scientific reports

Check for updates

OPEN

An atypical class of non-coding small RNAs is produced in rice leaves upon bacterial infection

Ganna Reshetnyak[1,7], Jonathan M. Jacobs[1,5,6,7], Florence Auguy[1], Coline Sciallano[1], Lisa Claude[1], Clemence Medina[1], Alvaro L. Perez-Quintero[1], Aurore Comte[1], Emilie Thomas[1], Adam Bogdanove[2], Ralf Koebnik[1], Boris Szurek[1], Anne Dievart[3,4], Christophe Brugidou[1], Severine Lacombe[1] & Sebastien Cunnac[1]✉

Non-coding small RNAs (sRNA) act as mediators of gene silencing and regulate plant growth, development and stress responses. Early insights into plant sRNAs established a role in antiviral defense and they are now extensively studied across plant–microbe interactions. Here, sRNA sequencing discovered a class of sRNA in rice (*Oryza sativa*) specifically associated with foliar diseases caused by *Xanthomonas oryzae* bacteria. *Xanthomonas*-induced small RNAs (xisRNAs) loci were distinctively upregulated in response to diverse virulent strains at an early stage of infection producing a single duplex of 20–22 nt sRNAs. xisRNAs production was dependent on the Type III secretion system, a major bacterial virulence factor for host colonization. xisRNA loci overlap with annotated transcripts sequences, with about half of them encoding protein kinase domain proteins. A number of the corresponding rice *cis*-genes have documented functions in immune signaling and xisRNA loci predominantly coincide with the coding sequence of a conserved kinase motif. xisRNAs exhibit features of small interfering RNAs and their biosynthesis depend on canonical components *OsDCL1* and *OsHEN1*. xisRNA induction possibly mediates post-transcriptional gene silencing but they do not broadly suppress *cis*-genes expression on the basis of mRNA-seq data. Overall, our results identify a group of unusual sRNAs with a potential role in plant–microbe interactions.

RNA silencing represents a fundamental mechanism of gene regulation in eukaryotes and orchestrates a wide range of basic cellular and organismal processes. In plants, RNA silencing coordinates major biological functions such as maintenance of genome integrity and inheritance, development and environmental responses[1,2]. Non-coding small RNAs (sRNAs) determine the specific ty of RNA silencing. sRNAs are typically 20–24 nucleotides (nt) in size and derive from a double stranded RNA (dsRNA) precursor that is diced into a single or multiple distinct sRNA duplexes by Dicer-like proteins (DCL). In plants, the released sRNAs duplexes are universally methylated on the 2′-OH of the 3′ terminal ribose of each strand by HEN1[3]. The mature sRNA duplex 'guide' strand is selectively loaded into an ARGONAUTE (AGO) protein component of RNA-Induced gene Silencing Complexes to mediate either transcriptional gene silencing (TGS) or post-transcriptional gene silencing (PTGS) according to nucleic acids sequence complementarity between the sRNAs and their target RNA[4]. TGS typically depends on 24 nt heterochromatic small interfering RNAs (het-siRNAs) and involves DNA methylation and chromatin reprogramming. PTGS, on the other hand, guided by 21 nt micro and siRNAs, leads to mRNA cleavage and mRNA translational inhibition[1].

Several major classes of endogenous plant regulatory sRNAs have been defi ed based on the specifics of their biogenesis pathways and their mode of action[3,5]: Micro RNAs (miRNAs) are arguably the best characterized class of sRNAs. Hairpins of *MIRNA* gene transcripts (pri-miRNA) are cleaved and diced by the plant DCL1 protein to generate a single miRNA duplex. Other hairpin small interfering RNAs (hp-siRNAs) also derive from a single-stranded RNA with inverted repeats that folds into a hairpin structure but whose processing does not follow the

[1]PHIM Plant Health Institute, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France. [2]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA. [3]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, 34398 Montpellier, France. [4]CIRAD, UMR AGAP Institut, 34398 Montpellier, France. [5]Present address: Department of Plant Pathology, The Ohio State University, Columbus, OH 43201, USA. [6]Present address: Infectious Disease Institute, The Ohio State University, Columbus, OH 43201, USA. [7]These authors contributed equally: Ganna Reshetnyak and Jonathan M. Jacobs. ✉email: Sebastien.Cunnac@ird.fr

precise pattern associated with miRNAs. The remaining main classes of sRNAs pertain to siRNAs whose dsRNA precursor is made up of two complementary RNA molecules. With the exception of Natural antisense transcript (NAT) siRNAs (NAT-siRNAs) that are produced from two independently transcribed but complementary precursors, siRNAs biogenesis usually involves an RNA-dependent RNA polymerase (RDR) catalyzed reaction. For example, phased secondary siRNAs (phasiRNAs) and *trans*-acting siRNAs are generated following the conversion of a single-stranded miRNA cleavage product to a double stranded precursor by RDR6[6].

Even if the role of sRNAs in plant biotic interactions is an area of intensive research, it is well-established that multiple plant RNA silencing pathways are also critical for plant immune responses to bacterial, fungal or oomycete pathogens[7]. In nature, healthy plants rely on their innate immune system to perceive and fi ht off pathogens. PAMP-triggered immunity (PTI) represents a fi st line of defense relying on the detection of Pathogen-Associated Molecular Patterns (PAMPs) such as bacterial flagellin or fungal chitin. Plasma membrane localized receptor-kinases and receptor-like proteins perceive PAMPs and trigger phosphorylation cascades that ultimately signal the deployment of defense mechanisms to hinder pathogen invasion[8]. Plant receptor-kinases (a.k.a. RLKs for Receptor-Like Kinases) have a monophyletic origin and have been classifi d into the RLK/Pelle kinase family[9,10]. Th s family also includes Receptor-Like Cytoplasmic Kinases (RLCKs) that lack the transmembrane and extracellular domains and function intracellularly as substrates of receptor-kinase complexes for signal transduction[11].

Several plant miRNAs whose expression is modulated in response to PAMPs were shown to be important to efficiently combat pathogen attacks downstream of the perception of PAMP signals[7]. Adapted pathogens have evolved mechanisms to counteract eukaryotic immune responses. For instance, many Gram-negative bacterial pathogens employ a Type Three Secretion System (T3SS) to inject effector proteins directly into eukaryotic host cells to modulate their physiology and suppress defenses[12]. Similar to Viral Suppressors of RNA silencing plant-pathogenic bacteria and oomycetes have been shown to use their repertoire of virulence effector proteins to dampen sRNA silencing activity and compromise plant immunity[13]. As a counter-defense mechanism, plants evolved resistance (R) proteins that recognize pathogen virulence factors through a mechanism called Effector-Triggered Immunity (ETI) which restricts pathogen growth and infection by triggering programmed cell death[14]. The role of sRNA as positive mediators of immunity is not limited to the PTI layer, and both TGS and PTGS also regulate the ETI layer[7].

Rice (*Oryza sativa*) is a major crop in global agriculture and stands as an economic and nutritional cornerstone in the developing world. Rice possesses RNA silencing pathways that are both analogous to *A. thaliana* and idiosyncratic[15,16]. *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) and *X. oryzae* pv. *oryzicola* (*Xoc*) are the pathogenic agents that cause major diseases of rice: bacterial leaf blight (BLB) and bacterial leaf streak (BLS), respectively. Even though *Xoo* and *Xoc* belong to the same *X. oryzae* (*Xo*) species, they form distinct phylogenetic groups that are further delimited as a function of their geographic origin and have contrasted infection strategies*: Xoo* enters through hydathodes and invades the xylem whereas *Xoc* uses stomata to colonize the plant mesophyll[17]. *Xo* strains employ virulence factors, including the T3SS, that contribute to disease development. Mutants lacking the T3SS are unable to translocate effector proteins into host cells and are non-virulent. These bacteria possess over 32 effectors, including a unique class of Transcription Activator-Like Effector (TALE). TALEs activate protein-encoding gene expression and reprogram the transcriptome of the host's cells by mimicking eukaryotic transcription factors[12].

Recently, the discovery of trans-kingdom RNAi or *trans*-species small RNA silencing, has shed a new light on the role of sRNAs as mobile effectors modulating the outcome of plant biotic interactions[7,18]. Pathogen effectors acting as Suppressors of RNA silencing and pathogen *trans*-species sRNAs are both delivered into the plant cell to prevent defense. However, while the former proteinaceous effectors suppress silencing, the latter ribonucleic effectors instead, exploit it. There is still limited evidence in support of a third virulence strategy where proteinaceous effectors would leverage RNA silencing to promote immune suppression and susceptibility. Recently however, a *Xoo* T3SS effector-induced rice miRNA was shown to exert PTGS on a *R* gene conferring quantitative resistance to BLB[19].

With a few exceptions[7], the available functional data on the role of endogenous plant sRNAs in regulating the disease process has focused almost exclusively on miRNAs and their role in promoting immunity. For example, in rice, the expression of multiple miRNAs changes in response to the blast fungus *Magnaporthe oryzae*[20] and *Xoo*[21,22]. Here, we aimed at examining all sRNAs associated with the disease process at a genome-wide scale and used an unbiased small RNA deep-sequencing data analysis approach on rice inoculated with *Xo* strains. Unusual 20–22 nt small RNAs were found to be generated from annotated protein coding loci in the *Xo* infected leaf in a T3SS-dependant manner and required the archetypal sRNA biogenesis pathways components *OsDCL1* and *OsHEN1*. Several xisRNAs mapped to *cis*-genes previously implicated in immune signaling, but did not appear to generally operate through PTGS of their *cis*-gene. Th s study provides insight into an intriguing group of sRNAs that may participate in gene silencing for plant disease development.

Results

Small RNA sequencing of rice leaves infected with Xoo BAI3 identifies sRNA loci that are upregulated in a T3SS-dependant manner. To gain genome-wide insight on host sRNA species that may be involved in the disease process, we generated sRNA Illumina sequencing data from RNA collected from the Japonica *O. sativa* rice cultivar Nipponbare leaves confronted by pathogenic Xoo. Briefly, sRNAs were analyzed from leaf samples taken 24 h post-infiltration (hpi) with the wild-type African Xoo reference strain BAI3, a non-virulent T3SS mutant, which is unable to inject virulence effectors (herein referred to as BAI3H) or a water control (H2O). We inferred potential sRNA generating loci based on sRNA reads coverage profile along the Nipponbare genomic sequences, and tested for differential accumulation of sRNA reads in treatment comparisons as summarized in Supplementary Fig. S1. Because the T3SS is an essential component of infection, we aimed

to defi e sRNA loci that were induced in a T3SS-dependent fashion. We required that Xanthomonas-induced small RNA (xisRNA) loci be signifi antly induced in both the BAI3 vs H2O and the BAI3 vs BAI3H comparisons but not in the BAI3H vs H20 comparison where no T3SS effect is expected to occur. To ensure that xisRNA loci would not correspond to repetitive sequences and would be enriched for potential regulatory non-coding sRNAs, we further required that ≥ 75% of the reads at these loci mapped uniquely in the genome and that ≥ 50% of the reads had a size between 20 and 24 nt. Applying these criteria to the set of 341 Differentially Expressed (DE) sRNAs loci resulted in a total of 64 xisRNA loci.

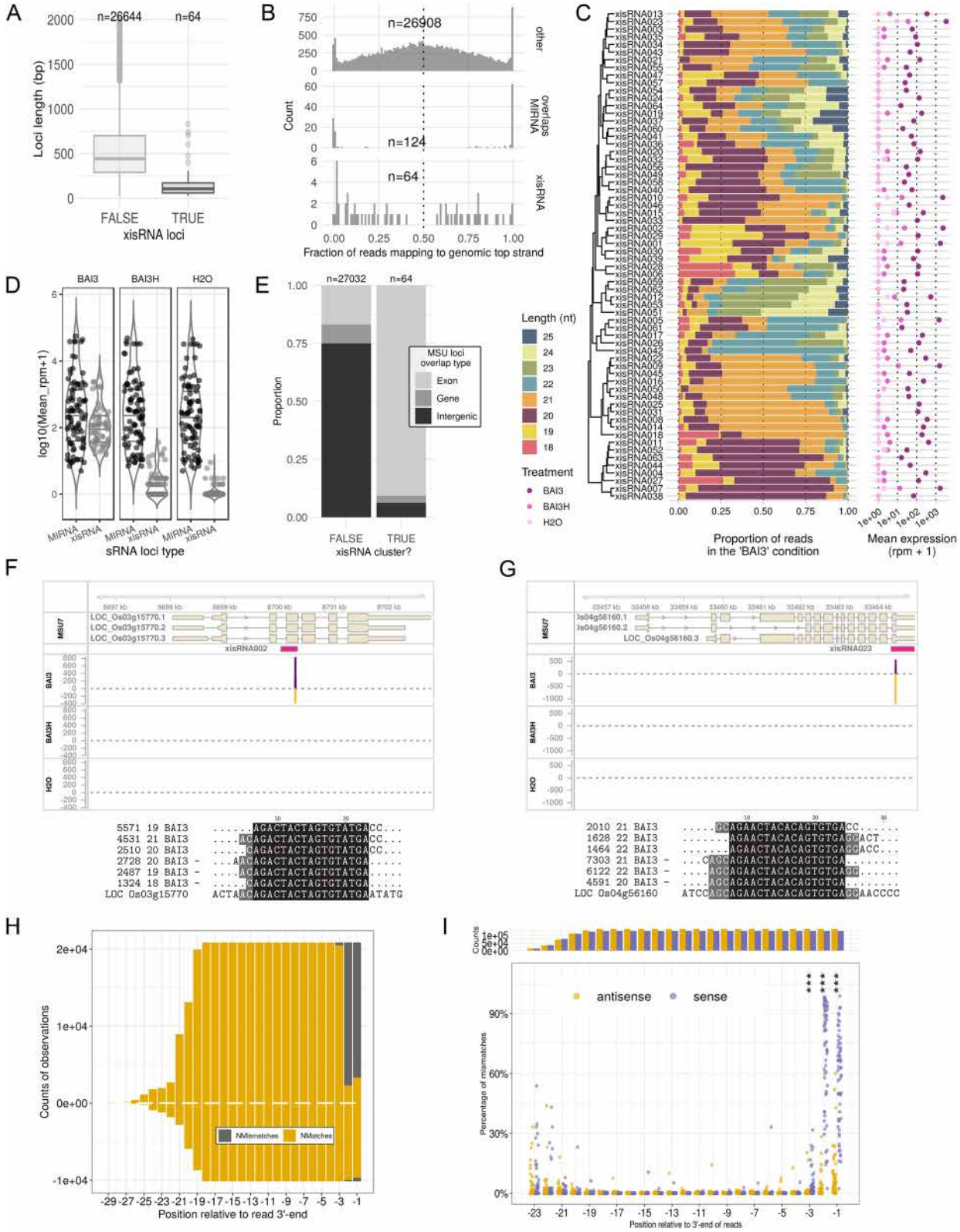
To characterize xisRNA loci, we fi st analyzed the distribution of xisRNA loci length (Fig. 1A) and found that with a median of 106 base pairs (bp), their length is signifi antly different from other experimentally defi ed sRNA loci (Two-sided Wilcoxon rank sum test: W = 1,547,014, p = 1.02e−27). In addition, inspecting the distribution of the proportion of reads mapping to the top genomic strand (Fig. 1B), revealed that, as expected, a vast majority of the experimental sRNA loci overlapping pri-miRNA genomic intervals have mapping strand ratios of 0 or 1. In contrast, the bulk of xisRNA and other experimental sRNA loci are composed of reads partitions that map to either strands of their cognate genomic locus, reminiscent of siRNA loci. xisRNA loci expression in non-inducing conditions is extremely low or even null (Fig. 1C,D). In contrast, upon BAI3 infection, xisRNA loci reads are rather abundant. Even if overall lower, their normalized expression is not at odd with pri-miRNA loci levels (Fig. 1D). The most highly expressed loci, namely, xisRNA023, 01, 02, 05, 07, 09 and 10 reach several thousand rpm in leaves inoculated with virulent BAI3 bacteria (Fig. 1C). Regarding read sizes distributions, no major global bias in *MIRNA* genes-derived reads sizes and coverage patterns were detected (Supplementary Fig. S2). While this suggests that read populations in our dataset are not anomalously biased, read length distribution for xisRNA loci was broader than for *MIRNA* genes (Fig. 1C). Furthermore, with a few exceptions (e.g. xiRNA059, 62, 12, 53 and 51), xisRNA read sizes ranged overall from 20 to 22 nt, which is comparable to rice siRNAs.

In order to relate xisRNA loci with existing rice miRNA annotation, we asked whether they overlap with *O. sativa MIRNA* gene loci annotated in MirBase but found no match. Consistent with this observation, none of the xisRNA loci was classified by ShortStack[23] as a possible MIRNA locus (category 'N14', 'N15' or 'Y'), suggesting that their features are incompatible with their being previously undocumented MIRNA genes. We also examined the overlap of experimental sRNA loci with annotated genes and transcripts in the Rice Genome Annotation Project (MSU7 annotation)[24]. As illustrated in Fig. 1E, the distribution of the overlap type with MSU loci is different for xisRNA loci as compared to other experimental sRNA loci. Notably, 58 out of 64 xisRNA loci (91%) overlap with an annotated MSU exon which is significantly greater than for other experimental sRNA loci (One-sided Fisher's Exact test adjusted p-value = 1.57e−37).

In summary, xisRNA loci are generally short and do not correspond to documented or putative *MIRNA* genes but a majority of them overlap with exons of annotated rice genes. Analogous to siRNAs, these loci are composed of 20–22 nucleotides-long reads mapping to both strands of the genomic sequence and that can also accumulate to substantial levels in the context of a T3SS-dependent *Xoo* BAI3 infection.

BAI3 xisRNAs loci generate single sRNA duplexes. A majority of xisRNA loci overlap with exons of annotated protein-coding genes. We therefore examined the coverage profile of mapped xisRNAs signatures along the Nipponbare genome sequence in regions encompassing annotated MSU loci (see individual coverage plots in our dataverse). For example, highly expressed and typical xisRNA002 and xisRNA023 loci display a T3SS-dependant ~ 25 bp-wide coverage peak comprised of reads mapping on an annotated exon in either sense or antisense orientation (Fig. 1F,G). To extent this observation, we inspected read mappings at these genomic locations (Supplementary Fig. S3) and created multiple alignments of the most abundant unique read signatures for each strand of the xisRNA loci (Fig. 1F,G). On these alignments, the pattern of complementarity between sense and anti-sense xisRNA major reads, with ~ 2 nt 3′-overhangs, is evocative of guide-passenger strands of sRNAs duplexes resulting from DCL processing. Another discernible feature is that the last two 3′-end nucleotides of major reads in the sense orientation relative to the annotated *cis*-transcript do not match the transcript sequence. To evaluate whether the presence of untemplated 3′-most nucleotides is a general feature of xisRNA sense reads, we computed the counts of matches and mismatches relative to the genomic sequence for individual positions along read sequences in each xisRNA loci and contrasted them as a function of read orientation (exemplifi d for xisRNA002 in Fig. 1H). Th s further enabled to derive the position- and strand-specific mismatch ratios distributions plotted in Fig. 1I for all xisRNA loci. Statistical tests for differences in xisRNA mismatch ratios for individual positions were signifi ant (adjusted p-value ≤ 0.05) only for the last fi st, second and third nucleotides (positions − 1, − 2 and − 3) where xisRNA loci displayed overall higher proportions of mismatches for sense reads than for antisense reads. The substitution matrix of genomic sequence versus reads sequences for positions − 2 and − 1 in Supplementary Fig. S4 indicates that xisRNA sense reads untemplated 3′ nucleotides are overwhelmingly biased for cytosines, independently of the nature of the underlying genomic nucleotide. To rule out the possibility that this predominance of 3′ untemplated nucleotides in sense xsiRNA reads would be a global artifact of our Illumina sequencing experiment, we conducted the same analysis for a subset of non-xisRNA experimental sRNA loci and did not conclude on any signifi ant differences in mismatch ratios (Supplementary Fig. S4). In contrast, testing differences in sense reads mismatch ratios between pri-miRNA transcripts and xisRNA loci indicated that these ratios were signifi antly higher for xisRNA than for pri-miRNA loci (Supplementary Fig. S4).

In short, we obtained evidence that while similar to non-coding small RNAs loci, xisRNA loci are essentially composed of complementary reads reminiscent of a single sRNA duplex with those reads in the sense orientation relative to an annotated *cis*-mRNA having a high proportion of untemplated cytosines in the last 2–3 3′ nucleotides.


A
n=26644
n=64
Loci length (bp)
FALSE
TRUE
xisRNA loci
B
n=26908
n=124
n=64
other
overlaps MIRNA
xisRNA
Count
Fraction of reads mapping to genomic top strand
C
Length (nt)
Treatment
BAI3
BAI3H
H2O
Proportion of reads in the 'BAI3' condition
Mean expression (rpm + 1)
D
BAI3
BAI3H
H2O
log10(Mean_rpm+1)
MIRNA
xisRNA
sRNA loci type
E
n=27032
n=64
Proportion
MSU loci overlap type
Exon
Gene
Intergenic
FALSE
TRUE
xisRNA cluster?
F
LOC_Os03g15770.1
LOC_Os03g15770.2
LOC_Os03g15770.3
xisRNA002
G
LOC_Os04g56160.3
xisRNA023
H
Counts of observations
NMismatches
NMatches
Position relative to read 3'-end
I
Counts
antisense
sense
Percentage of mismatches
Position relative to 3'-end of reads

◂**Figure 1.** Xanthomonas-induced small RNAs (xisRNA) signatures and loci possess distinctive attributes. (**A**) Box plot representation (median, 1st and 3rd quartile and 1.5 × IQR) of the distribution of the length of the experimental sRNA loci, contrasted as a function of whether they correspond to a xisRNA or not. Note that 388 loci were excluded from the plot because their length is above 2000 bp. (**B**) Distribution of the fraction of reads mapping to the genomic top strand for experimental sRNA loci contrasted as a function of whether the loci correspond to xisRNA, MIRNA or other loci. (**C**) Distribution of sRNA reads length and mean expression at the 64 xisRNA loci. The hierarchical clustering of xisRNA loci is based on the profiles of reads proportions for individual length values. Mean expression values in each treatment (shades of pink) are represented on a log10 scale. (**D**) Distribution of mean normalized expression in reads per million (rpm) for pri-miRNA loci (n = 87) versus xisRNA loci (n = 64) in libraries derived from various leaf inoculation treatments. (**E**) Proportions of the type of rice annotation features overlapping with experimental sRNA loci, contrasted as a function of whether sRNA loci correspond to a xisRNA or not. Overlaps are included in the counts if spanning at least 20 bp. Experimental sRNA loci overlapping both a gene and an exon annotation range are classifi d as 'Exon'. Those overlapping a gene annotation element only are classifi d as 'Gene'. Others are classifi d as 'Intergenic'. (**F**) Genome browser snapshot of the MSU7 LOC_Os03g15770 locus annotation overlapping with the xisRNA002 ShortStack segment and read coverage tracks (reads per million) for the various experimental treatments. Shorter exon boxes correspond to untranslated regions. Coverage of reads mapping to the top genomic strand is represented with positive values and a violet colored area whereas coverage of reads mapping to the opposite strand is represented with negative values and a golden colored area. The multiple sequence alignment below the genome browser graph corresponds to the three most abundant unique mapping signatures per strand of the ShortStack loci as well as the genomic sequence (LOC ID). In the sequence title, the two numbers indicate, respectively, the count of reads with this sequence in the "BAI3" treatment libraries and the length of the unique sequence. The "–" sign indicates that the original read sequences map to the complementary strand of the genomic locus. In this case, the displayed sequences are thus the reverse complement of the original read sequences. (**G**) Same as E but for the LOC_Os04g56160 locus region overlapping with xisRNA023. (**H**) Counts of nucleotides matching (Nmatches) or not matching (Nmismatches) the xisRNA002 loci genomic sequences along the read alignment region. Positions on the x-axis are expressed relative to the sRNA reads 3′-end (last nucleotide is − 1). Positive values are computed for reads mapping to the annotated sense strand of the LOC_Os03g15770 locus. Conversely, negative values correspond to reads mapping to the antisense strand of the annotated loci. (**I**) Distribution of genomic sequence mismatch ratios along relative positions within reads for xisRNA loci. Each data point represents the ratio of the number of mismatches over the total number of comparisons with the genomic sequence of the MSU annotated transcript for that position of all the reads mapping to a xisRNA locus in the sense orientation (violet) or antisense orientation (golden). Positions where a two-sided Wilcoxon signed rank test measuring differences in xisRNA mismatch ratios between read orientations were signifi ant (BY adjusted p-value ≤ 0.05) are marked with a "*" sign. Here, three "*" indicate an adj. P-val. ≤ 0.001. The bar graph on the top reports on the total count of comparisons for each orientation.

***X. oryzae* strains from multiple sublineages induce xisRNAs.** In order to independently validate the assumption that xisRNA NGS reads correspond to genuine sRNA species, we performed standard low molecular weight northern blots with probes designed to detect the moderately to highly expressed xisRNA duplex strands that are complementary to annotated *cis*-transcripts. In agreement with this view and as shown in Fig. 2A and Supplementary Fig. S5, for all tested xisRNA loci, following Nipponbare leaf infiltration with the BAI3 strain, a T3SS-dependent signal was detected no earlier that 24 hpi and gradually increased in intensity up to 72 hpi. We were curious to determine if the capacity to trigger xisRNA expression was specific to *Xoo* BAI3 alone or if it is a conserved feature of the phylogenetically diverse *Xo* species. Northern blot experiments were conducted with RNA samples extracted from leaves infiltrated with reference strains representative of the main genetic and pathotypic groups: African (BAI3, MAI1), Asian (PXO99A, KACC 10331) or American (X11-5A) *Xoo* strains as well as African (e.g. MAI10) or Asian *Xoc* strains (e.g. BLS256). As shown in Fig. 2A and Supplementary Fig. S6, *X. oryzae* strains were broadly capable of eliciting xisRNA accumulation with contrasting strength and specific ties. In qualitative terms, African *Xoo* and *Xoc* strains induced all tested xisRNAs with the exception of xisRNA007 which seems to be specific to African *Xoo*. The Asian *Xoo* strains and the American *Xoo* strain were less active and fainter but reproducible signals could be detected essentially for xisRNA001 and xisRNA002. In the case of PXO99A, these signals were also T3SS-dependent. While xisRNA023 was not tested for the US *Xo* strain X11-5A, xisRNA023 signals appeared to be specific to the *Xoo* pathovar as these sRNAs were readily detected in response to both Asian and African *Xoo* strains (Supplementary Fig. S6).

As a complementary way to establish that xisRNA induction is not limited to African *Xoo* infection, we exploited the dataset generated by Zhao et al.[22] where sRNAs expression temporal dynamics in *Xoo* PXO99 versus mock-inoculated rice leaves was profiled at early time points following leaf clipping inoculation. A second dataset consisting of sRNA sequencing data 24 hpi with the Azacytidine-resistant derivative of PXO99 strain, PXO99A from the Plant MPSS databases[25] was also included. The DE statistics of experimental sRNA loci whose expression changed in relevant treatment comparisons and that overlapped with BAI3 xisRNA loci are summarized in Fig. 2B. Coverage profiles of sRNA reads in xisRNA loci further indicate that PXO99-induced reads map exactly to the same location as BAI3-induced reads (see coverage plots in our dataverse). In addition to confi ming that PXO99 also induces xisRNA accumulation, the higher sensitivity of sRNA-seq applied to leaf clipping assays samples indicates that the most abundant xisRNAs accumulates as early as six hours following inoculation in an assay that more closely recapitulates a natural infection of xylem vessels as opposed to massive infiltration of the leaf mesophyll.

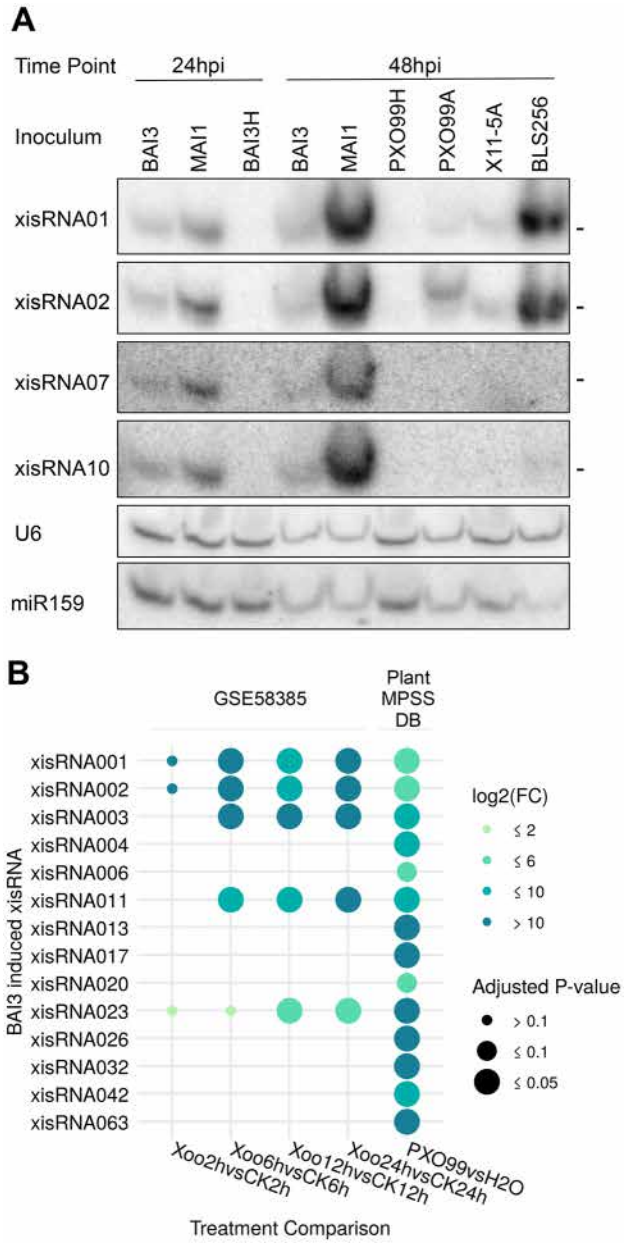


Figure 2. A subset of xisRNAs are detected in northern blots and exhibit contrasted expression patterns in response to phylogenetically diverse Xo strains. (**A**) Autoradiographs obtained after RNA gel northern blot analysis conducted with specific ligonucleotide probes hybridized to total RNA extracted from Nipponbare leaves infiltrated with the indicated strains and collected at the indicated time points (*hpi* hours post inoculation). The xisRNA probes detect RNAs complementary to the sense strand of the xisRNA *cis*-gene. The BAI3H and PXO99H strains correspond to a T3SS defective mutant derivative of the BAI3 and PXO99A virulent strains, respectively. The hyphens on the right of individual blots mark a size of 21 nt. Individual images derive from successive stripping, hybridization and detection rounds of the same membrane. Full-length blots are presented in Supplementary Fig. S15. Equivalent results were obtained in two other similar experiments. (**B**) Differential expression statistics of xisRNA loci found to be induced by the Asian Xoo PXO99 strain in public sRNA-seq datasets. The treatment comparisons (bottom labels) involve tissue samples inoculated with PXO99 versus a mock control at the indicated time point. The plot displays xisRNA loci with a log2(fold change) ≥ 1 and an BH adjusted p-values ≤ 0.1 in at least one comparison.

The broad xisRNA induction capacity in the genetic diversity of both *Xo* pathovars emphasizes the relevance of these sRNAs as evolutionary conserved molecular markers of early rice infection events and implies a potential functional role in this process.

A broader xisRNAs inventory. Having established that xisRNA accumulation is a widespread molecular feature of rice infection by *Xo* pathogens raised the question of whether other strains induced novel xisRNA loci that were not captured by our initial BAI3-specific sRNA-seq experiment. To address this, we performed a second set of rice sRNA-seq experiments, designated as the 'Diversity' dataset. African *Xoo* MAI1 strain was selected because in northern blot experiments it strongly induced xisRNAs. As representatives of the *Xoc* pathovar, the African BAI11 strain and the Asian BLS256, together with a T3SS mutant derivative (BLS256H), were also selected to investigate T3SS-dependency and *Xoc* pathovar specific xisRNAs. The setup of these experiments as well as the outcome of the primary genome segmentation and DE analysis are summarized in Supplementary Fig. S7 and Supplementary Table S2. Across both datasets, a total of 163 distinct loci met the criteria defi ed for xisRNA loci, 99 of which were not identifi d in the fi st 'BAI3' dataset. Supplementary Table S3 records expression measures, features of mapping sRNA reads and annotated *cis*-genes information for these loci while Fig. 3 depicts a graphical representation of this data for the 60 most highly expressed xisRNA loci. The clustering of DE patterns across comparisons (Fig. 3 and Supplementary Fig. S8) is complex but pointed to few main classes of xisRNA loci: 16 are broadly induced by all tested *Xo* strains (e.g. xisRNA01, 02), 22 are induced by African *Xoo* and *Xoc* strains (e.g. xisRNA009, 22) and 57 are specifi ally induced by the Asian Xoc BLS256.

In order to ascertain again that the sRNA reads mapping at dataset-specific xisRNA loci in this extended inventory were congruent with the hypothesis that xisRNAs derive from a duplex of sRNAs molecules, we inspected the coverage of all reads from both datasets. With the exception of xisRNA094 for which the shape of read coverage is not consistent with a sRNA duplex model, all other loci displayed a single major ~ 25 bp wide coverage peak. We therefore defined these shorter genomic intervals as 'xisRNA duplex loci' (see xisRNA loci-specific genome browser views in our dataverse). As shown in Supplementary Fig. S8 and Fig. 3, duplex loci reads complementary to the *cis*-gene strand have a proportion of 5′-T that is well above the proportion of Ts in the rice genome. A hallmark of sRNA loaded into rice AGO1s complexes is a strong sequence bias for 5′ Uracil[26]. This implies that the composition of *cis*-gene antisense xisRNA 5′-nucleotide is in line with their possible loading into an AGO1 complex.

The joint analysis of our sRNA-seq datasets generated a larger collection of xisRNA loci detected independently with diverse *Xo* strains and confi med that the main features of xisRNAs described in the BAI3 dataset were also conserved in the Diversity dataset. Furthermore, *cis*-gene antisense xisRNAs might be sorted into OsAGO1 complexes to silence *cis*-genes.

Many protein products of loci overlapping with xisRNAs harbor kinase domains and are involved in immune responses signaling. Of the 163 xisRNA duplex loci, 91% (149/163) overlapped with an MSU annotated gene (*cis*-gene) and 75% (122/163) were included within an annotated exon. To examine if those genes had related functions, we performed Plant GO Slim term enrichment analysis (Supplementary Fig. S9). In the Molecular Function ontology domain, we obtained a signifi ant enrichment (FDR ≤ 0.05) of the "RNA binding" term. Among several predicted proteins with RNA-binding domains, the presence of two genes involved in sRNA silencing pathways was especially noteworthy: the ARGONAUTE family member *OsAGO13*[16] associated with xisRNA064 and the DICER-LIKE family member *OsDCL4*/*SHO1*[27,28] associated with xisRNA022 (Fig. 3, Supplementary Table S3). In the Molecular Function GO category, the most significantly enriched terms are connected to "receptor activity" and "kinase activity", with 64 xisRNA loci overlapping genes (46%) annotated with this latter term. Consistent with GO enrichment analysis, the proportions of the different kinase groups in the xisRNA set is different from the whole genome composition (two-sided Fisher's Exact test p-value = 5e−04) and a one-sided post-hoc Fisher's exact test indicated that 5 subgroups (Extensin, WAK, DLSV, RLCK-IV, RLCK-VIIa-2) of the RLK/Pelle protein kinase family are signifi antly enriched in the xisRNA *cis*-gene products list (Holm method adjusted p-values < 0.05). Subfamilies such as DLSV, WAK, and RLCK-VIIa, were previously shown to be enriched in genes with a role in biotic stress responses based on either their function or expression profile[9,29]. To investigate in greater details if xisRNA *cis*-genes coding for kinase domain proteins are likely to participate in rice responses to biotic interactions, we constructed a phylogenetic tree of *Arabidopsis thaliana* and rice kinase domain-containing proteins (Supplementary Data S1). In this tree, we identifi d those instances where a xisRNA duplex locus was fully included within the mature rice transcript. As illustrated in Fig. 4, manual examination of individual subclades with associated xisRNAs revealed some interesting insight: fi st, 17 xisRNA have a *cis*-gene product in a subtree of RLCK-VIIa proteins (Fig. 4A). Most of the OsRLCK clustering with RIPK, an *Arabidopsis* RLCK involved in ETI signaling, stomatal defense and root development[30], have a cognate xisRNA. Still on Fig. 4A, in the BIK1/PBL1[30] clade, OsRLCK107, OsRLCK118 and OsRLCK176 have been previously shown to contribute to PTI or resistance to *Xoo* downstream of the OsCERK1, SDS2 or Xa21 receptors[31–33] and their transcript has a xisRNA duplex. Second, on Fig. 4B, the RLCK-VIIa proteins OsRLCK28 and OsRLCK55, cluster with the Arabidopsis decoy PBS1[34] and their mRNA overlap respectively with xisRNA007 and xisRNA044. OsRLCK55 and OsRLCK185 (also in this clade) interact with a *Xoo* type III effector. OsRLCK185 is essential for PAMP-triggered rice immune signaling[35]. Finally, a couple of the most highly and broadly induced xisRNAs are associated with subgroups of RLK/Pelle protein kinase of the DLSV subfamily (Fig. 4C): the xisRNA010-associated protein kinase domain is closely related to the *Arabidopsis* CRK2 protein which is required for PTI and defense against a bacterial pathogen[36]. xisRNA004-associated protein kinase domain is closely related to *Arabidopsis* AT1G07650/LMK1 which possesses cell death induction activity[37].

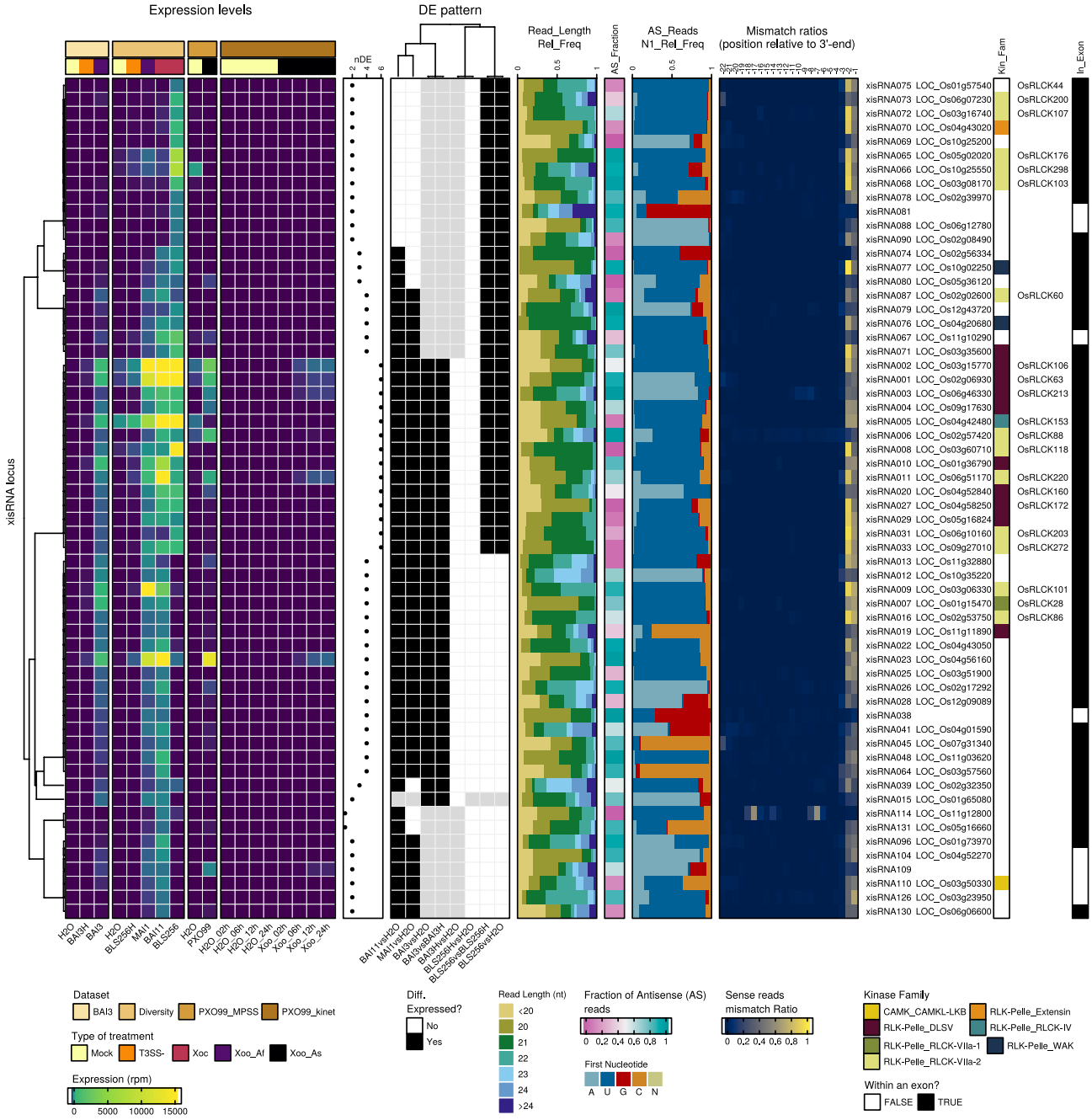


Figure 3. A summary of xisRNAs features in the analyzed sRNA-seq datasets. In the complex heatmap layout, the top 60 xisRNA loci with the highest maximum treatment mean expression across datasets are described row-wise. The fi st component, on the left, eports on expression (reads per million) in the various inoculation treatments of the different datasets. In the type of treatment legend, the "Xoo_Af" and "Xoo_As" labels stand for African *Xoo* and Asian *Xoo* strains, respectively. The DE patterns component shows whether xisRNA loci were detected as DE in various comparisons. The "nDE" dot plot depicts the number of comparisons where the xisRNA locus was deferentially expressed. Cells in gray indicate that the xisRNA locus was not considered in the dataset because the loci was not detected by ShortStack or because reads in this dataset did not pass our xisRNA loci features criteria. The subsequent plots on the right depicts the size distribution of reads ("Read_Length_Rel_Freq"), the fraction of antisense reads ("AS_Fraction"), the proportion of antisense reads with distinct 5′-end fi st nucleotide ("AS_Reads_N1_Rel_Freq") and the specific ucleotide mismatch ratios for relative positions −22 to −1 of antisense reads using xisRNA loci mapping reads from the "BAI3" and "Diversity" datasets. The MSU7 identifier f the gene overlapping with a xisRNA duplex locus is indicated together with the kinase subfamily into which the encoded protein is classifi d ("Kin_Fam." column) as well as the RLCK protein name, when relevant and available. Finally, the "In_Exon" column indicates whether the xisRNA duplex loci is fully included in an exon of an annotated mRNA of the gene. All numerical values reported in the graphic were computed using reads mapping to the xisRNA duplex loci. Note that the antisense polarity is defi ed relative to the orientation of the underlying *cis*-gene. To compute values that incorporate a notion of polarity for xisRNA loci that do not overlap with an annotated gene, reads mapping to the top strand were arbitrarily considered to be in the sense orientation.

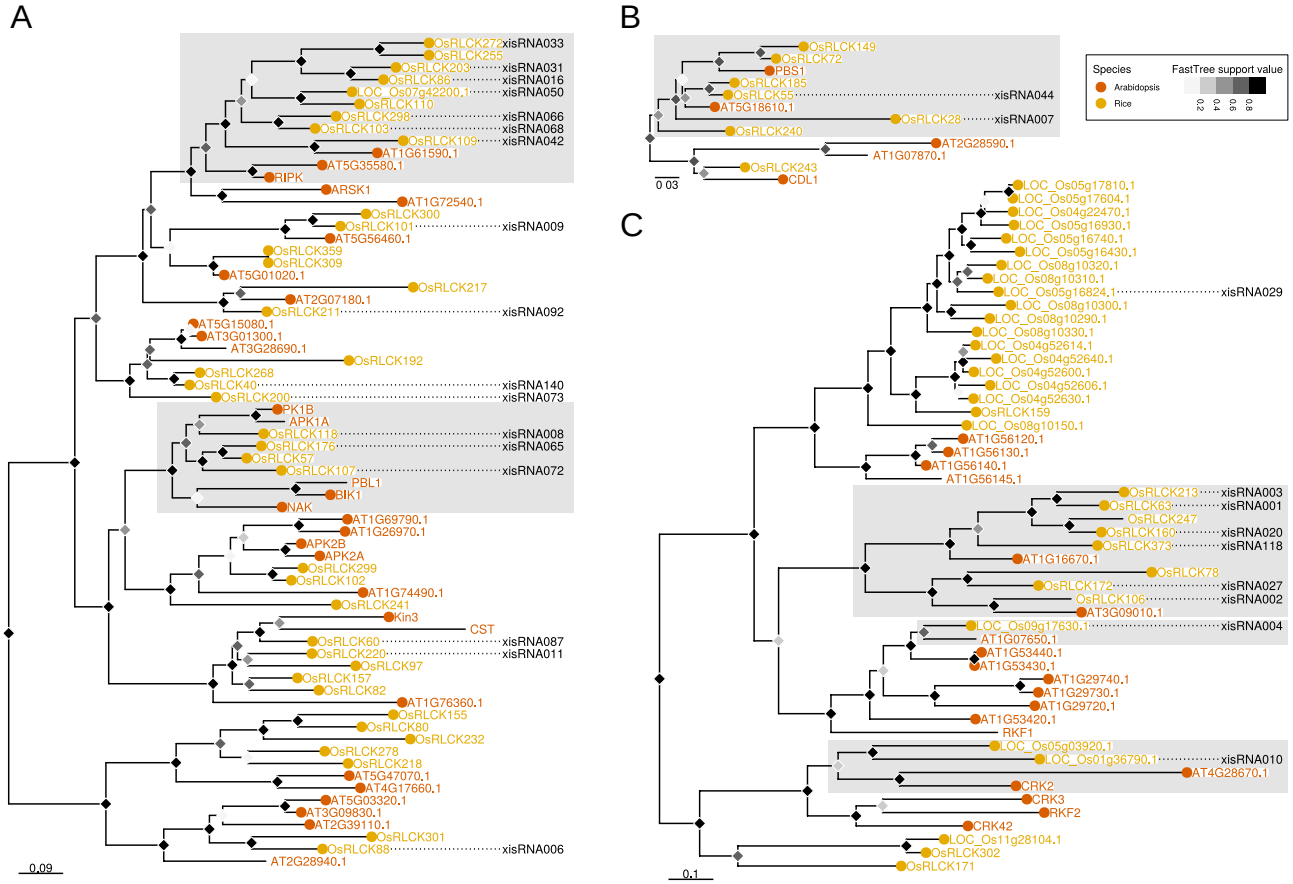


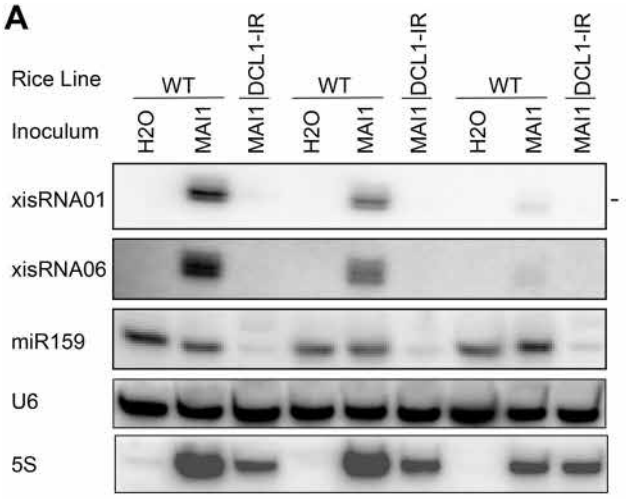
Figure 4. Multiple xisRNA loci *cis*-gene products with kinase domains are related to immune signaling pathways components in rice or Arabidopsis. The displayed phylogenetic trees correspond to selected sub-clades of proteins belonging to the RLCK-VIIa-2 (**A**), RLCK-VIIa-1 (**B**) and DLSV (**C**) subfamilies in the genome-wide phylogenetic tree (Supplementary File S1) constructed using the kinase domain of annotated proteins in the *Arabidopsis* Col-0 and Rice Nipponbare genomes. Dotted lines connect plant protein names with xisRNA identifie s when the xisRNA duplex loci is fully included in an exon of the underlying gene. Shaded rectangles highlight clades that are discussed in the text. The FastTree support value grey scale for tree nodes corresponds to local bootstrappings used by FastTree to estimate the tree's reliability. Scale bars correspond to substitutions per site.

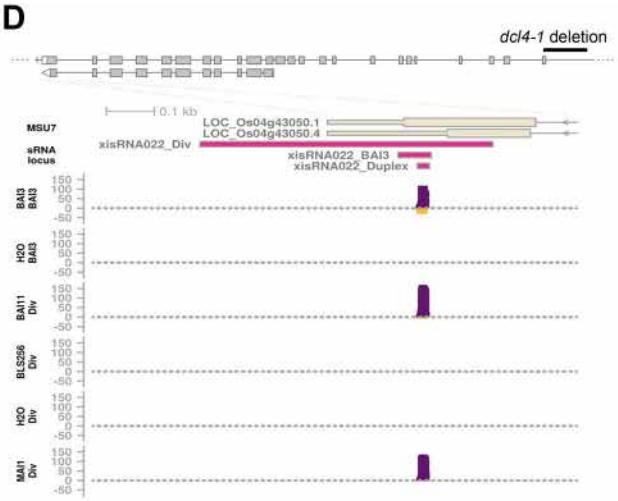
Overall, xisRNA duplex loci *cis*-genes are remarkably biased for loci encoding RLK and RLCK proteins. A number of protein sequences harboring a kinase domain and associated with xisRNA loci are phylogenetically or functionally linked to immune signaling. Thus, if xisRNA have regulatory activity, it is conceivable that they act in *cis* to dampen PTI signaling.

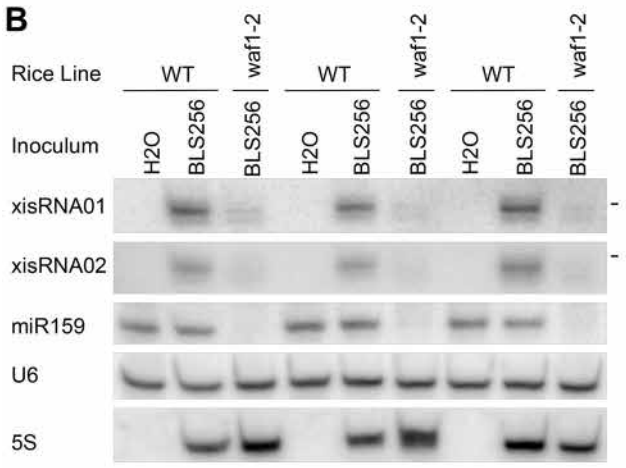
xisRNAs accumulation is dependent on canonical components of the rice regulatory sRNAs biogenesis machinery. As a fi st step in characterizing the molecular mechanisms underpinning xisRNA induction, we investigated the capacity of rice lines compromised for the activity of canonical regulatory sRNA synthesis pathway components to produce xisRNAs in response to bacterial infection.

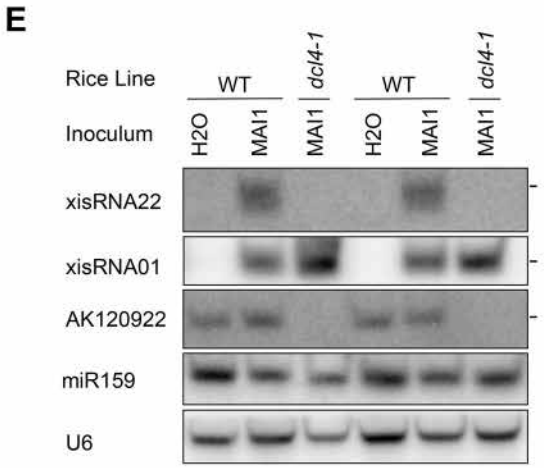
In rice, OsDCL1 is required for rice canonical ~ 21 nt miRNA biogenesis[38]. Because xisRNAs are 20–22 nt long, we tested the inverted repeat transgenic line OsDCL1-IR previously shown to be specifically silenced for the *OsDCL1*[39]. In northern blots, as expected in plants with *OsDCL1* activity defects, the MAI1-inoculated OsDCL1-IR individual plants had strongly reduced miRNA159 levels (Fig. 5A). The OsDCL1-IR plants also failed to respond to MAI1-inoculations and displayed xisRNA01 and xisRNA06 levels similar to water inoculated negative controls, which suggests that *OsDCL1* is required for the induction of these xisRNAs. We note that in the OsDCL1-IR individuals, the signals for the *X. oryzae* 5S rRNAs probe, acting as a proxy for MAI1 cells leaf density, tended to be lower than in Nipponbare controls. Th s indicates that *OsDCL1*-silenced plants may be more resistant to MAI1 infection.

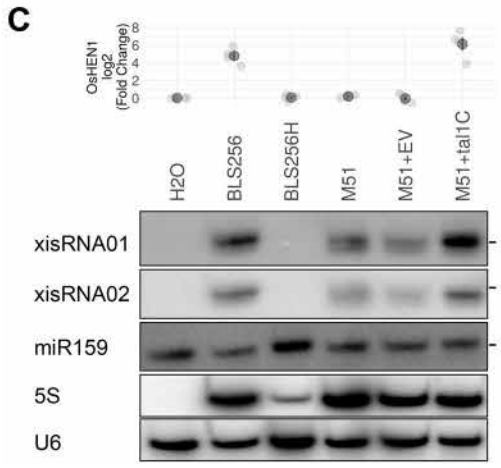
Canonical plant regulatory sRNA biogenesis pathways converge on the HEN1 methyltransferase[3]. The rice *wavy leaf1-2* (*waf1-2*) line with a null mutation in *OsHEN1* has been shown to be impaired in the accumulation of miRNAs and ta-siRNA[40]. To investigate whether *OsHEN1* activity is important for xisRNAs biogenesis, we infiltrated wild type Kinmaze and *waf1-2* individual plants with BLS256 and monitored xisRNA accumulation (Fig. 5B). The miR159 probe confi med that the *waf1-2* mutation greatly reduced miR159 accumulation in


A
Rice Line
WT
DCL1-IR
Inoculum
H2O
MAI1
xisRNA01
xisRNA06
miR159
U6
5S


D
dcl4-1 deletion
0.1 kb
MSU7
LOC_Os04g43050.1
LOC_Os04g43050.4
sRNA locus
xisRNA022_Div
xisRNA022_BAI3
xisRNA022_Duplex
BAI3 BAI3
H2O BAI3
BAI1 Div
BLS256 Div
H2O Div
MAI1 Div


B
Rice Line
WT
waf1-2
Inoculum
H2O
BLS256
xisRNA01
xisRNA02
miR159
U6
5S


E
Rice Line
WT
dcl4-1
Inoculum
H2O
MAI1
xisRNA22
xisRNA01
AK120922
miR159
U6


C
OsHEN1 log2 (Fold Change)
H2O
BLS256
BLS256H
M51
M51+EV
M51+tal1C
xisRNA01
xisRNA02
miR159
5S
U6

◂**Figure 5.** Rice sRNA biogenesis genes impact xisRNA accumulation. The images of autoradiographs obtained after RNA gel northern blot analysis conducted with specific ligonucleotide probes hybridized to total RNA extracted from leaves of wild type Nipponbare (WT) or a *OsDCL1*-silenced line (DCL1-IR) in (**A**), wild type Kinmaze (WT) or a *OsHEN1* mutant allele line (*waf1-2*) in (**B**) or Nipponbare in (**C**). Leaves were infiltrated with the indicated strains and collected 24 hpi. These results are representative of two (*OsDCL1*) and three (*OsHEN1* and Tal1c blots) replicate experiments. In (**C**), the plot on top of the gel images represent *OsHEN1* Q-RT-PCR expression data from four experiments (light gray point) with mean (dark gray points) and standard error (line ranges). The BLS256H strains correspond to a T3SS minus mutant derivative of BLS256. The "+" sign indicate that the strains were transformed with a plasmid expressing Tal1c or the corresponding empty vector (EV). Full-length blots are presented in Supplementary Fig. S15. (**D**) Diagram depicting the general exon–intron structure of *OsDCL4* and a close-up genome browser view of the last exon of the gene with the 3′ UTR. The *dcl4-1* deletion is not drawn to scale but span ~ 1.5 kb on the 5′-side of the gene which includes upstream promoter sequences up to the fi st 72 bp of exon 1[27]. The genome view represents reads coverage in various treatments, the MSU7 annotation, and the xisRNA022 locus defi ed in the "BAI3" and "Diversity" sRNA-seq datasets as well as its duplex locus. For details see the legend of Fig. 1F. (**E**) Northern blot analysis conducted on total RNA extracted at 48 hpi from leaves of homozygous wild type (WT) or homozygous *dcl4-1* individuals. Th s experiment was repeated three times with similar results. In (**A**), (**B**) and (**E**), lanes with the same labels correspond to biological replicates of total RNA samples extracted from a single infiltrated leaf area from individual independent plants. The hyphens on the right of individual blots mark a size of 21 nt. For each panel, individual blot images derive from successive stripping, hybridization and detection rounds of the same membrane.

homozygous mutant individuals. While BLS256 infiltration triggered xisRNA001 and xisRNA002 accumulation in wild type plants relative to water controls, this effect was very much attenuated in the *waf1-2* individuals indicating that, in addition to *OsDCL1*, *OsHEN1* activity is necessary for xisRNA induction. *OsHEN1* is probably universally upregulated by Asian *X. oryzae* TALEs including Tal1c from the Xoc BLS256 strain and Tal9a from the Xoo PXO99 strain[41,42]. To date, the functional signifi ance of this induction remains elusive. We therefore asked whether Tal1c-mediated *OsHEN1* induction during BLS256 infection affects xisRNA accumulation. As exemplified in Fig. 5C, compared to the wild type BLS256 strain, the xisRNA001 and xisRNA002 signals were slightly weaker in rice leaves infiltrated with the *tal1c* insertion mutant strain M51[41] in the BLS256 background. Conversely, expression of *tal1c* from a plasmid in the M51 background not only restored *OsHEN1* induction but also increased xisRNA001 and xisRNA002 levels relative to M51 complemented with an empty vector. Because the effect of *tal1c* activity on xisRNA accumulation were relatively mild and to test if *tal9a* has a similar effect, we produced a PXO99A strain derivative with a deletion of the entire tal9 TALE genes cluster, which contains 5 *tal* including the *tal9a* gene[43]. In an experiment summarized in Supplementary Fig. S10, while this mutant had lost the ability to induce *OsHEN1* and this ability could be restored by the *Tal1c* plasmid, xisRNA023 levels were not diminished in leaves infiltrated with the tal9 cluster deleted strains relative to PXO99A. Thus, we conclude that *OsHEN1* induction during *Xoc* BSL256 infection maximizes xisRNA accumulation but a comparable effect is not observed during *Xoo* PXO99A infection.

Overall, our epistasis analysis agrees with the view that xisRNAs production requires the canonical regulatory sRNA biogenesis pathways component *OsDCL1* and *OsHEN1*. As bona fide products of these pathways, xisRNAs may thus possess PTGS-related regulatory activity.

xisRNA accumulation is DCL4-independent but probably requires expression of the *cis*-gene. OsDCL4 is the preponderant rice DICER-LIKE responsible for the cleavage of dsRNA precursors that releases ~ 21 nt siRNAs associated with inverted repeat transgenes, hp-siRNAs and *TAS3* ta-siRNAs[27]. As noted above and shown in Fig. 5D, the African *X. oryzae*-specific xisRNA022 duplex is located on the last exon of LOC_Os04g43050 which precisely corresponds to *OsDCL4*/SHO1[16]. To test if *OsDCL4* is also involved in xisRNA synthesis, we examined the effect of the null *dcl4-1* allele on xisRNA accumulation in northern blot experiments. The *dcl4-1* mutation, identifi d in the L16S indica variety, carries a ~ 1.5 kb deletion (Fig. 5D)[27]. The AK120922 transcript can form a hairpin-like structure and has been shown to be processed by OsDCL4 to produce 21 nt hp-siRNAs[27]. A northern blot oligonucleotide was designed to detect one of the most abundant reads on the genomic region of AK120922 in our BAI3 dataset and was used as a probe to monitor AK120922-derived hp-siRNAs. As shown in Fig. 5E, while wild type and *dcl4-1* leaves expressed similar levels of miRNA159. The AK120922 signal detected in wild type individuals was lost in samples from the *dcl4-1* plants, thus confi ming the absence of OsDCL4 activity. The *dcl4-1* mutation did not impair xisRNA01 accumulation because the corresponding signal in MAI1 infiltrated mutant plants was not decreased relative to corresponding wild type plants. In contrast, MAI1-triggered accumulation of xisRNA022 in *OsDCL4* plants, disappeared in *dcl4-1* individuals.

From these results, we conclude that *OsDCL4* does not play a general role in xisRNAs production. However, the promoter region deleted in *dcl4-1* is specifi ally required for induction of the *OsDCL4*-associated xisRNA locus. *dcl4-1* plants fail to express a *OsDCL4* transcript[27] which potentially suggests that the expression of the protein-coding mRNA overlapping with the xisRNA duplex is required for double-stranded precursor biogenesis and consequently the production of the corresponding xisRNA.

Genome-wide transcriptomics data provides limited support for a general role of xisRNAs in *cis*-loci silencing. We have established that xisRNA loci rely on components of regulatory sRNA biosynthesis pathways to form duplexes of 20–22 nt molecules that predominantly map to exons of protein coding genes. It is therefore conceivable that xisRNAs act in *cis* to silence overlapping transcripts. To address the potential

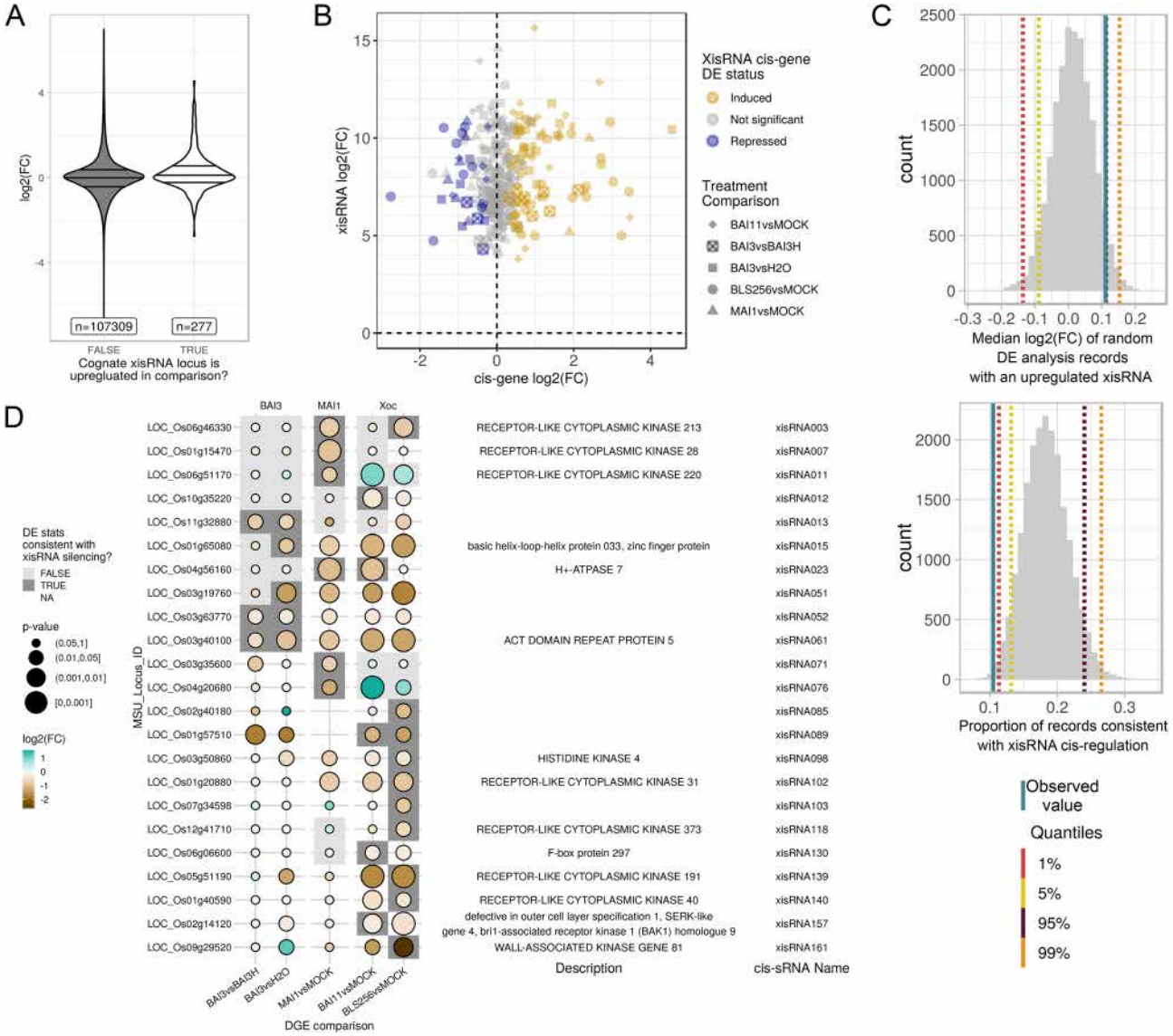


Figure 6. mRNA-seq data analysis points to a restricted set of xisRNA *cis*-genes as potential silencing targets. (**A**) Distribution of the $\log_2$ transform of the fold change ratio for DE analysis records (expressed gene by treatment comparison combinations) contrasted on whether or not the gene has a cognate xisRNA and this xisRNA is also up-regulated in the same treatment comparison. The horizontal lines refl ct the 25%, median and 75% quartiles. (**B**) log2(FC) of xiRNA versus cognate *cis*-gene in a relevant treatment comparison. For each treatment comparison, a *cis*-gene is considered to be 'Induced' if it has a log2FC > 0 and a p-value ≤ 0.05, 'Repressed' if the log2FC < 0 and the p-value ≤ 0.05 and 'Not signifi ant' otherwise. (**C**) Histogram distributions of the variables computed in 20,000 trials for randomization test. (**D**) Summary of the rice annotated *cis*-genes whose expression profile is consistent in at least one comparison with a possible negative regulatory effect of a cognate xisRNA. The 'DE stats consistent with xisRNA silencing status' is regarded as "TRUE" if in the corresponding comparison, the xisRNA has been detected as induced and if the gene DE record has log2FC < 0 and a p-value ≤ 0.05, "FALSE" if the xisRNA has been detected as induced but the gene DE record does not pass the log2FC or p-value criteria and "NA" (Not Available; white background) otherwise.

cis-regulatory activity of xisRNAs, we applied a standard transcriptomics data analysis approach based on the postulate that candidate target genes are expected to be downregulated in a treatment comparison where the corresponding sRNA is upregulated. We generated paired-end mRNA-seq data using samples from the set of experiments that were also used to obtain the BAI3 sRNA-seq dataset. In addition, we included sequences from Nipponbare samples collected 24 h after infiltration with the *Xoo* MAI1 strain and a water control from a previously published dataset[44]. Finally, sequences from Nipponbare samples collected 48 h after infiltration with the *Xoc* strains BLS256, BAI11 (a.k.a. CFBP7342), or the mock control were also retrieved from another published[45] dataset. Differential expression metrics were computed for comparisons involving samples inoculated with a virulent strain versus samples inoculated with the corresponding T3SS mutant strain or mock treatment.

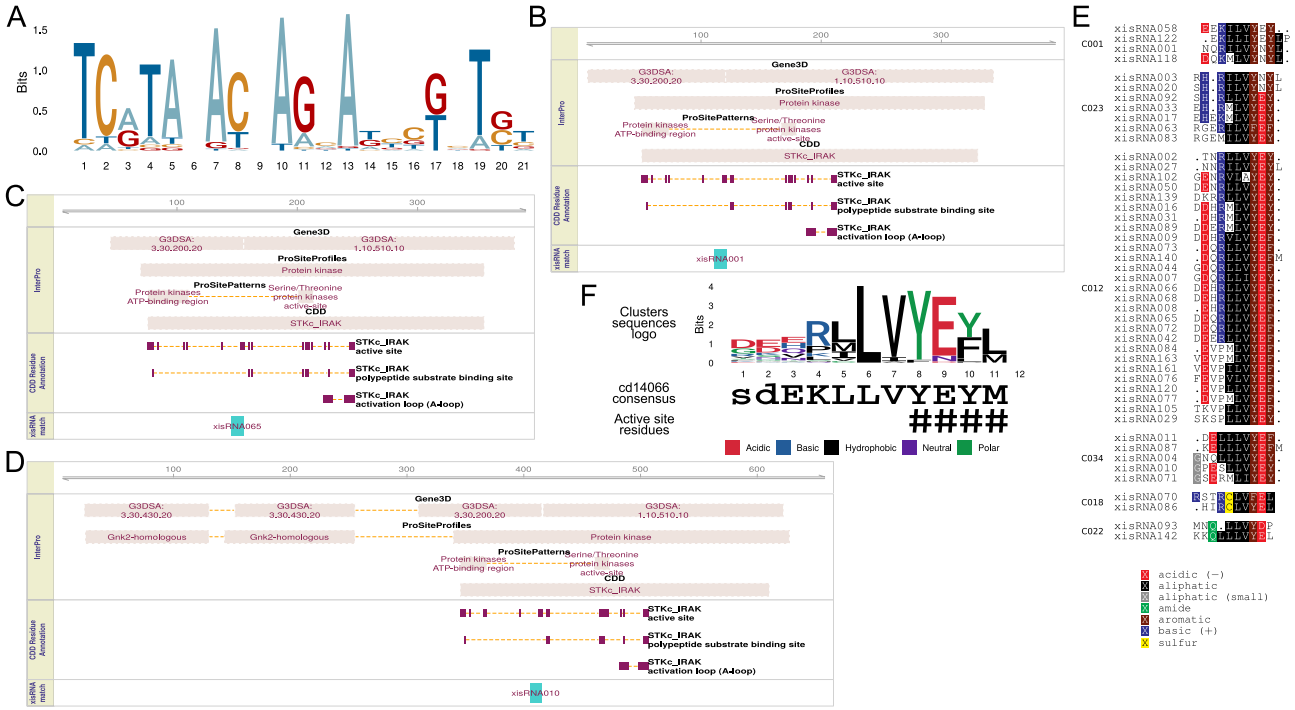


Figure 7. Analysis of xisRNA duplex sequences similarities. (**A**) Sequence logo representation of a multiple alignment of major antisense xisRNA duplex reads sequences belonging to the clusters defi ed in Supplementary Fig. S11. The sketches of InterPro domain hits in panel (**B**), (**C**) and (**D**) for respectively xisRNA001, xisRNA065 and xisRNA010 *cis*-gene products are representative and illustrative of the observation that xisRNA duplex coding sequences consistently coincide with a specific inase domain protein regions. (**E**) Multiple protein subregion sequences alignments for a subset of 6 individual clusters from those defi ed in Supplementary Fig. S12 (the clusters labels are consistent). (**F**) Sequence logo representation of a multiple alignment of the 47 amino acid sequences displayed in (**E**). The active site residues described for the NCBI CDD STKc_IRAK family (accession cd14066) are highlighted with hash symbols.

First, we examined the distribution of the log2 transform of the fold change ratio of genes in the various comparisons depending on whether an annotated transcript of the gene overlapped with a xisRNA duplex loci that is upregulated in the corresponding comparison (Fig. 6A). If upregulation of a cognate xisRNA generally suppressed *cis*-gene expression, we would expect the median log2(FC) of this set of DE records to be inferior to zero. The computed median log2(FC) is 0.111 and, on the contrary, is significantly superior to zero (one-sided Wilcoxon rank sum test: $W = 25{,}008$, $p = 8e{-}06$). Likewise, a two-sided Fisher's Exact test for a specific enrichment of genes with a cognate up-regulated xisRNA in repressed genes ($\log2(FC) < 0$ and p-value ≤ 0.05) across treatment comparisons produced a moderately significant p-value of 0.03 for the BAI11vsMOCK comparison only. Conversely, of the 277 DE records pertaining to genes having an up-regulated cognate xisRNA in the corresponding comparison only 29 (10.5%) meet these criteria for down-regulation (Fig. 6B).

To explore further whether a xisRNA *cis*-gene gene is more likely to be down-regulated when the corresponding xisRNA is up-regulated, we performed randomization tests (20,000 trials) where each xisRNA loci was assigned a random locus identifier as cognate *cis*-gene. The median log2(FC) of DE records of randomly assigned genes with an upregulated xisRNA as well as the proportion of these records that are consistent with negative regulation ($\log2(FC) < 0$ and p-value ≤ 0.05) were subsequently computed as above for the observed values. As shown in Fig. 6C, 95.1% of the median log2(FC) values in the random set were lower than the observed value. Similarly, 99.4% of the proportions of records that are consistent with negative regulation in the random set were higher than the observed value.

Contrary to our initial hypothesis, this analysis argues against the view that xisRNA generally direct *cis*-silencing. It cannot be ruled out however that xisRNA do occasionally act in *cis* to diminish transcript accumulation and we listed in Fig. 6D, 23 *cis*-gene–xisRNA pairs whose DE data is consistent with a regulatory effect.

xisRNAs duplexes exhibit limited nucleotide sequence similarity but those associated with kinase loci coincide with mRNA regions encoding a kinase active site motif. The fact that a noticeable fraction of xisRNAs is associated with genes encoding proteins with closely related kinase domains raises the question of xisRNA sequences relatedness. We used systematic pairwise alignments between xisRNA duplex loci sequences of the most abundant read (19–24 nt) mapping on the *cis*-locus antisense strand to perform hierarchical clustering. Of the 148 loci considered (xisRNA133 had no antisense major reads), 54 were grouped into 21 clusters of related xisRNA major reads where the number of variable positions ranges from 1 to 8 (Supplementary Fig. S11). Interestingly, most individual xisRNA duplex loci in these clusters overlap a kinase

sRNA class	Precursor	Preferred DCL	DCL processing duplexes	Size (nt)
het-siRNA	dsRNA dependent on RDR2 and Pol IV	DCL3	Unphased duplexes covering the precursor	23–24
miRNA	Single-stranded hairpin RNA	DCL1	Single duplex	21
hp-siRNAs	Single-stranded hairpin RNA	DCL4	Unphased duplexes covering the precursor	21
phasiRNA	dsRNA dependent on RDR6	DCL4	Phased duplexes covering the precursor	21
NAT-siRNA	Complementary pair of Pol II RNAs forming a dsRNA and often dependent on a RDR	Variable depending on locus (DCL1, DCL2 and/or DCL3)	Variable depending on locus	Variable depending on locus (21, 22 or 24)
xisRNA	dsRNA likely deriving from an underlying Pol II mRNA and an unidentifi d complementary molecule	DCL1	Single duplex	20–22

Table 1. A comparison of xisRNA properties against main documented classes of endogenous plant non-coding sRNAs.

encoding locus and the sequences clusters display weak overall similarities (relatively lower mismatch counts outside the diagonal of the matrix in Supplementary Fig. S11). To visualize these similarities, the multiple alignment of all major reads in the clusters was used to generate the consensus logo reproduced in Fig. 7A. In line with our earlier observation regarding 5′ nucleotides distribution of antisense reads, major reads frequently start with a "TC" (or "UC" for an RNA sequence) dinucleotide on their 5′-end. The consensus pattern further points to conserved positions that seem to be phased with a 3 nt off et, reminiscent of codon code degeneracy.

This pattern of conserved nucleotides in xisRNA sequences hints to the possibility that xisRNA duplex loci correspond to *cis*-transcripts regions encoding related protein motifs. To determine if xisRNA duplexes correspond to the coding sequences of documented domains, we searched InterProScan hits on *cis*-transcripts translation products and analyzed their overlap with xisRNA-associated protein subsequences (97 out of 122 that overlap an annotated transcript are also strictly within a CDS). Domains that repeatedly overlapped with a xisRNA-associated protein subsequences were connected with kinase activity. For example, 50 out of 97 overlap with a ProSite "Protein kinase" domain (PS50011). As illustrated in Fig. 7B,D, 39/97 xisRNA-associated protein regions overlapped with at least one residue matching with an "active site" features in the NCBI CDD database "STKc_IRAK" domain (cd14066). The corresponding STKc_IRAK conserved sites consensus motif 'YEYM' is involved in ATP binding according to its the CDD domain description. As a broader complementary approach, we measured the degree of similarity between translation products of xisRNA duplex loci sequences following a procedure similar to the one used above for xisRNA major reads sequences (Supplementary Fig. S12). While many clusters displayed no particular similarity with one another, a subset of six clusters with a total of 47 sequences appeared to be related as judged by dissimilarity score patterns off the matrix diagonal. In agreement with the InterPro domains overlap analysis, the consensus motif at the C-terminus of the xisRNA-associated protein sub-sequences clusters multiple alignment (LVYE[YF][LM]) match very well with the 'YEYM' consensus sequence of the "STKc_IRAK" ATP-binding motif in this region (Fig. 7E,F).

In summary, our analysis of xisRNA duplex antisense sequences similarities revealed that most xisRNAs display limited sequence conservation but defined, nonetheless, a few families containing similar reads. It also appears that about half of the xisRNA duplex overlapping a CDS coincide with sequences coding for a conserved RLK/Pelle kinase motif that is involved in ATP binding.

Discussion

Our understanding of the biology of plant regulatory sRNAs bears on a few paradigmatic pathways refl cting common biosynthesis routes and regulatory activities. However, computational approaches similar to the one that have been taken here are revealing the hidden diversity of atypical non coding small RNA types[46–48]. Here we report on novel rice 20–22 nt small non-coding RNA populations highly induced upon infection by a set of diverse *X. oryzae* strains. Only pathogenic bacteria possessing the T3SS virulence factor translocation system are competent for xisRNA induction.

The discovery of xisRNAs prompted an examination of rice genetic requirement for their biogenesis. xisRNA production appears to utilize host regulatory small RNA synthesis pathways components *OsDCL1* and *OsHEN1* but not *OsDCL4* (Fig. 5). Th s suggests that xisRNAs derive from precursors processed by these enzymes and argues in favor of the view that xisRNAs are genuine small silencing RNAs. *OsDCL1* is dispensable for siRNA originating from CentO repeats and inverted repeats transgenes but is strictly required for the production of miRNAs[38]. The ShortStack experimental sRNA loci classifi ation algorithm did not score any of the xisRNA loci as likely *MIRNA*. Our analysis of sRNA reads properties at the xisRNA loci further indicates that upon bacterial infection most of these loci generate sRNA reads mapping to both strands of the genome sequence. These observations are incompatible with the possibility that xisRNA loci correspond to canonical *MIRNA* genes. Previous work has identifi d differentially accumulating rice miRNA during Asian Xoo infection[21,22], it was therefore surprising that no xisRNA locus corresponded to a *MIRNA*.

Even if akin to siRNAs, xisRNA loci reads coverage supports the existence of a single xisRNA duplex at each locus (Figs. 1 and 6A and our dataverse). In contrast, endogenous siRNAs loci such as hp-siRNAs, phasiRNA, transgene-derived siRNAs, or 24 nt heterochromatic siRNAs generally give rise to several distinct siRNA duplexes that may be phased and that cover more or less uniformly the entire precursor sequence[5]. While xisRNAs do

not fit well into canonical classes of plant sRNAs (Table 1), they resemble a subset of DCL1-dependent 21 nt *cis*-nat-siRNAs from rice and Arabidopsis deriving from one specific site in the region of antisense transcripts overlap[49]. If extrapolated to all xisRNA loci, our interpretation that xisRNA022 accumulation requires a sense OsDCL4-encoding mRNA (Fig. 5) is consistent with this view. However, neither the reference rice Nipponbare annotation nor our BAI3-infected leaves mRNA-seq data (see coverage plots in our dataverse) argue for the widespread existence of an antisense messenger transcript complementary to the annotated protein coding mRNA. This antisense pol II RNA transcript could still be a very short-lived intermediate. Alternatively, the antisense strand of the xisRNA dsRNA precursor could be the product of RDR activity (Supplementary Fig. S14).

Interestingly, polarized signatures of untemplated 3′ terminal nucleotides were detected only in xisRNA reads with a sense orientation relative to the orientation of the overlapping annotated gene (Fig. 1 and Supplementary Figs. S3 and S4). The RDR2 strand of P4R2 RNAs and their DCL3-processing heterochromatic siRNA products carry a single 3′ end untemplated random nucleotide as a result of RDR2 nucleotidyl transferase activity[50]. 3′ to 5′ exonucleolytic truncation and/or 3′ tailing occurs pervasively in sRNAs of an *Arabidopsis hen1* mutant but has also been detected in wild type plants[51]. In this context, the tailing of unmethylated 3′ ends is accomplished by the HESO1 and URT1 uridyltransferases but other types of nucleotide additions occur in a *Arabidopsis hen1-2 heso1-2 urt1-3* triple mutant[52]. The pattern of substitution in xisRNA reads untemplated 3′ nucleotides is biased in favor of cytosines (Supplementary Fig. S4) which would require other nucleotidyltransferases[53].

In spite of an overall predominance of 20–22 nt reads, our sequencing data indicated that xisRNAs sizes vary remarkably (Fig. 1C). DCL1 dices dsRNA precursor in homogeneously sized 21 nt duplexes except for those 22 nt miRNAs that trigger phasiRNA production. If xisRNAs are indeed *OsDCL1*-dependent, this raises the issue of the origin of xisRNA size variability**.** Just like other plant small RNAs, xisRNA require HEN1 activity for their proper accumulation (Fig. 5B). The presence of noticeable proportions of reads < 21 nt for a few xisRNA loci may indicate that they underwent 3′ to 5′ trimming. Th s observation and the distinctive preeminence of 3′ non-templated nucleotides in sense reads may further indicate that xisRNA undergo 3′ to 5′ exonucleolytic truncation and that 2′-O methylation by the OsHEN1 methyltransferase is a rate limiting step in xisRNA biosynthesis. The observed effect of OsHEN1 induction by Tal1c (Fig. 5C) that maximizes xisRNA accumulation would be consistent with this conjecture.

Our results about xisRNA features and biogenesis argue in favor of the idea that they possess PTGS-related regulatory activity. The observed enrichment of 5′ uracyl in antisense xisRNAs (Fig. 3 and Supplementary Fig. S8) is a hallmark of sRNAs loaded into OsAGO1 complexes and supports this notion. xisRNA duplex loci generally overlap with annotated rice gene bodies (91%) and are included within annotated exons of these genes (75%). However, our mRNA-seq data analysis provided little support for a general role of xisRNAs in the suppression of *cis*-gene expression (Fig. 6). One potential pitfall of our analysis is that for the MAI1 and BLS256 strains no T3SS mutant inoculated samples were included and we thus may have overlooked PTI-regulated *cis*-genes induced by a T3SS minus strain. To demonstrate xisRNAs silencing activity, future work will need to address whether they are detected in AGO1 complexes and if they act in *trans* through partial complementarity with target transcript or if their regulatory activity mostly lies in translational inhibition of target genes similar to plant 22-nt siRNAs[54].

xisRNA *cis*-genes are enriched for RLK/Pelle kinase family proteins encoding loci (Fig. 3). In the absence of evidence for a *cis*- or *trans*-regulatory activity, it is difficult to interpret the significance of our results regarding the overlap between the xisRNA duplex loci and the coding sequence of a conserved kinases ATP-binding site (Fig. 7). miRNAs frequently target several members of gene families[2]. For example, the miR482/2118 superfamily targets several plant R genes at the level of a conserved P-loop coding region[6]. Probably evolving under different selective constraints, plant parasitic *Cuscuta* mobile sRNAs can be grouped into superfamilies that also target conserved host protein coding sequences[55].

We initially identified xisRNAs in experiments with the African *Xoo* strain BAI3 (Fig. 1) but subsequently realized that a subset of xisRNA duplex loci are also induced by other strains of this species and that other subsets of xisRNA duplex loci seem to respond differentially depending on the genetic background of the bacterial strains (Figs. 2, 3 and Supplementary Fig. S6). These observations underscore the importance of this phenomenon in a disease context and implies that it is evolutionary conserved across the *Xo* species diversity and across pathovars causing two distinct diseases. The dependence on a functional T3SS for xisRNA accumulation in a compatible interaction is an indication that xisRNA are functionally connected to rice susceptibility. Although the *in planta* growth defect of a T3SS strain rather than the specific activity of T3SS-dependent virulence effectors could also explain this result, PXO99 triggers xisRNA production at a very early stage (6 hpi) of leaf infection (Fig. 2B) in a T3SS-dependant fashion (Fig. 2A). This approximately correspond to the time frame at which the earliest T3SS virulence effectors dependent effects on the host transcriptome are usually detected[56,57]. Even if some extent of *Xo* multiplication is conceivably necessary for xisRNA appearance, it is not sufficient as exemplified by several xisRNA that only respond to a subset of virulent strains (Fig. 2A and Supplementary Fig. S6). Consequently, we favor the hypothesis that Xo T3SS virulence effectors are directly implicated in xisRNA biogenesis.

Rice mutant lines in sRNA biogenesis pathways components suffer a range of severe reproductive and developmental phenotypes that preclude rigorous testing in classical susceptibility assays for *Xo* diseases. We showed that Tal1c-mediated *OsHEN1* induction contributes mildly to xisRNA accumulation (Fig. 5C). Th s indicates that an individual Type III effectors acts on xisRNA production and further describes a fi st function for this TALE. A BLS256 derivative strain with a *tal1c* mutation is nevertheless as virulent as the wild type[58]. This could be due to an inadequacy of the assay to detect subtle phenotypes or this effect on xisRNA accumulation does not result in altered virulence. In line with previous studies on rice susceptibility to *M. oryzae*[39], our northern blot data on the *OsDCL1* silenced plants (Fig. 5A) does provide circumstantial evidence that this gene promote BLB susceptibility.

Many of the xisRNA duplex *cis*-genes encoding kinase domain proteins have documented or highly likely functions in rice immune signalling (Fig. 4) but, as discussed above, the level of the *cis*-transcript is rarely reduced in the presence of the cognate xisRNA. There are however exceptions, such as the xisRNA007 *cis*-gene

(Fig. 6) that belongs to the OsRLCK185 clade (Fig. 4B). Of the non-kinase loci overlapping xisRNAs, xisRNA023 is probably the most conspicuous. Not only because it is one of the most highly expressed in response to all strains except Xoc BLS256 (Fig. 3), but also because its *cis*-gene, *OSA7*, which is downregulated in the presence of xisRNA023 (Fig. 6), codes for a plasma membrane H^+-ATPase proton pomp[59] and is the rice orthologue of *Arabidopsis AHA1* and *AHA2* genes that have important functions in a range of physiological processes including immune signaling[60].

xisRNAs are evocative of eukaryotic pathogens' mobile sRNAs that suppress plant immunity genes expression and favor compatibility in Cross-Kingdom RNAi[7]. However, it is doubtful that xisRNA or their precursors are of bacterial origin because their sequences poorly match the genomic sequences of the bacterial triggers (Supplementary Fig. S13). Taken altogether, our results can be formalized in a working model (Supplementary Fig. S14) postulating that *Xo* pathovars use T3SS-translocated virulence effectors to co-opt rice sRNA biosynthesis and activity pathways to produce an unusual class of small silencing RNAs that alter plant immune signaling in order to promote bacterial infection. An alternative working model that appears less likely is that rice generates xisRNAs that direct bacterial gene silencing to mitigate bacterial invasiveness similar to what has been reported for a bacterial plant pathogen[61] and gut bacteria[62].

Th s study describes the discovery of xisRNAs. Future efforts will investigate in greater details their molecular activity and biological function with the long term goal to leverage this basic knowledge to improved disease control strategies in the field.

Methods

Bacterial strains and plant inoculation. Th *Xo* strains used in this study are described in Supplementary Table S4. The isolates were maintained in 15% glycerol at − 80 °C. All *Xoo* strains were cultivated on PSA (10 g/l peptone, 10 g/l sucrose, 1 g/l glutamic acid, 15 g/l Bacto Agar), supplemented with kanamycin (50 µg/ml), rifampicin (100 µg/ml) and/or tetracycline (5 µg/ml) when needed for the propagation of mutant and complemented strains. For plant infiltration, frozen bacterial stocks were spread on Petri dishes with PSA and incubated at 28 °C. On the morning of the experiment, bacterial suspensions were prepared in distilled water and injected with a 1 ml needleless syringe on the abaxial face of the youngest fully expended leaf of 4 weeks old rice plants. Bacterial suspensions were adjusted at $OD_{600nm} = 0.5$ except for infiltration of *waf1-2* homozygous plants that were found in preliminary experiments to require a three fold more concentrated suspension in order to deliver an equivalent number of bacteria in the leaf mesophyll at t_0. After infiltration, plants were kept in the same green-house growth conditions until sample collection.

Plant material and growth conditions. Th s study involved only laboratory experiments with cultivated rice parental lines and complied with relevant institutional, national, and international guidelines and legislation. Plants were grown under greenhouse conditions with cycles of 12 h of light at 28 °C and 80% relative humidity and 12 h of dark at 25 °C and 70% relative humidity. The following rice genotypes were used: cultivar Nipponbare[24], the DCL1-IR transgenic line #43-6 in the Nipponbare background[39], Kinmaze variety and *waf1-2* mutant derivative[40] and a heterozygous *dcl4-1* line[27]. For experiments on the effect of *dcl4-1*, the wild type controls consisted of individuals in the progeny of selfed *dcl4-1* heterozygous parents that were genotyped as homozygous wild type at the *OsDCL4* loci whereas *dcl4-1* plants were confi med to be homozygous for the mutant allele.

For PCR genotyping with primers listed in Supplementary Table S5 and the GoTaq G2 DNA Polymerase (Promega) according to manufacturer's instructions, template rice genomic DNA was extracted from leaf samples using a standard MATAB-based protocol[63].

RNA extraction. Unless stated otherwise, ~ 2 cm fragments were collected from a leaf of three independently infiltrated plants at the specifi d time point after inoculation and pooled in 2 ml microtubes containing 3 steel beads and immediately frozen in liquid nitrogen. For routine northern blot and Q-RT-PCR experiment, total RNA was extracted using 1 ml of Tri reagent (MRC) following the manufacturer's instructions and resuspended in nuclease-free water. For Illumina sequencing, the total RNA was treated with the TURBO DNA-free Kit (ThermoFisher) following instructions from the manufacturer and further purifi d with the mirVana Isolation Kit (ThermoFisher).

RNA next-generation sequencing. The sRNA-seq data for the BAI3 dataset was generated by the Fasteris company (Geneva, Switzerland). Briefly, sRNA libraries were prepared with total RNA extracted from duplicate samples collected at 24 hpi using the Illumina TruSeq small RNA kit with polyacrylamide gel purifi d 18–25 nt sRNA fractions and sequenced on a portion of a Illumina HiSeq 2000 lane, 1 × 50 single-reads. The sRNA-seq data for the Diversity dataset was generated by the Novogene company (Hong Kong, China) from total RNA extracted from triplicate samples collected at 24 hpi using a NEBNext Small RNA Library Prep Set for Illumina kit optimized by Novogene on the supplied total RNAs. Ultimately, size selected 18–47 bp inserts were sequenced on a Illumina HiSeq (1 × 50 bp). The 'BAI3' mRNA stranded sequencing was performed on a HiSeq 2000 lane (2 × 100 bp) by the Fasteris company (Geneva, Switzerland) with libraries prepared following a dir-mRNA-dUTP protocol that includes poly-adenylated transcripts selection followed by cDNA library construction using Illumina TruSeq Stranded mRNA Library Prep kit. Demultiplexing of sequencing reads was performed by the sequence providers.

Northern blots and quantitative RT-PCR. Low molecular weight northern blots were performed using previously described standard procedures[64]. Briefly, 10 µg of total RNA in a volume of 40 µl in 1 × gel loading

buffer II (Ambion) were loaded on a 17.5% denaturing acrylamide gel and electrophoresed at 80 V, transferred to a Hybond NX membrane (Amersham) and crosslinked with UV. The membrane was then pre-hybridized 30 min with PerfectHyb buffer (Sigma-Aldrich). The radioactively labeled probe was first prepared with γ[^{32}P] ATP and DNA oligonucleotides listed in Supplementary Table S5 with a T4 polynucleotide kinase kit (Promega), purified on MicroSpin G-25 (GE Healthcare) column and denatured 2 min at 95 °C. The RNAs on the membrane were hybridized with the probe overnight at 40 °C in PerfectHyb buffer and washed three times with a 2X SSC 0.1% SDS solution for 5 min. Radioactive signals were imaged using a Typhoon phosphorimager (Amersham). Contrast adjustments were applied to blot hybridized with individual probes in a linear fashion and faithfully represent original signal information. The position of the 21 nt marker was determined by superposing images for the xisRNA blots with the corresponding one for miRNA159. Before hybridization with a new probe, northern membranes were stripped with three 10 min washes in near-boiling 0.1% SDS and re-equilibrated in 2 × SSC.

For RT-qPCR, after contaminant DNA removal with the TURBO DNA-free Kit (ThermoFisher), 1 µg RNA was reverse transcribed into cDNA using SuperScriptIII (ThermoFisher) and random hexamers as specified by the manufacturer. The reactions were diluted 5 times in water before quantitative real-time PCR on a LightCycler LC480 (Roche), using SYBR-based Mesa Blue qPCR Mastermix (Eurogentec) and the oligonucleotides primers listed in Supplementary Table S5. Transcript levels in a sample were calculated using the ΔΔCt method from two technical replicate reactions from the same cDNA preparation using the *OsEF-1α* gene for normalization.

Deep sequencing data primary analysis. All primary analysis steps were conducted separately for each individual dataset. The first stage of our sRNA sequencing data processing pipeline included a 3′ adaptor trimming step performed with AdapterRemoval (v. 2.2.2—https://github.com/MikkelSchubert/adapterremoval) using these key parameters : "--mm 5 --minadapteroverlap 3 --minlength 18 --maxlength 50". Illumina reads were mapped onto the nuclear rice genome with bwa aln[65] version 0.7.17 with default parameters. sRNA loci inference was performed with ShortStack (v. 3.8.5)[23] using the following key parameters: "--dicermin 20 --dicermax 24 --mincov 2 rpm --pad 75 --foldsize 400 --strand_cutoff 0.8". Read counts were obtained using the summarizeOverlaps function from the GenomicAlignments Bioconductor package. Experimental sRNA loci that accumulated less than 30 total reads across all libraries in the dataset were excluded. Likewise, sRNA loci that overlapped with rice loci with feature types "tRNA", "rRNA" or "pseudogenic_tRNA" in the gff annotation file for NCBI RefSeq assembly accession GCF_001433935.1 were excluded from differential expression (DE) analysis.

For replicated datasets, standard DE analysis was performed with the DESeq2 R Bioconductor package[66]. For normalization, size factors were estimated using the "iterate" method and the "shorth" function to compute a location for a sample. Dispersions were fitted using the "parametric", by default, or the "local" method, in case estimates did not converge. Two-tailed Wald test were used for DE tests where the alternative hypotheses was that the absolute value of the log2 transform of the fold change ratio is greater than 2. Adjusted p-values were computed with the Benjamini and Hochberg method and considered significant when equal or below 0.05.

For unreplicated datasets, DE analysis used the DESeq package[67] with default parameters except that a "blind" value for the method parameter and a value "fit-only" for the sharingMode parameter were applied instead in the calls to the estimateDispersions function. As above, dispersions were fitted using the "parametric", by default, or the "local" method, in case estimates did not converge. Benjamini and Hochberg adjusted p-values were computed for the null hypothesis that the value of the log2 transform of the fold change ratio is equal to 0. The latest version of our primary sRNA processing workflow was written into a Snakemake pipeline and is available on gitHub (https://github.com/Aucomte/sRNAmake).

For the analysis of differential gene expression using mRNA-seq data, we applied the ARMOR pipeline[68] with the following parameter values specified in the config file: additional_salmon_index: "-k 31", additional_salmon_quant: "--seqBias --gcBias". Note that whenever available in the sample metadata information, a 'experiment' term was included in the statistical model for sRNA loci and gene DE analysis in order to block for experimental batch effects.

Secondary bioinformatic analysis. Rice Nipponbare genomic data[24] including genome sequence, gene annotation and transcripts sequences were downloaded from the MSU website. MIRNA annotation was obtained from MirBase (v22)[69]. Unless otherwise stated, pri-miRNAs loci considered in these analysis correspond to those reference rice sRNA loci annotations in the Plant small RNA gene annotations database[46] that are flagged as "MIRNA" ("type" field), that are also listed in miRBase ("source" field has value "miRBase21") and for which an overlapping experimental sRNA loci has been identified in all the datasets considered in this study.

We performed Singular Enrichment Analysis for GO Slim Gene ontology terms in the "Rice MSU7.0 nonTE gene" dataset on the AgriGO website (http://bioinfo.cau.edu.cn/agriGO/—date: 10/09/2020) with the list of *cis*-gene identifiers and default parameter values (Fisher statistical test, Yekutieli p-value correction).

xisRNA *cis*-gene kinase annotation in Fig. 3 and Supplementary Table S3 was based on the kinase classification of Lehti-Shiu and Shiu[9] and the rice OsRLCK gene family classification of Vij et al.[70].

To build the rice and Arabidopsis kinase phylogenetic tree available in Supplementary Data S1, the hmmsearch program[71] was used to search for Hidden Markov Model (HMM) kinase profile (PF00560) into the rice MSU7 and Arabidopsis TAIR10 proteomes. Kinase domain sequences were aligned using MAFFT[72]. This alignment was cleaned with TrimAl configured to remove sites with more than 80% of gaps[73]. Based on this alignment, a phylogenetic tree was computed with the Fasttree method (approximately-maximum-likelihood method)[74].

The search for domain hits in the InterPro database[75] used InterProScan linux executable and dataset (v. 5.41-78.0) as well as the Panther data (v. 14.1) with default parameters on the set of MSU7 annotated protein with an underlying xisRNA duplex region fully included within the coding sequence.

The search for major xisRNA reads similarities in bacterial genomes used the blastn tool of the BLAST + suite (v. 2.10.1) with these relevant parameters: "-task \"blastn-short\" -max_target_seqs 5 -perc_identity 80". The blast database was created with the following strain and NCBI assembly accessions: BAI3 (GCA_003031385.1), MAI1 (GCA_003031365.1), BLS256 (GCA_000168315.3), BAI11 (GCA_000940825.1).

The secondary bioinformatic and statistical analysis that generated the graphics and results included in this study have been essentially conducted using the R Statistical programming language and packages, notably from the Bioconductor project[76]. The corresponding code and output files have been deposited in a dataverse with the https://doi.org/10.23708/RXIXM3.

Data availability

Deep sequencing data was downloaded from SRA using the following NCBI BioProject accessions: 'MAI1' mRNA-seq dataset (PRJNA427491), 'Xoc' mRNA-seq dataset (PRJNA280380), PXO99 infection time course sRNA-seq dataset (PRJNA252424). The PXO99A sRNA-seq dataset was downloaded from the Plant MPSS databases (https://mpss.danforthcenter.org/web/php/pages/library_info.php?SITE=rice_sRNA&showAll=true). The sequences generated in the course of this study were deposited to SRA under the following NCBI BioProject accessions: 'BAI3' sRNA-seq dataset (PRJNA552731), 'Diversity' sRNA-seq dataset (PRJNA552775), 'BAI3' mRNA-seq dataset (PRJNA679478).

Code availability

The R code used to generate the material described in this study as well as the R session file and additional plots are stored in a dataverse and persistently available at https://doi.org/10.23708/RXIXM3.

Received: 17 June 2021; Accepted: 29 November 2021
Published online: 17 December 2021

References

1. Borges, F. & Martienssen, R. A. The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* **16**, 727–741 (2015).
2. Song, X., Li, Y., Cao, X. & Qi, Y. MicroRNAs and their regulatory roles in plant–environment interactions. *Annu. Rev. Plant Biol.* **70**, 489–525 (2019).
3. Bologna, N. G. & Voinnet, O. The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. *Annu. Rev. Plant Biol.* **65**, 473–503 (2014).
4. Fang, X. & Qi, Y. RNAi in plants: An argonaute-centered view. *Plant Cell* **28**, 272–285 (2016).
5. Axtell, M. J. Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* **64**, 137–159 (2013).
6. Liu, Y., Teng, C., Xia, R. & Meyers, B. C. PhasiRNAs in plants: Their biogenesis, genic sources, and roles in stress responses, development, and reproduction. *Plant Cell* **32**, 3059–3080 (2020).
7. Huang, C.-Y., Wang, H., Hu, P., Hamby, R. & Jin, H. Small RNAs—Big players in plant–microbe interactions. *Cell Host Microbe* **26**, 173–182 (2019).
8. Couto, D. & Zipfel, C. Regulation of pattern recognition receptor signalling in plants. *Nat. Rev. Immunol.* **16**, 537–552 (2016).
9. Lehti-Shiu, M. D. & Shiu, S.-H. Diversity, classifi ation and function of the plant protein kinase superfamily. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2619–2639 (2012).
10. Dievart, A., Gottin, C., Périn, C., Ranwez, V. & Chantret, N. Origin and diversity of plant receptor-like kinases. *Annu. Rev. Plant Biol.* **71**, 131–156 (2020).
11. Liang, X. & Zhou, J.-M. Receptor-like cytoplasmic kinases: Central players in plant receptor kinase-mediated signaling. *Annu. Rev. Plant Biol.* **69**, 267–299 (2018).
12. White, F. F., Potnis, N., Jones, J. B. & Koebnik, R. The type III effectors of *Xanthomonas*. *Mol. Plant Pathol.* **10**, 749–766 (2009).
13. Staiger, D., Korneli, C., Lummer, M. & Navarro, L. Emerging role for RNA-based regulation in plant immunity. *New Phytol.* **197**, 394–404 (2013).
14. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
15. Arikit, S., Zhai, J. & Meyers, B. C. Biogenesis and function of rice small RNAs from non-coding RNA precursors. *Curr. Opin. Plant Biol.* **16**, 170–179 (2013).
16. Kapoor, M. *et al.* Genome-wide identifi ation, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics* **9**, 451 (2008).
17. Niño-Liu, D. O., Ronald, P. C., Bogdanove, A. J. & Niño-Liu, D. *Xanthomonas oryzae* pathovars: Model pathogens of a model crop. *Mol. Plant Pathol.* **7**, 303–324 (2006).
18. Johnson, N. R. & Axtell, M. J. Small RNA warfare: Exploring origins and function of trans-species microRNAs from the parasitic plant Cuscuta. *Curr. Opin. Plant Biol.* **50**, 76–81 (2019).
19. Jiang, G. *et al.* A rice NBS-ARC gene conferring quantitative resistance to bacterial blight is regulated by a pathogen effector-inducible miRNA. *Mol. Plant* https://doi.org/10.1016/j.molp.2020.09.015 (2020).
20. Baldrich, P. & San Segundo, B. MicroRNAs in rice innate immunity. *Rice* **9**, 1–9 (2016).
21. Jia, Y. *et al.* Characteristic dissection of *Xanthomonas oryzae* pv. *oryzae* responsive microRNAs in rice. *Int. J. Mol. Sci.* **21**, 785 (2020).
22. Zhao, Y. T. *et al.* Dynamic and coordinated expression changes of rice small RNAs in response to *Xanthomonas oryzae* pv. *oryzae*. *J. Genet. Genomics* **42**, 625–637 (2015).
23. Axtell, M. J. ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA N. Y. N* **19**, 740–751 (2013).
24. Kawahara, Y. *et al.* Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice N. Y. N* **6**, 4 (2013).
25. Nakano, M. *et al.* Plant MPSS databases: Signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Res.* **34**, D731–D735 (2006).
26. Wu, L. *et al.* Rice MicroRNA effector complexes and targets. *Plant Cell* **21**, 3421–3435 (2009).
27. Liu, B. *et al.* *Oryza sativa* dicer-like4 reveals a key role for small interfering RNA silencing in plant development. *Plant Cell* **19**, 2705–2718 (2007).
28. Nagasaki, H. *et al.* The small interfering RNA production pathway is required for shoot meristem initiation in rice. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14867–14871 (2007).
29. Lehti-Shiu, M. D., Zou, C., Hanada, K. & Shiu, S.-H. Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol.* **150**, 12–26 (2009).

30. Lu, D. *et al.* A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 496–501 (2010).
31. Ao, Y. *et al.* OsCERK1 and OsRLCK176 play important roles in peptidoglycan and chitin signaling in rice innate immunity. *Plant J. Cell Mol. Biol.* **80**, 1072–1084 (2014).
32. Zhou, X. *et al.* Four receptor-like cytoplasmic kinases regulate development and immunity in rice. *Plant Cell Environ.* **39**, 1381–1392 (2016).
33. Li, Z. *et al.* OsRLCK 57, OsRLCK107 and OsRLCK118 positively regulate chitin- and PGN-induced immunity in rice. *Rice* **10** (2017).
34. Pottinger, S. E. & Innes, R. W. RPS5-mediated disease resistance: Fundamental insights and translational applications. *Annu. Rev. Phytopathol.* **58**, 139–160 (2020).
35. Yamaguchi, K. *et al.* A receptor-like cytoplasmic kinase targeted by a plant pathogen effector is directly phosphorylated by the chitin receptor and mediates rice immunity. *Cell Host Microbe* **13**, 347–357 (2013).
36. Kimura, S. *et al.* CRK2 and C-terminal phosphorylation of NADPH oxidase RBOHD regulate reactive oxygen species production in Arabidopsis. *Plant Cell* **32**, 1063–1080 (2020).
37. Li, X. *et al.* Protein phosphorylation dynamics under carbon/nitrogen-nutrient stress and identifi ation of a cell death-related receptor-like kinase in Arabidopsis. *Front. Plant Sci.* **11**, 377 (2020).
38. Liu, B. *et al.* Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. *Plant Physiol.* **139**, 296–305 (2005).
39. Zhang, D. *et al.* Repression of microRNA biogenesis by silencing of OsDCL1 activates the basal resistance to *Magnaporthe oryzae* in rice. *Plant Sci.* **237**, 24–32 (2015).
40. Abe, M. *et al.* WAVY LEAF1, an ortholog of Arabidopsis HEN1, regulates shoot development by maintaining MicroRNA and trans-acting small interfering RNA accumulation in rice. *Plant Physiol.* **154**, 1335–1346 (2010).
41. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
42. Pérez-Quintero, A. L. *et al.* An Improved method for TAL effectors DNA-binding sites prediction reveals functional convergence in TAL repertoires of *Xanthomonas oryzae* strains. *PLoS One* **8** (2013).
43. Ji, Z. *et al.* Interfering TAL effectors of *Xanthomonas oryzae* neutralize R-gene-mediated plant disease resistance. *Nat. Commun.* **7**, 13435 (2016).
44. Tran, T. T. *et al.* Functional analysis of African *Xanthomonas oryzae* pv. *oryzae* TALomes reveals a new susceptibility gene in bacterial leaf blight of rice. *PLoS Pathog.* **14**, e1007092 (2018).
45. Wilkins, K. E., Booher, N. J., Wang, L. & Bogdanove, A. J. TAL effectors and activation of predicted host targets distinguish Asian from African strains of the rice pathogen *Xanthomonas oryzae* pv. *oryzicola* while strict conservation suggests universal importance of five TAL effectors. *Front. Plant Sci.* **6**, 536 (2015).
46. Lunardon, A. *et al.* Integrated annotations and analyses of small RNA-producing loci from 47 diverse plants. *Genome Res.* **30**, 497–513 (2020).
47. Hardcastle, T. J., Müller, S. Y. & Baulcombe, D. C. Towards annotating the plant epigenome: The Arabidopsis thaliana small RNA locus map. *Sci. Rep.* **8**, 6338 (2018).
48. Polydore, S. & Axtell, M. J. Analysis of RDR1/RDR2/RDR6-independent small RNAs in Arabidopsis thaliana improves MIRNA annotations and reveals unexplained types of short interfering RNA loci. *Plant J. Cell Mol. Biol.* **94**, 1051–1063 (2018).
49. Zhang, X. *et al.* Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. *Genome Biol.* **13**, R20 (2012).
50. Singh, J., Mishra, V., Wang, F., Huang, H.-Y. & Pikaard, C. S. Reaction mechanisms of Pol IV, RDR2, and DCL3 drive RNA channeling in the siRNA-directed DNA methylation pathway. *Mol. Cell* **75**, 576-589.e5 (2019).
51. Wang, F., Johnson, N. R., Coruh, C. & Axtell, M. J. Genome-wide analysis of single non-templated nucleotides in plant endogenous siRNAs and miRNAs. *Nucleic Acids Res.* **44**, 7395–7405 (2016).
52. Wang, X. *et al.* Synergistic and independent actions of multiple terminal nucleotidyl transferases in the 3′ tailing of small RNAs in Arabidopsis. *PLoS Genet.* **11**, e1005091 (2015).
53. Tomita, K. & Yamashita, S. Molecular mechanisms of template-independent RNA polymerization by tRNA nucleotidyltransferases. *Front. Genet.* **5**, 36 (2014).
54. Wu, H. *et al.* Plant 22-nt siRNAs mediate translational repression and stress adaptation. *Nature* **581**, 89–93 (2020).
55. Johnson, N. R., de Pamphilis, C. W. & Axtell, M. J. Compensatory sequence variation between trans-species small RNAs and their target sites. *eLife* **8**, e49750 (2019).
56. Thilmony, R., Underwood, W. & He, S. Y. Genome-wide transcriptional analysis of the *Arabidopsis thaliana* interaction with the plant pathogen *Pseudomonas syringae* pv. tomato DC3000 and the human pathogen *Escherichia coli* O157:H7. *Plant J. Cell Mol. Biol.* **46**, 34–53 (2006).
57. Navarro, L. *et al.* The transcriptional innate immune response to flg22. Interplay and overlap with Avr gene-dependent defense responses and bacterial pathogenesis. *Plant Physiol.* **135**, 1113–1128 (2004).
58. Cernadas, R. A. *et al.* Code-assisted discovery of TAL effector targets in bacterial leaf streak of rice reveals contrast with bacterial blight and a novel susceptibility gene. *PLoS Pathog.* **10**, e1003972 (2014).
59. Toda, Y. *et al. Oryza sativa* H+-ATPase (OSA) is involved in the regulation of dumbbell-shaped guard cells of rice. *Plant Cell Physiol.* **57**, 1220–1230 (2016).
60. Elmore, J. M. & Coaker, G. The role of the plasma membrane H+-ATPase in plant-microbe interactions. *Mol. Plant* **4**, 416–427 (2011).
61. Singla-Rastogi, M. *et al.* Plant small RNA species direct gene silencing in pathogenic bacteria as well as disease protection. *bioRxiv* 863902 (2019). https://doi.org/10.1101/863902
62. Teng, Y. *et al.* Plant-derived exosomal microRNAs shape the gut microbiota. *Cell Host Microbe* **24**, 637-652.e8 (2018).
63. Romero, L. E. *et al.* Major QTLs control resistance to rice hoja blanca virus and its vector *Tagosodes orizicolus*. *G3* (*Bethesda, Md*) **4**, 133–42 (2014).
64. Blevins, T. Northern blotting techniques for small RNAs. in *Plant Epigenetics: Methods and Protocols* (eds. Kovalchuk, I. & Zemp, F. J.) 87–107 (Humana Press, 2010). https://doi.org/10.1007/978-1-60761-646-7_9
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics (Oxf., Engl.)* **25**, 1754–1760 (2009).
66. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
67. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
68. Orjuela, S., Huang, R., Hembach, K. M., Robinson, M. D. & Soneson, C. ARMOR: An Automated Reproducible MOdular Workfl w for Preprocessing and Differential Analysis of RNA-seq Data. *G3* (*Bethesda, Md*) **9**, 2089–2096 (2019).
69. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: From microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).
70. Vij, S., Giri, J., Dansana, P. K., Kapoor, S. & Tyagi, A. K. The receptor-like cytoplasmic kinase (OsRLCK) gene family in rice: Organization, phylogenetic relationship, and expression during development and stress. *Mol. Plant* **1**, 732–750 (2008).

71. Eddy, S. R. A new generation of homology search tools based on probabilistic inference. *Genome Inform. Int. Conf. Genome Inform.* **23**, 205–211 (2009).
72. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
73. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxf., Engl.)* **25**, 1972–1973 (2009).
74. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
75. Mitchell, A. L. *et al.* InterPro in 2019: Improving coverage, classifi ation and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).
76. Gentleman, R. C. *et al.* Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

Acknowledgements

GR and AP-Q and were supported by a doctoral fellowships awarded by the French Ministry of Research and Higher Education (Ministére de l'Enseignement Supérieur et de la Recherche) and the Erasmus Mundus Action 2 PANACEA, PRECIOSA program of the European Community, respectively. JMJ was supported by an NSF Postdoctoral Fellowship in Biology (DBI-1306196). Th s work was funded by the ANR with a grant (ANR-14-CE19-0009-01) to SC. The authors acknowledge the IRD itrop (https://bioinfo.ird.fr/) HPC platform at IRD Montpellier for providing HPC resources. We thank Dr. Blake Meyers and Dr. Bing Yang for granting access to their sRNA deep sequencing dataset in the Plant MPSS databases. We are particularly grateful to our colleagues who shared seeds of rice genetic material used in this study: Dr. Bo Zhou for the OsDCL1-IR transgenics, Dr. Xiaofeng Cao for the *dcl4-1* line and Dr. Itoh Jun-Ichi for the Kinmaze variety and *waf1-2* mutant.

Author contributions

G.R., J.M.J., F.A., C.S., L.C., C.M., A.L.P.Q., E.T. performed lab experiments. A.C., A.D., S.C. developed computer code. A.B., S.L. contributed materials. G.R., J.M.J., F.A., A.D., C.B., S.L., S.C. analysis and interpretation of results. G.R., J.M.J., A.L.P.Q., R.K., B.S., C.B., S.L., S.C. study conception and design. S.C. wrote an initial version of the manuscript that was subsequently critically revised by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-03391-9.

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.