimes occur is indicated at the right. The absence of PCR amplicons with dominant markers is indicated as “none”.

**Figure 5. Pup1 haplotype in diverse rice genotypes.**
Eighty-one diverse rice varieties and breeding lines were genotyped using the Pup1-specific markers indicated. Based on the Pup1 core marker set (highlighted with broken lines; see text for details), a UPGMA cluster analysis was conducted using Powermarker v3.25 (Sokal and Michener 1958; Liu and Muse 2005). Three main groups (I, II, III) with different allelic composition at the Pup1 locus were identified. The different allele calls (Kasalath, non-Kasalath, and heterozygous) are color-coded as indicated in the legend. The rice genotypes analyzed are adapted to different rice agro-ecosystems (upland (UL): aerobic, unfavorable, and drought-prone; lowland/irrigated (LL): anaerobic, generally favorable conditions) as shown in the panel at the right. For details on the genotypes analyzed see text and Supplemental Table 1.

**Figure 6. Pup1 core markers.**
A GGT 2.0 analysis (van Berloo 2008) was conducted with the marker data of the eighty one genotypes analyzed with the indicated Pup1-specific markers. The r2 values of the individual markers are represented as a heat plot (black color: $r^2 = 1$). The approximate physical location of the markers is indicated on the black line, which represents the Kasalath Pup1 locus. Two haploblocks can be identified that define the core marker set (bold lines). The marker K29-2 is not associated with the other two markers for the OsPupK29-1 gene (broken line; see text for details).

Figure 7. Field evaluation and selection of IR64-Pup1 and IR74-Pup1 breeding lines.
Sister lines with (+Pup1) and without (−Pup1) the tolerant Pup1 locus and control plants were grown under irrigated field conditions in P-deficient soil (−P) and in a P-fertilized parallel control plot in three replicates. Phenotypic data from the BC$_2$F$_3$ populations were collected in 2010 (a). For the final selection of IR64-Pup1 and IR74-Pup1 plants, contrasting BC$_2$F$_4$ sister lines were genotyped with selected Pup1-core markers (b, top). Representative plants are shown in (b). Data were analyzed by paired t-test (95%). Significance levels: *: 0.05>p≥0.01, **: 0.01>p≥0.001, ***: 0.001>p. No asterisk: not significant.

Figure 8. Field evaluation of three Indonesian Pup1 breeding lines
Advanced generations (BC$_3$F$_4$) of Pup1 breeding populations of three Indonesian upland varieties were grown in a field in Sukabumi (West Java, Indonesia) under rainfed conditions in 2010. The data show average grain yield per plant obtained with (P+; Y-axis) and without (P−; X-axis) P-fertilizer application. Grain yield of the Pup1 recipient parents is highlighted in the graphs. Standard errors are shown as horizontal and vertical lines. The gray bar on top of the graphs signifies the Pup1 locus of the recipient parent. Kasalath alleles detected with the indicated Pup1 markers are shown as black lines.

Supplemental Figure 1. Graphical genotype of Indonesian Pup1 introgression lines
Pup1 breeding lines of three Indonesian rice varieties (B=Batur, S=Situ Bagendit, D=Dodokan) were genotyped with a set of polymorphic SSR markers (47 to 61 SSRs depending on the population). The schematic overview of the data represents the genotype of several plants analyzed from the different breeding populations. The presence of the tolerant Pup1 locus was determined using Pup1 foreground markers.
<table>
<thead>
<tr>
<th>gene name</th>
<th>annotation</th>
<th>TIGR6 BLASTp</th>
<th>E value</th>
<th>RT-PCR primers</th>
<th>gene expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>OsPupK01-1</td>
<td>Hypothetical protein</td>
<td>No significant similarity; located in intergenic region between LOC_Os12g26260 and LOC_Os12g26270</td>
<td>4.10E-07</td>
<td>5'-ACACGTACAATGTCCACCGGG-3' 5'-ATGTCAACGCGACGTTACG-3'</td>
<td>No</td>
</tr>
<tr>
<td>OsPupK04-1</td>
<td>Highly similar to alpha-DOX2 (LOC_Os12g26290)</td>
<td>Alpha-DOX2, putative, expressed</td>
<td>0</td>
<td>5'-GGGATATCAAGCTTGTGGTG-3' 5'-GAATGCTGTTTCGCTTATGG-3'</td>
<td>Yes</td>
</tr>
<tr>
<td>OsPupK05-1</td>
<td>Located within OsPupK04, expressed</td>
<td>Transposon protein, putative, CACTA, En/Spm sub-class</td>
<td>0.97</td>
<td>5'-AGTACAGTCCCGGCTACATC-3' R:CCGAGATCTGGTCCTCAATA-3'</td>
<td>Yes</td>
</tr>
<tr>
<td>OsPupK20-2</td>
<td>Highly similar to dirigent protein, putative, expressed (LOC_Os12g26380)</td>
<td>Dirigent, putative, expressed</td>
<td>3.50E-100</td>
<td>5'-CTGGACCTTGACCCCAATGTA-3' 5'-TCTGATGGAGTTCGAGGT-3'</td>
<td>Yes</td>
</tr>
<tr>
<td>OsPupK29-1</td>
<td>Similar to expressed proteins LOC_Os12g26390 and LOC_Os12g26410</td>
<td>LOC_Os12g26390, expressed LOC_Os12g26410, expressed</td>
<td>0</td>
<td>5'-CCATGCATCCAATTCTTGT-3' 5'-ATGAGCCCGATTAGAATG-3'</td>
<td>Yes</td>
</tr>
<tr>
<td>OsPupK46-2</td>
<td>Most similar to protein kinase genes (LOC_Os01g49580, LOC_Os01g49614, LOC_Os08g24630)</td>
<td>LOC_Os01g49580, expressed LOC_Os01g49614, expressed LOC_Os08g24630, expressed</td>
<td>1E-93</td>
<td>5'-AGGAAAGATGGCTTGTGCTTG-3' 5'-TCCACAAACAGTTGCTGTC-3'</td>
<td>Yes</td>
</tr>
<tr>
<td>OsPupK53-1</td>
<td>Insignificant similarity to zinc finger CCCH type family protein, expressed (LOC_Os05g10670)</td>
<td>Zinc finger CCCH type family protein, putative, expressed</td>
<td>6.40E-37</td>
<td>5'-TTCAGGTTCCTTCGGCGACAC-3' 5'-TAGACGAGACGTCGTGAC-3'</td>
<td>No</td>
</tr>
<tr>
<td>OsPupK67-1</td>
<td>Highly similar to aspartic proteinase precursor nepenthesin-1 precursor (LOC_Os12g262470)</td>
<td>Aspartic proteinase nepenthesin-1 precursor, putative, expressed</td>
<td>9.50E-211</td>
<td>5'-CACGAGCTATTTTGTGGTGGG-3' 5'-GGTACGAACATTTGCGGAC-3'</td>
<td>No</td>
</tr>
</tbody>
</table>
Table 2. Primer sequences of gene-specific Pup1 markers.

<table>
<thead>
<tr>
<th>marker name</th>
<th>Pup1 gene model</th>
<th>marker type</th>
<th>physical location (bp)</th>
<th>Nipponbare chromosome 12</th>
<th>band size (bp)</th>
<th>primer sequence (5’–3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td>OsPupK01-1</td>
<td>co-dominant</td>
<td>Kasalath AB458444.1</td>
<td>96175–96299</td>
<td>125 / 122</td>
<td>F: AGTCTGGAATGGCAACACTCTGCG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TGCTAGCTCATTTGCCCTTACGC</td>
</tr>
<tr>
<td>K5</td>
<td>OsPupK05-1</td>
<td>co-dominant</td>
<td>Nipponbare chromosome 12</td>
<td>272 / 280</td>
<td></td>
<td>F: ATTCAGAATGCAGGCGAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCCCTGTAACATGCTGCG</td>
</tr>
<tr>
<td>K20−1</td>
<td>OsPupK20-2</td>
<td>co-dominant</td>
<td>Kasalath</td>
<td>15410254–15410496</td>
<td>240 / 243</td>
<td>F: TCAGGATGGAATCACTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TGCTCAACACACACCTG</td>
</tr>
<tr>
<td>K20−1</td>
<td>OsPupK20-2</td>
<td>CAPS</td>
<td>(Mse1)</td>
<td>201 / 243</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K20−2</td>
<td>OsPupK20-2</td>
<td>co-dominant</td>
<td>Kasalath</td>
<td>15409652–15410981</td>
<td>982 / 995</td>
<td>F: TCAAAAAATTTCTCTAGTCTAAGCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TTGGCTGATCACTCTGC</td>
</tr>
<tr>
<td>K20−2</td>
<td>OsPupK20-2</td>
<td>CAPS</td>
<td>(Bsp1286I)</td>
<td>K: 231+349+402</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OsPupK20-2</td>
<td>CAPS</td>
<td>(Bsp1286I)</td>
<td>N: 413+582</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K29−1</td>
<td>OsPupK29-1</td>
<td>co-dominant</td>
<td>Kasalath</td>
<td>205067–205287</td>
<td>212 / 206</td>
<td>F: ATGGCCAGCGGATGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: GTCAGTAGTCACCACGAG</td>
</tr>
<tr>
<td>K29−2</td>
<td>OsPupK29-1</td>
<td>co-dominant</td>
<td>Kasalath</td>
<td>204398–204616</td>
<td>291 / 212</td>
<td>F: CCCGTCTGCTTCTACCTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCCCCGTCAAGCAAAATCT</td>
</tr>
<tr>
<td>K29−3</td>
<td>OsPupK29-1</td>
<td>co-dominant</td>
<td>Kasalath</td>
<td>202698–202933</td>
<td>236 / 248</td>
<td>F: TTGGCTCAAGATGCCTGCTATG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCTCAAGATGCAACACTG</td>
</tr>
<tr>
<td>K41</td>
<td>OsPupK41-1</td>
<td>dominant</td>
<td>Kasalath</td>
<td>262050–262431</td>
<td>382 / –</td>
<td>F: TGATGATCTCTAGGACAGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCAAGGTTGCTCTGCTGTA</td>
</tr>
<tr>
<td>K42</td>
<td>OsPupK42-1</td>
<td>dominant</td>
<td>Kasalath</td>
<td>267154–268071</td>
<td>918 / –</td>
<td>F: CCCGAGATGTCACAGAAAGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: AGTGATGTCGGCTCTGCTGAT</td>
</tr>
<tr>
<td>K43</td>
<td>OsPupK43-1</td>
<td>dominant</td>
<td>Kasalath</td>
<td>268590–269501</td>
<td>912 / –</td>
<td>F: TCAGCAGCATAGCAAGATT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: GGGGAAAGACAGATGCAATG</td>
</tr>
<tr>
<td>K45</td>
<td>OsPupK45-1</td>
<td>dominant</td>
<td>Kasalath</td>
<td>274072–274344</td>
<td>276 / –</td>
<td>F: TGATGATCTCTAGGACAGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCAAGGTTGCTCTGCTGTA</td>
</tr>
<tr>
<td>K46−1</td>
<td>OsPupK46-2</td>
<td>dominant</td>
<td>Kasalath</td>
<td>275710–276232</td>
<td>523 / –</td>
<td>F: TGAGATAAGCCGCTCAAGATCTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: AAGGACACCACTCTCAGAAC</td>
</tr>
<tr>
<td>K46−2</td>
<td>OsPupK46-2</td>
<td>dominant</td>
<td>Kasalath</td>
<td>276371–276597</td>
<td>227 / –</td>
<td>F: AGGAGATGCTGGCTGGCTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCTACACAAACAGTGTGCTC</td>
</tr>
<tr>
<td>K48</td>
<td>OsPupK48-1</td>
<td>dominant</td>
<td>Kasalath</td>
<td>282795–283640</td>
<td>847 / –</td>
<td>F: AGGAGATGCTGGCTGGCTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCTACACAAACAGTGTGCTC</td>
</tr>
<tr>
<td>K52</td>
<td>OsPupK52-1</td>
<td>dominant</td>
<td>Kasalath</td>
<td>300870–301374</td>
<td>505 / –</td>
<td>F: AAGGACACCACTCTCAGAAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCTACACAAACAGTGTGCTC</td>
</tr>
<tr>
<td>K59</td>
<td>OsPupK59-1</td>
<td>dominant</td>
<td>Kasalath</td>
<td>324843–325392</td>
<td>550 / –</td>
<td>F: GGGGAAAGACAGATGCAATG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R: TCTACACAAACAGTGTGCTC</td>
</tr>
</tbody>
</table>

F: forward primer; R: reverse primer
Figure 1. *Pup1* gene models and position of markers.
The genomic *Pup1* region of the tolerant donor Kasalath (a) is defined by the markers T5-4 and Ba76H14_7154 (Heuer et al., 2009; accession number AB458444.1) and characterized by a large insertion/deletion (INDEL, gray line) with adjacent regions partially conserved in Nipponbare (black lines). The position and name of the developed co-dominant and dominant markers are indicated. Markers K4 and K67 were not polymorphic. The targeted gene models and marker position within the models are illustrated in b-e. The putative gene *OsPupK05-1* is located within an intron of gene *OsPupK04-1*, which is highly similar to the Nipponbare gene *Os12g26290* (b). For the dirigent-like gene model *OsPupK20-2* (c), two markers were developed and PCR products can be digested with *Bsp*1286I (open triangles) and *Mse*I (gray triangle). The hypothetical protein gene *OsPupK29-1* is partially similar to two genes in Nipponbare and the markers target exon 1, exon 2, and an intron, respectively (d). For the protein kinase *OsPupK46-2*, the markers amplify a region flanking the protein-kinase domain (K46-1) and the 3′-region (K46-2), respectively (e). Black boxes in the gene models indicate exons and lines indicate introns and untranslated regions. Conserved protein domains are highlighted in gray.
Figure 2. RT-PCR gene expression analysis.
The Pup1 genes indicated were analyzed by RT-PCR using RNA samples derived from roots and shoots of plants grown in P-deficient soil with (+P) and without (−P) P-fertilizer application were analyzed. Two different tolerant Nipponbare-Pup1 near-isogenic lines with the Kasalath Pup1 locus (NIL24-4 = 1+; NIL14-4 = 2+). Nipponbare (NB) and a NIL14-4 sister line without the tolerant Pup1 locus (NIL14-6 = “−”) were used as intolerant controls. cDNA samples were analyzed by PCR using the cycle number indicated on the right. GAPDH was included as a control. H2O: water control.
Figure 3
Figure 3. Sequence comparisons of marker amplicons.

PCR amplicons of five markers were sequenced in the rice varieties indicated to determine the specificity of the primers and validate sequence polymorphisms. For OsPupK20-2, a non-synonymous SNP at +620 bp was identified that changed the amino acid sequence in exon 2 (T: Nipponbare, C: Kasalath) (a, left) and created a Bsp1286I site in Kasalath-type alleles but not in non-Kasalath (N-type) alleles. In the adjacent intron, one SNP (+660 bp) and one INDEL (+665–671 bp) were identified. The SNP creates an Msel restriction site (a, right). For OsPupK29-1, a polymorphic region with one INDEL and four SNPs is amplified by the marker K29-1 (b). Another INDEL in OsPupK29-1 is targeted by the marker K29-3 (c). PCR products of the dominant marker K46-2 (stop codon is highlighted in gray) were sequenced in three varieties that possess the OsPupK46 gene (d). Only polymorphic regions of the sequence regions are shown.
Figure 4. PCR amplicons of Pup1 gene-specific markers.
Seven representative rice varieties were genotyped with nine co-dominant (top panels) and eleven dominant Pup1-specific markers (bottom panels). The markers target a total of twelve Kasalath Pup1 gene models. PCR products K20-1 and K20-2 were digested with MseI (K20-1\textsubscript{Mse}) and Bsp1286I (K20-2\textsubscript{Bsp}), respectively, to enhance resolution. K: Kasalath; N: Nipponbare; B: Batur; D: Dodokan; SB: Situ Bagendit; H\textsubscript{2}O: water control. The size of the DNA fragments in Kasalath and Nipponbare (K/N) and unspecific amplicons that sometimes occur is indicated at the right. The absence of PCR amplicons with dominant markers is indicated as “none”.
Figure 5. *Pup1* haplotype in diverse rice genotypes.

Eighty-one diverse rice varieties and breeding lines were genotyped using the *Pup1*-specific markers indicated. Based on the *Pup1* core marker set (highlighted with broken lines; see text for details), a UPGMA cluster analysis was conducted using Powermarker v3.25 (Sokal and Michener 1958; Liu and Muse 2005). Three main groups (I, II, III) with different allelic composition at the *Pup1* locus were identified. The different allele calls (Kasalath, non-Kasalath, and heterozygous) are color-coded as indicated in the legend. The rice genotypes analyzed are adapted to different rice agro-ecosystems (upland (UL): aerobic, unfavorable, and drought-prone; lowland/irrigated (LL): anaerobic, generally favorable conditions) as shown in the panel at the right. For details on the genotypes analyzed see text and Supplemental Table 1.
Figure 6. *Pup1* core markers.
A GGT 2.0 analysis (Van Berloo 2008) was conducted with the marker data of the eighty one genotypes analyzed with the indicated *Pup1*-specific markers. The $r^2$ values of the individual markers are represented as a heat plot (black color: $r^2 = 1$). The approximate physical location of the markers is indicated on the black line, which represents the Kasalath *Pup1* locus. Two haploblocks can be identified that define the core marker set (bold lines). The marker K29-2 is not associated with the other two markers for the *OsPupK29-1* gene (broken line; see text for details).
Figure 7. Field evaluation and selection of IR64-Pup1 and IR74-Pup1 breeding lines.
Sister lines with (+Pup1) and without (−Pup1) the tolerant Pup1 locus and control plants were grown under irrigated field conditions in P-deficient soil (−P) and in a P-fertilized parallel control plot in three replicates. Phenotypic data from the BC2F3 populations were collected in 2010 (a). For the final selection of IR64-Pup1 and IR74-Pup1 plants, contrasting BC2F4 sister lines were genotyped with selected Pup1-core markers (b, top). Representative plants are shown in (b). Data were analyzed by paired t-test (95%). Significance levels: *: 0.05>p≥0.01, **: 0.01>p≥0.001, ***: 0.001>p. No asterisk: not significant.
Advanced generations (BC$_2$F$_4$) of $Pup1$ breeding populations of three Indonesian upland varieties were grown in a field in Sukabumi (West Java, Indonesia) under rainfed conditions in 2010. The data show average grain yield per plant obtained with (P$+$; Y-axis) and without (P$-$; X-axis) P-fertilizer application. Grain yield of the $Pup1$ recipient parents is highlighted in the graphs. Standard errors are shown as horizontal and vertical lines. The gray bar on top of the graphs signifies the $Pup1$ locus of the recipient parent. Kasalath alleles detected with the indicated $Pup1$ markers are shown as black lines.
Supplemental Figure 1. Graphical genotype of Indonesian Pup1 introgression lines

Pup1 breeding lines of three Indonesian rice varieties (B=Batur, S=Situ Bagendit, D=Dodokan) were genotyped with a set of polymorphic SSR markers (47 to 61 SSRs depending on the population). The schematic overview of the data represents the genotype of several plants analyzed from the different breeding populations. The presence of the tolerant Pup1 locus was determined using Pup1 foreground markers.