Population genomics of the rice landrace Acuce reveals exceptional dynamics of immune receptors

Stéphane De Mita[2,6], Shengchang Duan[1,3,6], Xiang Li[1,3,6], Rémi Pélissier[2], Laura Mathieu[2], Jingyu Li[1,3], Linna Ma[1,3], Zihang Li[1,3], Jingjing Liao[1,3], Xiahong He[1,5], Youyong Zhu[1,3], Chengyun Li[1,3,*], Yang Dong[1,4,*], Jean-Benoit Morel[2,*], Huichuan Huang[1,3,*]

[1]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming, China

[2]PHIM Plant Health Institute, Université de Montpellier, INRAE, CIRAD, Institut Agro, IRD, Montpellier, France

[3]Key Laboratory for Agro-Biodiversity and Pest Control of Ministry of Education, Yunnan Agricultural University, Kunming, China

[4]Yunnan Research Institute for Local Plateau Agriculture and Industry, Kunming 650201, China

[5]Key Laboratory of Forest Resources Conservation and Utilization in the Southwest Mountains of China Ministry of Education, Southwest Forestry University, Kunming, Yunnan, China

[6]These authors contributed equally: Stéphane De Mita, Shengchang Duan and Xiang Li

[*]: Correspondence:

Chengyun Li, lichengyun@ynau.edu.cn

Yang Dong, loyalyang@163.com

Jean-Benoit Morel, jean-benoit.morel@inrae.fr

Huichuan Huang, absklhhc@gmail.com

ORCID

Shengchang Duan: 0000-0003-2728-8406

Xiang Li: 0000-0003-4913-2162

Rémi Pélissier: 0000-0001-7168-4727
Laura Mathieu: 0009-0003-4083-1298
Jean-Benoit Morel: 0000-0003-1988-956X
Huichuan Huang: 0000-0002-8400-7116

Keywords
landrace, diversity, immune receptor, pathogen resistance, *Oryza sativa*

Summary

In cultivated plants such as *Oryza sativa*, landraces harbour high levels of diversity, in sharp contrast with genetically uniform modern varieties. The Yuanyang terraces (China) are renowned for traditional farming involving landraces and low disease incidence. To understand the molecular basis of these resistance levels, we explored the genomic diversity of Acuce, one of the oldest landraces in the region. Based on whole-genome resequencing of over 200 Acuce plants, we evaluated the diversity of immunity and agronomical genes with respect to the rest of the genome. In parallel, we analysed the genetic basis of pathogen resistance and agronomical traits. Acuce exhibits high levels of diversity, reaching 38% of the worldwide levels. We uncovered an excess of diversity and signatures of directional selection associated with immunity genes, in contrast with agricultural traits. Furthermore, cultures of mixed genotypes exhibited superior performances for both pathogen resistance and agricultural traits. We showed that, within one single landrace, traditional farmer practices selected optimal agricultural traits while maintaining diversity enabling pathogen resistance. This study elucidates the complex interplay between genetic diversity, natural selection related to pathogens, and selection for agricultural traits within traditional landraces, providing insights into sustainable agricultural practices in diverse cropping systems.

Introduction

Landraces represent a traditional way of maintaining diversity in cultivated systems (McCouch et al., 2020). While they do not result from deliberate improvement, landraces are both locally adapted and genetically diverse, but are still identified and unambiguously delineated by farmers (Harlan, 1975; Villa et al., 2005). Local or regional landraces of rice (*Oryza sativa*) have been analysed to some extent in Borneo and Madagascar islands (Mather et al., 2010; Thomson et al., 2009). However, there is only limited knowledge on how diversity is structured within individual landraces, and none at the genomic scale (Pusadee et al., 2009). Specifically, it is not known whether individual landraces are homogeneous or, instead, harbour substructure with internally differentiated groups. Moreover, the molecular functions that are genetically diversified in landraces are not known. Furthermore, the question of whether landraces can be engineered in modern selection programs has not been addressed to date.

The Yuanyang terraces (Yunnan Province, China, hereafter YYT) were shaped by the agricultural practices of the Hani ethnic group (Wang, 1999). They were classified as a "cultural landcape" by UNESCO in 2013 and are known for being home to a wide range of rice landraces (Jiao et al., 2012). Quite importantly, the available records indicate that the YYT only suffers 1% losses caused by disease (Li, 1990), dramatically less than the 32% losses at the world level (Savary et al., 2019). High genetic diversity of some YYT landraces has been documented in previous reports using a limited number of molecular markers (Gao et al., 2012; Zhang et al., 2007), along with signatures of substructure. Recently, an analysis combining RenSeq targetting genes involved in immunity and genotyping-by-sequencing, has shown extensive levels variation and signatures of balancing selection in immunity genes within four landraces cultivated in the YYT and belonging to the XI group (standing for Xian / Indica) of *Oryza sativa* (Gladieux et al., 2024). Among them, Acuce (literally “God rice” in the Hani language and “Moon rice” in Mandarin) is a prevalent landrace in YYT (Ali et al., 2023; Alonso et al., 2019). It has been grown in YYT for more than 130 years (Figure S1A) and appears to have remained immune to the major fungal pathogen *Pyricularia oryzae* (Ali et al., 2023). Although cultural practices in YYT can be hypothesized to provide an efficient disease control, it is not clear whether this effect is due to a particular content of genes mediating immunity. A genome-wide survey of nucleotide diversity is required to

analyse the structure of diversity within landraces and the effect of disease-mediated selective forces on genes involved in immunity. Within the YYT, the genomic diversity is analysed across 91 landraces by Huang et al. (2025) and here we focus on Acuce to provide a detailed survey of diversity occurring within a single landrace. The Acuce landrace offers a unique opportunity to understand how the definition of a landrace by farmers is consistent with the structuring of genomic diversity, the molecular basis of stable and robust pathogen resistance, and the phenotypic and genotypic traits through which high intra-landrace diversity is expressed.

In this study, we relied on over 200 Acuce plants sampled from various villages of the YYT (Figure S1B) to analyse the structure of genomic diversity based on Illumina genome resequencing. These data were compared with the genome of representatives from other landraces of the YYT and accessions from worldwide collections (Figure S1E). We conducted a detailed analysis of genomic diversity structure, compared regions involved in pathogen resistance with the rest of the genome and analysed footprints of selection. In addition, we provide hints about genotype-phenotype relationship by examining genetic variance of disease resistance traits in comparison to agronomical traits. We put results concerning Acuce in perspective by examining a variety obtained through modern breeding, HongYang 3 (hereafter HY3; Figure S1C), which was introduced in 2014 in the YYT.

Material and Methods

Sampling of Acuce, HY3 and other rice genomes

Acuce and HY3 plants were sampled from a small sector (~8×13 km) in the Yuanyang County (Yunnan, China; Figure S1B). This area is similar to that described by (Jiao et al., 2012). A total of 227 Acuce and 30 HY3 plants were included in the sample. The 227 Acuce samples include 89 single-seed descents that have been propagated for 8 years starting from seeds collected in Qingkou village. This samples are referred to as inbred. All other Acuce and HY3 were collected in 2015 and 2016 (Figure S1D). The landrace or variety status was based on input from local farmers. In order to clarify the organization of the diversity with Acuce with respect to larger groups, we included 252 samples from Huang et al. (2025): 192 belonging to YYT clusters (including 4 Acuce samples) and 60 genomes representative of rice world diversity (50 from the XI group and 10 from the GJ group, the latter standing for Geng /

Japonica) (Wang et al., 2018). Overall, 509 genomes were included in the analysis (Table S1), and *Oryza barthii* (accession number: 65489_Ob_516W1) (Cubry et al. , 2018) was used as an outgroup.

Field phenotyping

The 89 inbred Acuce lines were sown and transplanted as pure stands in three separate fields in Qingkou village in 2022 (Figure S3A). Plant height, tiller number, and thousand-seed weight were measured in the fields. Four indicators associated with micro-organism associated diseases (blast, neck blast, brown spot, and smut) were surveyed using a grading scale ranging from level 0 to level 5 during the field survey. Subsequently, we converted the disease grades to disease indices by adopting the method from Chiang et al. (2017). Six Acuce inbred lines were selected (A15, A36, A49, A69, A74, and A87) and assembled into 6 binary mixtures. They were evaluated for three years. For each mixture, one line was considered as focal and was scored for phenotype, the other line(s) of the mixture being treated as neighbours. For each focal line and year, the number of plants scored for establishing phenotypes ranged from 195 to 441. We used linear modelling in R to scrutinize the impact of year, field, pure/mix, and focal plant on the measured metrics using the model equation: log(1+value) ~ year + field + condition × focal. Subsequently, the R package lsmeans was utilized to compute the least squares means for plotting. The p-values were extracted using lsmeans and the pairs function in R. The R package FactoMineR (Version: 2.9) (Lê et al., 2008) was employed for PCA analysis with default parameters. For correlation calculations using Pearson's correlation coefficient, we utilized the R package WGCNA (Version: 1.72.1) (Langfelder and Horvath, 2008). The *t*-test was applied to examine the significance of differences in data adhering to a normal distribution, while the Wilcoxon test was employed for the remaining non-normally distributed data.

DNA sequencing and mapping on reference genome

Young leaves were collected from plants and snap-frozen in liquid nitrogen. Total DNA was extracted uisng the DNAsecure plant kit (TIANGEN, Beijing). For each accession, 2 μg of genomic DNA were used to construct a sequencing library following the manufacturer's instructions using the NEBNext Ultra DNA Library Prep Kit (NEB Inc., USA). Paired-end sequencing libraries with an insert size of approximately 400 bp were sequenced on an

Illumina HiSeq 4000 sequencer. The average coverage per genome was 14.5X (Table S11). Paired-end resequencing reads were cleaned using fastp (Version: 0.12.2) (Chen et al., 2018). Cleaned paired-end reads were mapped to the XI rice genome Shuhui498 (alternatively named R498) reference genome (Du et al., 2017) with BWA (Version: 0.7.10-r789) (Li and Durbin, 2009). SAMtools (Version: 1.3.1) (Danecek et al., 2021) was used to convert the mapping results into BAM format. Duplicated reads were marked with the Picard package (picard.sourceforge.net, Version: 2.1.1) after sorting BAM files using the Picard package. After BWA alignment, the reads around indels were realigned, realignment was performed with the Genome Analysis Toolkit (GATK, Version: 3.3-0-g37228af) (McKenna et al., 2010) in two steps. The first step used the RealignerTargetCreator package to identify regions where realignment was needed, and the second step used IndelRealigner to realign the regions identified in the first step, which produced a realigned BAM file for each accession. Variation detection followed the best practice workflow recommended by GATK (McKenna et al., 2010). The SNP variants were called for each accession using GATK HaplotypeCaller. A joint genotyping step for comprehensive variations union was performed on the gVCF files. During the filtering step, the SNP filter expression was set as "QD < 5.0 || MQ < 40.0 || FS > 100.0 || SOR > 3.0 || MQRankSum < -5.0 || ReadPosRankSum < -5.0 || QUAL < 30". SNPs with none bi-allelic were removed. The SNPs were then cleaned for hidden paralogy as described by Huang et al. (2025).

Phylogenetic analysis

The pairwise distance matrix was computed between 510 samples (including *Oryza barthii*) over all SNPs as the proportion of pairwise differences among all SNPs where the given pair of samples have non-missing genotypes. The neighbor-joining tree was computed using the neighbor program of the PHYLIP package (Version: 3.698) (Felsenstein, 1993). *O. barthii* was used as outgroup to root the tree using EggLib (Version: 3.1) (Siol et al., 2022) and the tree representation was realized using the online tool iTol (Version: 5) (Letunic and Bork, 2021).

Structure analysis

Structure analysis was performed using two complementary methods: sNMF (Version: 2.0), as available within the R package LEA (Version: 3.14.0) (Frichot et al., 2014) and Admixture

(Version: 1.3) (Alexander et al., 2009). Both programs were used with K ranging from 1 to 15 and 10 replicates. sNMF was run with the parameters: alpha=10, iterations=1000, tolerance=0.000001 and Admixture with cv=10. The analysis was run on 494 samples, excluding the 10 GJ samples, as well as three Acuce (B73, 396, and A3) and two worldwide (304 and B170) samples that appear at a basal position of the phylogenetic tree, indicating a potential taxonomic error. After thinning, 27,522 diallelic sites were included in the analysis. Based on correlations of assignment coefficients, we used *ad hoc* Python code to map clusters with each other over the replicates, the two methods, and the successive values of K to obtain a consistent set of clusters. Individuals were assigned to the cluster with the highest coefficient if it was at least 0.75, or if the gap with the second-best cluster was larger than 0.20.

Analysis of nucleotide diversity

Only diallelic SNPs were considered in the analysis. Diversity statistics were computed using EggLib (Version: 3.1) (Siol et al., 2022). For each group, we computed the average number of non-missing samples (nseff), number of polymorphic sites (S), total nucleotide diversity (π), average unbiased heterozygosity (He), and Tajima's D (Tajima, 1989), which is an index of bias in the allele frequency spectrum. To measure population differentiation, we computed F_{ST} as in Weir and Cockerham (1984), G'_{ST} (Hedrick, 2005) and Jost's D (Jost, 2008), the latter being designed to be less sensitive than the former two to within-population levels of diversity, and both unstandardized (Dxy) and net (Da) pairwise distance (Nei, 1987).

Acuce genome construction

The Acuce A15 inbred line was used as a reference for PacBio sequencing. A total of 50 mg DNA was used to construct sequencing libraries. Sequencing was conducted on PacBio Sequel platform, and the resulting reads were assembled using MECAT2 (Version: 2019.3.14) (Xiao et al., 2017). The contig was polished using Pilon (Version: 1.21) (Walker et al., 2014) with Illumina reads from the A13 inbred line. For chromosome-scale assembly, Hi-C data were prepared by NowBio Biotechnology Co., Ltd. (Yunnan, China), involving extraction of high molecular weight DNA, digestion with DpnII, biotinylation, dilution, and random ligation of digested fragments. Biotinylated fragments were then enriched and sheared to 300–500 bp for sequencing on Illumina NovaSeq platform. The polished sequences were further scaffolded using Hi-C data based on 3d-dna pipeline (Dudchenko et al., 2017). First, Hi-C

reads were mapped to the polished assembly with juicer (Version: 1.5.6) (Durand et al., 2016), and then a candidate chromosome-length assembly was generated automatically using the 3d-dna pipeline (Version: 180922) (Dudchenko et al., 2017). Manual review and refinement of the candidate assembly was performed using Juicebox Assembly Tools (Version: 1.9.1) (Durand et al., 2016) for quality control and interactive correction. The genome assembly was then reviewed using 3d-dna (Dudchenko et al., 2017), according to manual adjustment. Using the modified 3d-dna and Juicebox workflow, 12 chromosomes were anchored. We obtained an initial 390.72 Mb Acuce genome assembly from 46.84 Gb PacBio single-molecule sequencing data with contig N50=5.67 Mb. After polishing the assembly, we anchored 385.15 Mb (98.43%) sequences into 12 pseudo-chromosomes using 60.84 Gb Hi-C data.

RNA sequencing

An Iso-Seq library for leaf tissue was prepared according to the Isoform Sequencing protocol (Iso-Seq) using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03). The library was then sequenced using the PacBio Sequel System. The sequenced data were managed using the SMRT Analysis software suite (PacBio, Version: release_6.0.0.47841), resulting in a set of full-length transcripts.

Repeat and gene annotation

Tandem repeats across the genome were identified using Tandem Repeat Finder (TRF, Version: 4.09) (Benson, 1999). Transposable elements (TEs) were detected using both homology-based methods, including RepeatMasker (Version: open-4.0.9) (Chen, 2004) against Repbase (Jurka et al., 2005) (Repbase Release 20181026; http://www.girinst.org/repbase/index.html) and de novo approaches such as LTR FINDER (Version: 1.07) (Xu and Wang, 2007) and ltrharvest in GenomeTools (Version: 1.5.10) (Gremme et al., 2013). Results from LTR_retriever (Version: 2.8) (Ou and Jiang, 2018) and RepeatModeler (Version: 2.0, http://repeatmasker.org/) were integrated to enhance TEs predictions.

Gene annotation was consolidated using EVidence Modeler (EVM) (Haas et al., 2008), combining predictions from Augustus (Version: 2.5.5) (Stanke et al., 2008), Genescan (Version: 2015-10-31) (Burge and Karlin, 1997), GlimmerHMM (Version: 3.0.2) (Majoros et

al., 2004), and SNAP (Version: 2006-07-28) (Korf, 2004) with homology-based annotations derived from protein sequences of related *Oryza* species to consolidate ab initio gene predictions and homologous method annotations into a final gene set. For ab initio gene prediction, Augustus (Version: 2.5.5) (Stanke et al., 2008) was run with the trained model of rice. Three additional ab initio gene prediction software were also used: Genescan (Version: 2015-10-31) (Burge and Karlin, 1997), GlimmerHMM (Version: 3.0.2) (Majoros et al., 2004), and SNAP (Version: 2006-07-28) (Korf, 2004). Protein sequences of the accessions Nipponbare (Goff et al., 2002), 93-11 (Yu et al., 2002), Shuhui498 (R498) (Du et al., 2017), *Oryza barthii (*Jacquemin et al., 2013*)*, *O. glaberrima* (Wang et al., 2014), and *O. nivara* (Jacquemin et al., 2013) species were obtained and used for homology-based gene annotation. GeneWise (Version: 2.2.0) (Birney and Durbin, 2000) was then used to predict gene structure within each protein-coding region. Based on all the above-mentioned annotation results, a weighted and non-redundant gene set was generated by merging all of the gene models with Acuce A15 EVM (Version: r2012-06-30625). Finally, the full-length transcripts were fed into PASA pipeline (Version: 2.4.1) (Haas et al., 2008) together with the EVM results for gene structure refinement and alternatively spliced isoform annotation. The genes were filtered if the proportion of repeat sequences in the gene sequences was greater than 40% or if the CDS sequence length was less than 150 bp. Gene functions were assigned according to the best match alignment using blastp against KEGG databases. InterProScan functional analysis and Gene Ontology IDs were obtained using InterProScan (Version: 5.53-87.0) (Zdobnov and Apweiler, 2001).

The pathways to which the genes belonged were derived from the matching genes in KEGG. The completeness of predicted genes was assessed using BUSCO (Version: 5.4.4) (Simão et al., 2015) with embryophyte_odb10 lineage dataset (2020-09-10).

AIGs and NLRs comparisons

The pairwise gene synteny between the R498 reference genome and Acuce line 15 was identified using MCscan (Python version, JCVI package, Version: 1.2.7) (Tang et al., 2008). AIGs were then extracted from Wang et al. (2018). NLRs were identified using full-length Homology-based R-gene Prediction (HRP) pipeline (Andolfo et al., 2022). Because NLRs are known to be highly variable, SNP calling was performed with the Acuce genome as reference according to above method. The nucleotide heterozygosity of each SNP was computed using

VCFtools software (Version: 0.1.13) (Danecek et al., 2021). The nucleotide diversity of a gene was calculated as nucleotide divergence of all SNPs in the gene region divided by gene length. Gene presence/absence variation was characterized using the SGSGeneLoss package (Version: 0.1) (Golicz et al., 2015) based on above sorted BAM files. For a given gene in a given accession, if less than 20% of its exon regions were covered by at least two reads (minCov = 2, lostCutoff = 0.2), this gene was treated as absent in that accession, otherwise it was considered present.

Search for positive selection

SweeD (Version: 3.3.1) (Pavlidis et al., 2013) was applied to SNP matrix to detect selective sweeps based on the CLR test to detect signatures of artificial or natural selection in each rice group with grid number set as chromosome length / 1000. A 50 kb sliding window approach was applied to quantify polymorphism levels (π, pairwise nucleotide variation as a measure of variability) by using VCFtools software (Version: 0.1.13) (Danecek et al., 2021). The impact of NLRs presence on π was assessed using a Wilcoxon test. The fold change was calculated between the average π of regions containing NLRs and those without. These analyses were conducted on both populations, Acuce and XI, for the whole genome and each chromosome. The CLR threshold value was set to select the 0.5%.

Balancing selection

Balancing selection was identified using Beta2 statistic in BetaScan (Siewert and Voight, 2020) applied to SNPs matrix with a sliding window size of 1000 bp, the substitutions and ancestral alleles were called using *O. rufipogon* as outgroup. The mutation rate was calculated as $4 \times Ne \times \mu$ based on the site frequency spectrum of SNPs in putatively neutrally evolving regions of the genome, where μ is the locus neutral mutation rate, Ne is the effective population size. The threshold for Beta2 score was set as top 1%.

GWAS analysis

GWAS was performed on rice brown spot using field data. BLUP (best linear unbiased predictors) were calculated using a generalized linear mixed model implemented in the glmer function from the R package lme4 (Bates et al., 2015), with a logit link function and a binomial distribution of errors:

$$Yijk = \mu + \pi + Lj + PLij + \varepsilon ijk$$

where Yijk was the severity of the rice brown spot from plot i, line j and repetition k, π was the random effect of the plot i, Lj was the random effect of the line j, and PLij was the interaction between the plot i and the line j. A GWAS was then performed with the BLUP values based on SNP set called with Acuce genome using Multiple Loci Mixed linear Model (MLMM) in GAPIT (Lipka et al., 2012). The first five PCA values (eigenvectors) were used as fixed effects in the model to correct for stratification. Considering that 27,522 SNPs were used, we defined the whole-genome significance cutoff with -log10 (P) = 5.

Results

Acuce samples display significant variation for pathogen susceptibility but not for agronomical traits

We first examined the distribution of phenotypic traits expressed by Acuce samples in the field, considering agricultural traits as well as disease resistance. This analysis was conducted using the subset of 89 inbred lines. We observed negative correlations between, on one hand, levels of damages caused by the two main rice pathogens (blast and smut fungi) and, on the other hand, thousand-seed weight, and yield (Figure S2), suggesting that fungal pathogens are potential drivers of plant fitness. This is consistent with current knowledge that pathogens are major, worldwide determinants of plant yield (Savery et al., 2019). As a whole, major agronomic traits such as plant height, tiller number, and thousand-seed weight exhibited low variability in Acuce inbred lines (Figure 1A, B). With respect to these agronomic traits, levels of disease resistance to three major fungal pathogens monitored were more variable. This high variability was also observed under controlled conditions for blast (Table S1), supporting the hypothesis that the observed variability of resistance in the field is due to genetic variation and not to heterogeneous levels of pathogen pressure. Thus, although major traits easily tractable by farmers were quite homogeneous in Acuce, disease resistance traits appeared to be highly variable.

Mixing different genotypes results in superior performances in the field

Under field conditions, we tested the hypothesis that emerging phenotypes in mixtures of Acuce inbred lines could favour the maintenance of genetic diversity by increasing the fitness of the population. Artificially reconstituting diversity by combining two-by-two Acuce inbred lines weakly (0.1 to 0.7%) but significantly increased plant height and thousand-seed weight

(Figure 1C), with some mixtures being more effective than others (Figure S3B). Mixtures tended to reduce disease levels: the incidence of smut (number of plants showing symptoms) was reduced by 23% in mixtures compared to pure stands (Figure S3D), and the average severity (defined as the level of symptoms) was also reduced by 17% (Figure 1C). Due to discretization of source data and smaller dataset size, this result not significant overall. However, it was significant in certain pairs of Acuce lines, with a severity reduced by up to 62% in some cases (Figure S3C). Although disease suppression, in particular incidence levels, is expected in cultivar mixtures (Mundt, 2002), it usually requires a minimum size of the plots to be observed, which was not the case in our experimental design consisting of small plots. Rather, we propose that the mere mixing of Acuce lines generates emergent properties, at the plant level and measured at the severity level, on traits that are key components for yield. This is consistent with observations made under controlled conditions where disease severity to pathogens was modulated by plant-plant interactions in Acuce mixtures (Pélissier et al., 2023).

Population structure at the genome level

The genomes of 227 Acuce (including 89 Acuce inbred lines) and 30 HY3 samples were compared to representative genomes of 192 YYT and 50 worldwide XI accessions (Table S1; Figure S1E). We identified a total of 20,731,390 single nucleotide polymorphisms (SNPs) that were considered for analysis. To reduce the computational burden, only 27,522 representative sites were used for structure analysis. The optimization criterion suggested that K=9 was the most appropriate for biological interpretation (Figure 2A, Figure S4) allowing the identification of three sets of Acuce samples (with each including inbred lines) belonging to different clusters (hereafter called “families”; Table S1; Figure S4C). Phylogenetic analysis grouped accessions according to their cluster assignation (Figure 2B). Acuce accessions were dominant in clusters A and I while many others belonged to cluster E, mixed with many non-Acuce YYT1 accessions. The rest of Acuce accessions, all of them originating from a single village (Tur Guo Zhai), belonged to cluster F, along with YYT1 accessions and one HY3 sample. We ignored this group of accessions because the single-village origin was indicative of a potential labelling error, and we defined the three families of Acuce accessions after the assignation to the A, I, and E clusters. HY3 formed a unique, different group, in agreement with its known pedigree (Figure S1C). In the phylogenetic tree, the three Acuce families and

the YYT4 cluster were imbedded into the YYT1 cluster, and Acuce accessions were included into both YYT1 and YYT4-dominated clades, in agreement with Huang et al. (2025) who found Acuce accessions in both these clusters. Using structure analysis and phylogenetic trees, we discarded samples labelled as Acuce that clearly clustered with non-Acuce landraces as they may result from mislabelling or seed contamination which is common in the YYT (Huang et al., 2025). Similarly, we excluded samples labelled as HY3 that were not associated with the relevant cluster. As a result, we defined conservative groups containing 108 "true" Acuce (excluding inbred), 87 "true" inbred Acuce, and 23 "true" HY3 (Figure S4D). The following analyses were based on these three groups and two sets of reference samples: worldwide XI samples and samples from other YYT landraces (denoted "YYT").

Levels of nucleotide diversity

The comparison of diversity levels between groups showed that diversity within Acuce was as high as 38% of the levels found in the entire XI subspecies (Table S2), consistently with previous studies that showed that landraces harbour high levels of diversity in YYT (Gao et al., 2012). On average, nucleotide diversity values were 28% higher in Acuce than in inbred and 71% higher than in the recently introduced HY3 population. Quite remarkably, the range of differentiation values (estimated with the Jost's D) between Acuce families overlapped those between worldwide XI clusters (Figure S1B and Table S5). There was no strong difference in nucleotide diversity among Acuce families when considering villages or clusters of villages (Table S2). Considering villages with sufficient sample sizes, the differentiation varied from 1.1% to 6.3%, much below differentiation between Acuce families (Table S2 reported in Figure S1B). This small difference between villages is consistent with the strong social habit of exchanging seeds in YYT (Dedeurwaerdere and Hannachi, 2019).

Genome-wide signatures of selection

We performed scans for signatures of selection (either natural or artificial). Pervasive selective sweeps were detected along the genome using SweeD for both Acuce (including inbred), and HY3 groups, targeting genes encoding proteins belonging to a variety of functional categories (Tables S3 and S4). The CLR cut-off values selecting the top 0.5% of regions were as follows: 1795.98 for Acuce family 1, 3263.83 for Acuce family 2, 1725.57 for inbred family 1, 2708.91 for inbred, 1639.96 for Acuce and 2540.68 for HY3. Interestingly,

there was little overlap between the functional categories and genes detected in the two populations. This is expected if the two varieties underwent independent breeding histories (Figure S5).

We identified genes targeted by balancing selection using BetaScan. We found a significant association between genes targeted by balancing selection and gene ontology only in the case of Acuce. All identified ontologies were associated with Nucleotide-binding domain and Leucine-rich Repeat genes (NLRs; Table S6) which are major determinants of pathogen resistance (Ngou et al., 2022). The mode of evolution of NLRs, involving in particular extensive presence/absence variation might blur the signal. Therefore, we performed a specific analysis of NLRs diversity using an *ad hoc* genome reference from one of the Acuce samples using PacBio sequencing (Table S7 and Figure S7). We identified 452 NLRs in this *de novo* Acuce genome (Table S8), a rather small number compared to other known rice genomes (e.g. 712 NLRs were identified in the R498 genome) but consistent with the analysis of Gladieux et al. (2024) who identified 460 NLRs in Acuce. Thirteen NLRs were identified under balancing selection in Acuce, 4 times more than expected by chance (Table S6). None of these NLRs corresponded to known resistance genes, suggesting that Acuce landrace achieves durable resistance with a yet unknown set of rapidly evolving resistance genes. Besides, we performed a GWAS analysis based on field phenotyping data with respect to resistance to several common diseases affecting rice. We identified two SNPs on chromosome 2 nearby a gene coding for a Leucine-rich repeat protein associated with brown spot resistance (Figure S6). This result still needs to be confirmed with further study, but this finding demonstrates the potential of Acuce as a resource for the identification of determinants of disease resistance. Quite interestingly, the recently introduced HY3 modern population also showed some NLRs under balancing selection. This suggests that this pattern of selection is connected to cultural practices applied in the YYT, from where HY3 samples have been collected.

Generation and annotation of a Acuce reference genome

A *de novo* genome assembly for Acuce accession A15 was generated in order to provide a relevant reference for surveying presence-absence variation of NLR genes in Acuce samples (Table S12). The annotation procedure achieved 98.2% completeness of the BUSCO core embryophyta genes (Table S12c). We predicted ~215.41 Mb (55.05%) repetitive elements in

the Acuce genome (Table S12d), among which 212.86 Mb repeats were transposable elements (TEs; Table S12e). We predicted 40,011 protein-coding genes (41,619 transcripts) following a combination of homology and ab initio methods (Table S12f), which captured 96.8% complete BUSCOs in embryophyte_obd10 (Table S12g). 67% (26,803) genes were assigned to at least one function in InterPro, GO or KEGG (Table S12h).

Genomic diversity provides Acuce with diverse immunity factors

We compared the diversity of NLRs to other agronomically important genes (AIGs) thanks to the *de novo* A15 genome. AIGs involved in leaf architecture, tillering and flowering were identified by database homology in the Acuce genome (Table S9). Fifty-seven of them were unambiguously found in Acuce according to the gene synteny and blastp check (Table S9). Overall, NLRs were much more diverse than AIGs in Acuce, inbred Acuce and HY3 (Figure 3A; Table S8). In addition to single polymorphic sites, we looked for presence/absence variations: Acuce had an average of 446 NLRs with a variance of 13.7 (449 for inbred, variance 8) as compared with 435 with a variance of 11.3 for HY3 (Figure 3B; Table S10). There was therefore a slightly elevated number of NLRs in Acuce when compared with HY3. Typical feature of diversity landscape for the Acuce landrace and XI genomes is shown in Figure 3C and Figure 3D, showing large-scale variation in both levels of diversity and presence of NLRs (Figure S8, Table S13). A significant association of diversity with regions containing NLRs was found for both populations (Figure 3E), with a stronger enrichment in Acuce (fold change=1.51) compared to XI genomes (fold change=1.37), for some but not all chromosomes (Figure 3F).

Discussion

Landraces are considered to be untapped reservoirs for future breeding efforts (McCouch et al., 2020). They have long been known to be more successful than modern cultivars under stressful conditions (Dwivedi et al., 2016; Newton et al., 2010). Our in-depth genomic analysis of Acuce showed that, despite visible homogeneity for agricultural traits likely under direct stabilizing selection by farmers, very high levels of genomic diversity were maintained within this landrace. This diversity is specifically inflated in immunity receptors, resulting in high phenotypic variation in disease resistance traits, consistenly with patterns observed for other landraces cultivated in the YYT (Gladieux et al., 2024). The analysis of HY3 indicates

that modern breeding can result in substantial levels of diversity, albeit at a lower level than those observed in the traditional Acuce landrace. Through signatures of balancing selection and the results of GWAS analysis, we provide several lines of evidence hinting that this diversity allows Acuce plants to respond to selection imposed by pathogens, similarly to other YYT landraces (Huang et al., 2025).

The high levels of diversity observed within Acuce can be traced back to the origins of the landrace and may have been maintained continuously by agricultural practices or, alternatively, might be fuelled by recurrent or permanent gene flow, as suggested by the sharing of ancestry groups with other landraces. Deciphering the origins of Acuce diversity requires further study. Our results clearly point to strong levels of artificial stabilizing selection over agricultural traits, both at the genomic and phenotypic levels. In contrast, selection on disease resistance did not exhaust polymorphism of resistance genes. Crop protection against pathogens may be under direct selection by farmers in Acuce, but heterogeneous disease incidence (in both time and space) might prevent complete fixation of single alleles at resistance loci, ultimately allowing for the coexistence of different alleles at immunity receptor genes.

Our results show that, in the field, several traits related to yield, which in some respect may be considered as fitness, are improved by simply mixing two genomes. Diversity at genotypic level has been shown to reduce disease risks and prevalence in agro-systems (Mundt, 2002). For instance, mixing two varieties has been proven to confer resistance to rice blast fungus (Zhu et al., 2000). The combination of two or more resistance genes in separate varieties has been classically invoked to explain this property and the sustainability of such agro-systems (Borg et al., 2018; Garrett et al., 2009). To some extent, disease control was also reduced in some pairwise mixtures tested here. This observation reinforces the hypothesis that direct plant-plant interactions can modulate disease resistance of individual plants (Pélissier et al., 2023; Pélissier et al., 2021) and thus add another possible level of disease regulation in cultivar mixtures besides known processes (Borg et al., 2018). Our experimental design only considered pairwise mixtures, still much simpler than the complex assemblages observed in field condition. In particular, the fitness benefit is likely to depend on the matching between, on the one hand, immunity receptor alleles expressed by genotypes composing the mixture

and, on the other hand, the actual presence of corresponding pathogen genotypes. The recent release of the HY3 modern composite population, which is depleted of a dozen immune receptors, is an opportunity to challenge the stability of disease control achieved by modern populations in the field. In conclusion, the Acuce landrace exhibits extensive genomic diversity, especially in immune receptors, offering potential advantages in disease resistance and crop protection.

Acknowledgements

We thank Elisabeth Fournier and Henry Adreit for their support in collecting plants in the field. This study was supported by National Key R&D Program of China (2023YFE0107500, 2023YED1400800), National Natural Science Foundation of China (31801792, 31960554), Yunnan Fundamental Research Projects (202401AS070002) and Yunnan Xingdian Talent Support Program. It was conducted under the MoBiDiv (https://www6.inrae.fr/mobidiv, funded by the ANR program PPR-CPA ANR-20-PCPA-0006), the SuMCrop INRAE Metaprogram, the ANR MUSE (ANR-16-IDEX-0006 AMUSER) and the INRAE-YAU LIA PLANTOMIX projects.

Competing interests

The authors declare no competing interests.

Author contributions

HH and JBM conceived and designed the study. CL and YZ collected the genetic material. HH, XL and LM processed the samples and performed the DNA extraction. XL, LM, RP, JL and JL phenotyped the inbred lines in the field. YZ and XH contributed to the historical understanding of Yuanyang terraces. YD generated the original sequencing data. SDM, SD, XL, RP, HH and JBM analyzed the data and prepared the figures. LM helped for GWAS analysis and statistical analysis. SDM, SD, XL, YD, HH and JBM wrote the manuscript draft, the co-authors reviewed and edited the manuscript. All authors read and approved the final manuscript.

Data availability

SNP tables of all individuals sequenced in this study after mapping to the two reference genomes are available through an open web server:

- http://teabase.ynau.edu.cn/data/tea/pangeome/rice/Table_S12.vcf.gz,
- http://teabase.ynau.edu.cn/data/tea/pangeome/rice/Table_S14.vcf.gz.

References

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome research **19**:1655-1664. https://doi.org/10.1101/gr.094052.109.

Ali, S., Gladieux, P., Ravel, S., Adreit, H., Meusnier, I., Milazzo, J., Cros-Arteil, S., Bonnot, F., Jin, B., Dumartinet, T., et al. (2023). Evolution of the rice blast pathogen on spatially structured rice landraces maintains multiple generalist fungal lineages. Mol. Ecol. **32**:2519-2533. https://doi.org/10.1111/mec.16927.

Alonso, P., Gladieux, P., Moubset, O., Shih, P.-J., Mournet, P., Frouin, J., Blondin, L., Ferdinand, R., Fernandez, E., Julian, C., et al. (2019). Emergence of Southern Rice Black-Streaked Dwarf Virus in the Centuries-Old Chinese Yuanyang Agrosystem of Rice Landraces. Viruses **11**:985. https://doi.org/10.3390/v11110985.

Andolfo, G., Dohm, J.C., and Himmelbauer, H. (2022). Prediction of NB-LRR resistance genes based on full-length sequence homology. The Plant Journal **110**:1592-1602. https://doi.org/10.1111/tpj.15756.

Bates, D., Mächler, M., Bolker, B., and Walker S. (2015). Fitting linear mixed-effects models ising lme4. Journal of Statistical Software **67**:1-48. https://doi.org/10.18637/jss.v067.i01.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research **27**:573-580. https://doi.org/10.1093/nar/27.2.573.

Birney, E., and Durbin, R. (2000). Using GeneWise in the Drosophila annotation experiment. Genome Research **10**:547-548. https://doi.org/10.1101/gr.10.4.547.

Borg, J., Kiær, L.P., Lecarpentier, C., Goldringer, I., Gauffreteau, A., Saint-Jean, S., Barot, S., and Enjalbert, J. (2018). Unfolding the potential of wheat cultivar mixtures: A meta-analysis perspective and identification of knowledge gaps. Field Crop. Res. **221**:298-313. https://doi.org/10.1016/j.fcr.2017.09.006.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA11Edited by F. E. Cohen. Journal of Molecular Biology **268**:78-94. https://doi.org/10.1006/jmbi.1997.0951.

Chen, N. (2004). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Current Protocols in Bioinformatics **5**:4.10.11-14.10.14. https://doi.org/10.1002/0471250953.bi0410s05.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**:i884-i890. https://doi.org/10.1093/bioinformatics/bty560.

Chiang, K.S., Liu, H.I., and Bock, C.H. (2017). A discussion on disease severity index values. Part I: warning on inherent errors and suggestions to maximise accuracy. Ann. Appl. Biol. **171**:139-154. https://doi.org/10.1111/aab.12362.

Cubry, P., Tranchant-Dubreuil, C., Thuillet, A.C., Monat, C., Ndjiondjop, M.N., Labadie, K., Cruaud, C., Engelen, S., Scarcelli, N., Rhoné, B., Burgarella, C.,

Dupuy, C., Larmande, P., Wincker, P., François, O., Sabot, F., Vigouroux, Y. (2018). The rise and fall of African rice cultivation revealed by analysis of 246 new genomes. Current Biology **28**:2274-2282.e6. https://doi.org/10.1016/j.cub.2018.05.066

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience **10**:giab008. https://doi.org/10.1093/gigascience/giab008.

Dedeurwaerdere, T., and Hannachi, M. (2019). Socio-economic drivers of coexistence of landraces and modern crop varieties in agro-biodiversity rich Yunnan rice fields. Ecological Economics **159**:177-188. https://doi.org/10.1016/j.ecolecon.2019.01.026.

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., Ma, B., Qi, M., Li, Y., Zhao, X., et al. (2017). Sequencing and de novo assembly of a near complete indica rice genome. Nature Communications **8**:15324. https://doi.org/10.1038/ncomms15324.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science **356**:92-95. https://doi.org/10.1126/science.aal3327.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Systems **3**:95-98. https://doi.org/10.1016/j.cels.2016.07.002.

Dwivedi, S.L., Ceccarelli, S., Blair, M.W., Upadhyaya, H.D., Are, A.K., and Ortiz, R. (2016). Landrace Germplasm for Improving Yield and Abiotic Stress Adaptation. Trends in Plant Science **21**:31-42. https://doi.org/10.1016/j.tplants.2015.10.012.

Felsenstein, J. (1993). PHYLIP (phylogeny inference package), version 3.5 c (Joseph Felsenstein.).

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. Genetics **196**:973-983. https://doi.org/10.1534/genetics.113.160572.

Gao, D., Mao, R., and Zhu, Y. (2012). Comparative studies on intra-varietal heterogeneity between Rice landraces and improved varieties. Rice Genomics and Genetics **3**:25-32. https://doi.org/10.5376/rgg.2012.03.0005.

Garrett, K.A., Zúñiga, L.N., Roncal, E., Forbes, G.A., Mundt, C.C., Su, Z., and Nelson, R.J. (2009). Intraspecific functional diversity in hosts and its effect on disease risk across a climatic gradient. Ecol. Appl. **19**:1868-1883. https://doi.org/10.1890/08-0942.1

Gladieux, P., van Oosterhoot, C., Fairhead, S., Jouet, A., Ortiz, D., Ravel, S., Shrestha R.K., He, X., Zhu, Y., Morel, J.B., Huang, H., Kroj, T., and Jones, J.D.G. (2024). Extensive immune receptor repertoire diversity in disease-resistant rice landraces. Curr. Biol. **34**:3983-3995. https://doi.org/10.1016/j.cub.2024.07.061.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., et al. (2002). A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. japonica). Science **296**:92-100. https://doi.org/10.1126/science.1068275.

Golicz, A.A., Martinez, P.A., Zander, M., Patel, D.A., Van De Wouw, A.P., Visendi, P., Fitzgerald, T.L., Edwards, D., and Batley, J. (2015). Gene loss in the fungal canola pathogen Leptosphaeria maculans. Functional & Integrative Genomics **15**:189-196. https://doi.org/10.1007/s10142-014-0412-1.

Gremme, G., Steinbiss, S., and Kurtz, S. (2013). GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. IEEE/ACM

Transactions on Computational Biology and Bioinformatics **10**:645-656. https://doi.org/10.1109/TCBB.2013.68.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biology **9**:R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Harlan, J.R. (1975). Our Vanishing Genetic Resources. Science **188**:618-621. https://doi.org/10.1126/science.188.4188.618.

Hedrick, P.W. (2005). A STANDARDIZED GENETIC DIFFERENTIATION MEASURE. Evolution **59**:1633-1638. https://doi.org/10.1111/j.0014-3820.2005.tb01814.x.

Huang, H., Duan, S., Li, X., Ma, L., He, X., Zhu, Y., Li, C., Dong, Y., Morel, J.-B., and Gladieux, P. (2025). Genome sequencing of rice landraces from the Yuanyang terraces uncovers ancient and diverse lineages of Indica rice. bioRxiv. https://doi.org/10.1101/2025.09.09.675106.

Jacquemin, J., Bhatia, D., Singh, K., and Wing, R.A. (2013). The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Current Opinion in Plant Biology **16**:147-156. https://doi.org/10.1016/j.pbi.2013.02.014.

Jiao, Y., Li, X., Liang, L., Takeuchi, K., Okuro, T., Zhang, D., and Sun, L. (2012). Indigenous ecological knowledge and natural resource management in the cultural landscape of China's Hani Terraces. Ecol. Res. **27**:247-263. https://doi.org/10.1007/s11284-011-0895-3.

Jost, L.O.U. (2008). GST and its relatives do not measure differentiation. Mol. Ecol. **17**:4015-4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research **110**:462-467. https://doi.org/10.1159/000084979.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics **5**:59. https://doi.org/10.1186/1471-2105-5-59.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics **9**:559. https://doi.org/10.1186/1471-2105-9-559.

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software **25**:1 - 18. https://doi.org/10.18637/jss.v025.i01.

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research **49**:W293-W296. https://doi.org/10.1093/nar/gkab301.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**:1754-1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, Y. (1990). Yuanyang county chronicles (Gui Zhou National Press).

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. Bioinformatics **28**:2397-2399. https://doi.org/10.1093/bioinformatics/bts444.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics **20**:2878-2879. https://doi.org/10.1093/bioinformatics/bth315.

Mather, K.A., Molina, J., Flowers, J.M., Rubinstein, S., Rauh, B.L., Lawton-Rauh, A.M.Y., Caicedo, A.L., McNally, K.L., and Purugganan, M.D. (2010). Migration,

isolation and hybridization in island crop populations: the case of Madagascar rice. Mol. Ecol. **19**:4892-4905. https://doi.org/10.1111/j.1365-294X.2010.04845.x.

McCouch, S., Navabi, Z.K., Abberton, M., Anglin, N.L., Barbieri, R.L., Baum, M., Bett, K., Booker, H., Brown, G.L., and Bryan, G.J. (2020). Mobilizing crop biodiversity. Molecular plant **13**:1341-1344. https://doi.org/10.1016/j.molp.2020.08.011.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research **20**:1297-1303. https://doi.org/10.1101/gr.107524.110.

Mundt, C.C. (2002). USE OF MULTILINE CULTIVARS AND CULTIVAR MIXTURES FOR DISEASE MANAGEMENT. Annual Review of Phytopathology **40**:381-410. https://doi.org/10.1146/annurev.phyto.40.011402.113723.

Nei, M. (1987). Molecular evolutionary genetics (Columbia university press).

Newton, A.C., Akar, T., Baresel, J.P., Bebeli, P.J., Bettencourt, E., Bladenopoulos, K.V., Czembor, J.H., Fasoula, D.A., Katsiotis, A., Koutis, K., et al. (2010). Cereal landraces for sustainable agriculture. A review. Agron. Sustain. Dev. **30**:237-269. https://doi.org/10.1051/agro/2009032.

Ngou, B.P.M., Ding, P., and Jones, J.D.G. (2022). Thirty years of resistance: Zig-zag through the plant immune system. The Plant Cell **34**:1447-1478. https://doi.org/10.1093/plcell/koac041.

Ou, S., and Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons Plant Physiology **176**:1410-1422. https://doi.org/10.1104/pp.17.01310.

Pavlidis, P., Živković, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. Molecular Biology and Evolution **30**:2224-2234. https://doi.org/10.1093/molbev/mst112.

Pélissier, R., Violle, C., and Morel, J.-B. (2021). Plant immunity: Good fences make good neighbors? Current Opinion in Plant Biology **62**:102045. https://doi.org/10.1016/j.pbi.2021.102045.

Pélissier, R., Ballini, E., Temple, C., Ducasse, A., Colombo, M., Frouin, J., Qin, X., Huang, H., Jacques, D., Florian, F., et al. (2023). The genetic identity of neighboring plants in intraspecific mixtures modulates disease susceptibility of both wheat and rice. PLOS Biology **21**:e3002287. https://doi.org/10.1371/journal.pbio.3002287.

Pusadee, T., Jamjod, S., Chiang, Y.-C., Rerkasem, B., and Schaal, B.A. (2009). Genetic structure and isolation by distance in a landrace of Thai rice. Proceedings of the National Academy of Sciences **106**:13880-13885. https://doi.org/10.1073/pnas.0906720106.

Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. Nature Ecology & Evolution **3**:430-439. https://doi.org/10.1038/s41559-018-0793-y.

Siewert, K.M., and Voight, B.F. (2020). BetaScan2: Standardized Statistics to Detect Balancing Selection Utilizing Substitution Data. Genome Biology and Evolution **12**:3873-3877. https://doi.org/10.1093/gbe/evaa013.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210-3212. https://doi.org/10.1093/bioinformatics/btv351.

Siol, M., Coudoux, T., Ravel, S., and De Mita, S. (2022). EggLib 3: A python package for population genetics and genomics. Molecular Ecology Resources **22**:3176-3187. https://doi.org/10.1111/1755-0998.13672.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics **24**:637-644. https://doi.org/10.1093/bioinformatics/btn013.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585-595. https://doi.org/10.1093/genetics/123.3.585.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., and Paterson, A.H. (2008). Synteny and Collinearity in Plant Genomes. Science **320**:486-488. https://doi.org/10.1126/science.1153917.

Thomson, M.J., Polato, N.R., Prasetiyono, J., Trijatmiko, K.R., Silitonga, T.S., and McCouch, S.R. (2009). Genetic Diversity of Isolated Populations of Indonesian Landraces of Rice (*Oryza sativa* L.) Collected in East Kalimantan on the Island of Borneo. Rice **2**:80-92. https://doi.org/10.1007/s12284-009-9023-1.

Villa, T.C.C., Maxted, N., Scholten, M., and Ford-Lloyd, B. (2005). Defining and identifying crop landraces. Plant genetic resources **3**:373-384. https://doi.org/10.1079/PGR200591.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE **9**:e112963. https://doi.org/10.1371/journal.pone.0112963.

Wang, M., Yu, Y., Haberer, G., Marri, P.R., Fan, C., Goicoechea, J.L., Zuccolo, A., Song, X., Kudrna, D., Ammiraju, J.S.S., et al. (2014). The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nature Genetics **46**:982-988. https://doi.org/10.1038/ng.3044.

Wang, Q. (1999). On the culture of the terraced fields (Yunnan University Press).

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature **557**:43-49. https://doi.org/10.1038/s41586-018-0063-9.

Weir, B.S., and Cockerham, C.C. (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution **38**:1358-1370. https://doi.org/10.2307/2408641.

Xiao, C.-L., Chen, Y., Xie, S.-Q., Chen, K.-N., Wang, Y., Han, Y., Luo, F., and Xie, Z. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nature Methods **14**:1072-1074. https://doi.org/10.1038/nmeth.4432.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research **35**:W265-W268. https://doi.org/10.1093/nar/gkm286.

Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., et al. (2002). A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. indica). Science **296**:79-92. https://doi.org/10.1126/science.1068037.

Zdobnov, E.M., and Apweiler, R. (2001). InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17**:847-848. https://doi.org/10.1093/bioinformatics/17.9.847.

Zhang, H., Sun, J., Wang, M., Liao, D., Zeng, Y., Shen, S., Yu, P., Mu, P., Wang, X., and Li, Z. (2007). Genetic structure and phylogeography of rice landraces in Yunnan, China, revealed by SSR. Genome **50**:72-83. https://doi.org/10.1139/g06-130.

Zhu, Y., Chen, H., Fan, J., Wang, Y., Li, Y., Chen, J., Fan, J., Yang, S., Hu, L., Leung, H., et al. (2000). Genetic diversity and disease control in rice. Nature **406**:718-722. https://doi.org/10.1038/35021046.

Supporting information

Fig. S1 Origins of Acuce and HY3 genomes.

Fig. S2 Field phenotypes of Acuce inbred lines.

Fig. S3 Phenotypes in binary Acuce mixtures.

Fig. S4 Structure analysis of Acuce and HY3 genomes.

Fig. S5 Separate breeding histories between Acuce and HY3.

Fig. S6 GWAS analysis of brown spot resistance in the Acuce landrace.

Fig. S7 Heatmap of Hi-C interactions for Acuce pseudo-chromosomes.

Fig. S8 Diversity feature of Acuce and XI genomes.

Table S1 Origin and associated metadata for all rice lines used in this study.

Table S2 Diversity statistics per population of Acuce and other local and indica genomes.

Table S3 Genes under purifying selection in true Acuce, Indred and HY3 sub-groups.

Table S4 Gene ontology enrichment for genes under purifying selection.

Table S5 Pairwise comparisons of Auce and other local and indica genomes.

Table S6 Genes under balancing selection in true Acuce, Indred and HY3 sub-groups.

Table S7 Genome annotation of the Acuce line 15, including NLR genes.

Table S8 Nucleotide diversity at AGI and NLR genes.

Table S9 Agronomically important genes (AIG) in Acuce genome.

Table S10 Presence/Absence polymorphism for AIG and NLR genes in Acuce and HY3.

Table S11 Genome statistics.

Table S12 Statistics of the Acuce genome assembly.

Table S13 Table for average diversity.

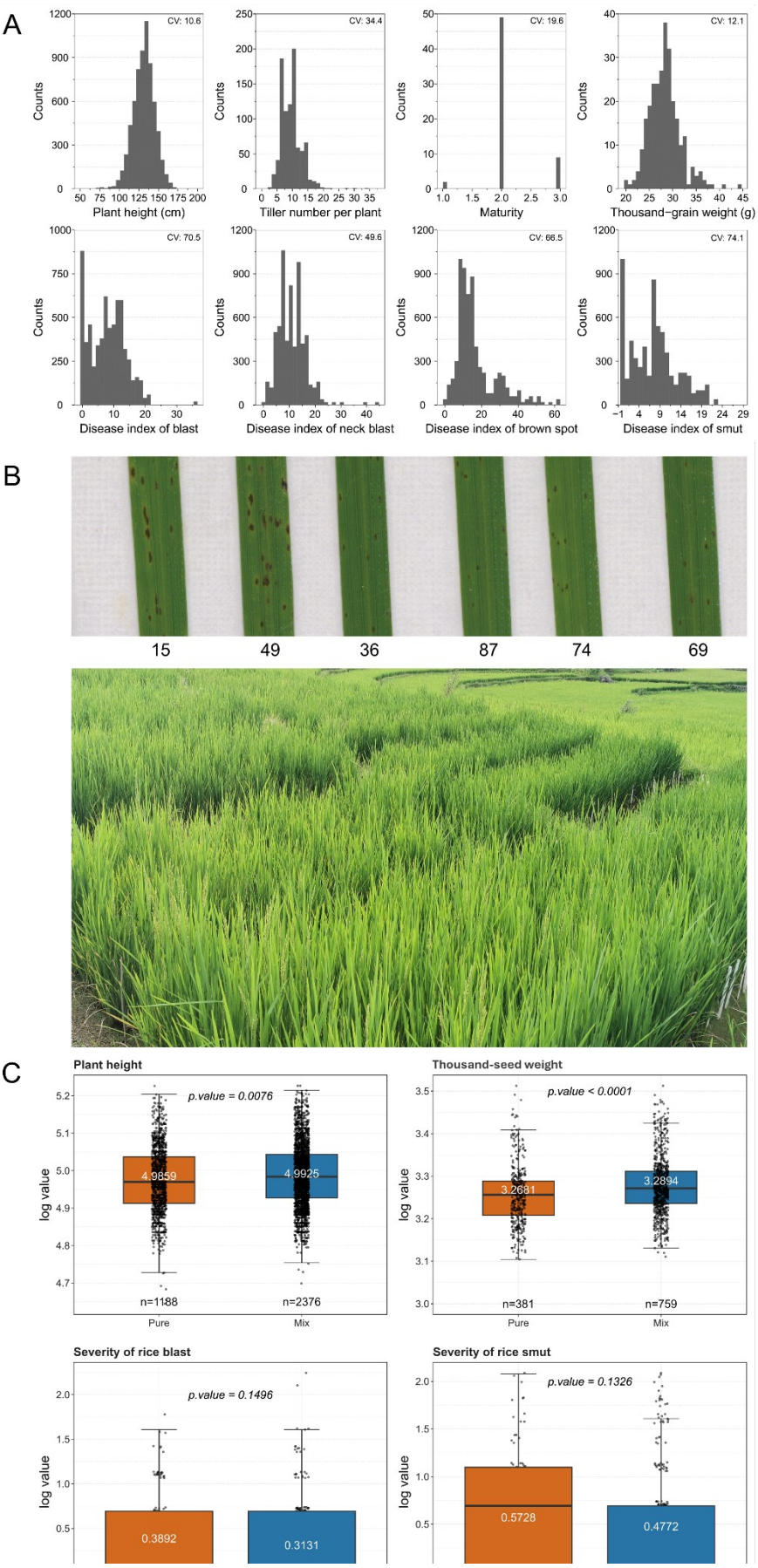

A
CV: 10.6
Counts
Plant height (cm)
CV: 34.4
Tiller number per plant
CV: 19.6
Maturity
CV: 12.1
Thousand−grain weight (g)
CV: 70.5
Disease index of blast
CV: 49.6
Disease index of neck blast
CV: 66.5
Disease index of brown spot
CV: 74.1
Disease index of smut
B
15
49
36
87
74
69
C
Plant height
p.value = 0.0076
log value
4.9859
4.9925
n=1188
n=2376
Pure
Mix
Thousand-seed weight
p.value < 0.0001
3.2681
3.2894
n=381
n=759
Severity of rice blast
p.value = 0.1496
0.3892
0.3131
Severity of rice smut
p.value = 0.1326
0.5728
0.4772

Figure 1. Phenotypic diversity in the Acuce landrace and emerging properties in binary mixtures. Field phenotypes of 89 inbred Acuce lines grown in pure stands (A and B) or in binary mixtures (C). (A) Distribution of growth phenotypes are weakly variable (coefficient of variation values are indicated on each panel) compared to disease phenotypes. (B) Examples of Acuce lines grown in the field displaying some phenotypical differences (forefront) and paddy fields in the background showing the homogeneity of the landraces grown by farmers. The upper panel shows representative symptoms of brown spot under laboratory conditions with the isolate CH9015 used for GWAS analysis (Figure S6). (C) Six binary mixtures of selected Acuce inbred lines were phenotyped in the field for three years. The average values of severity in pure stand and mixtures are shown. Significant differences were evaluated using *t*-test for plant height or thousand-seed weight and Wilcox test for incidence of disease.

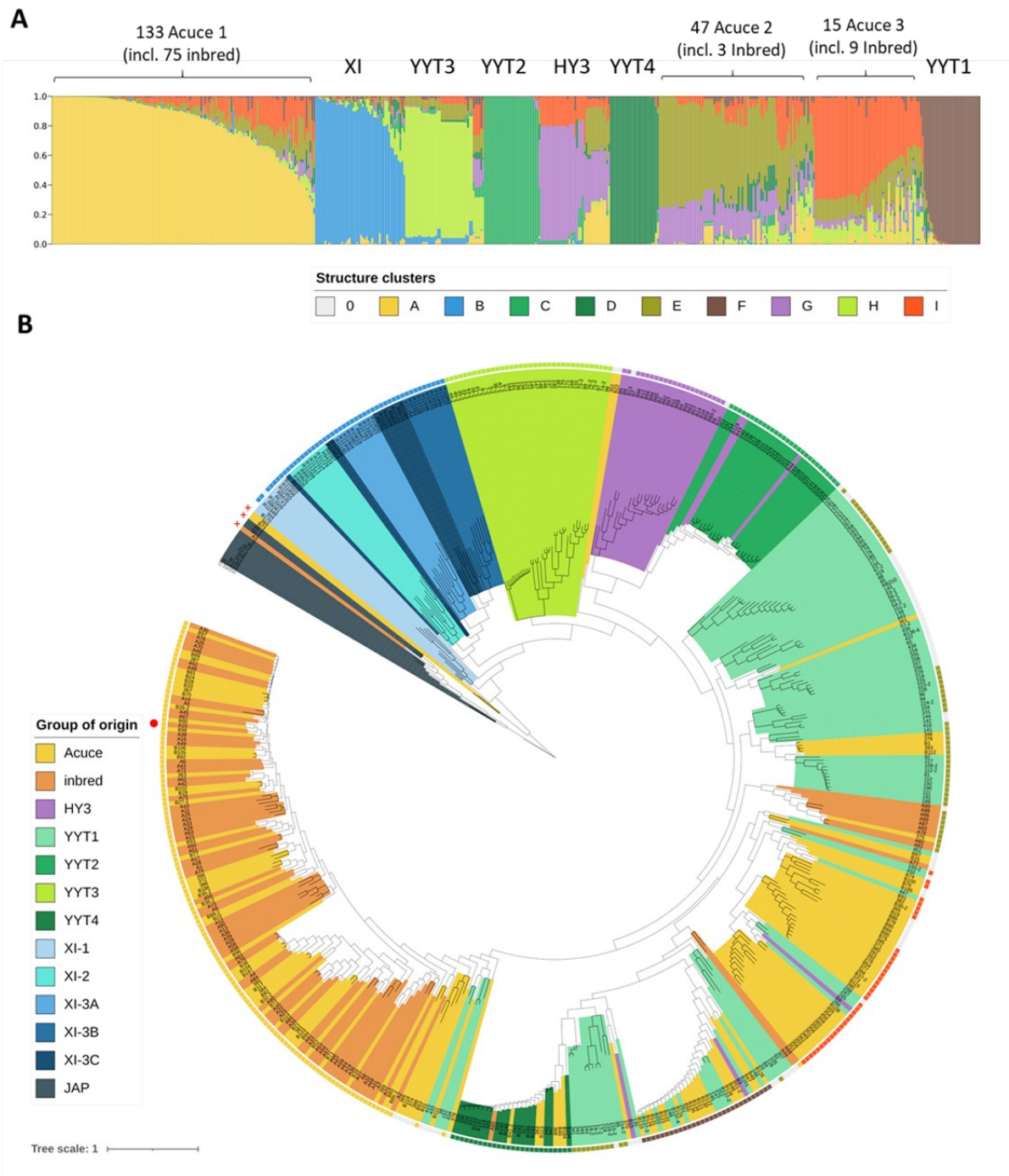
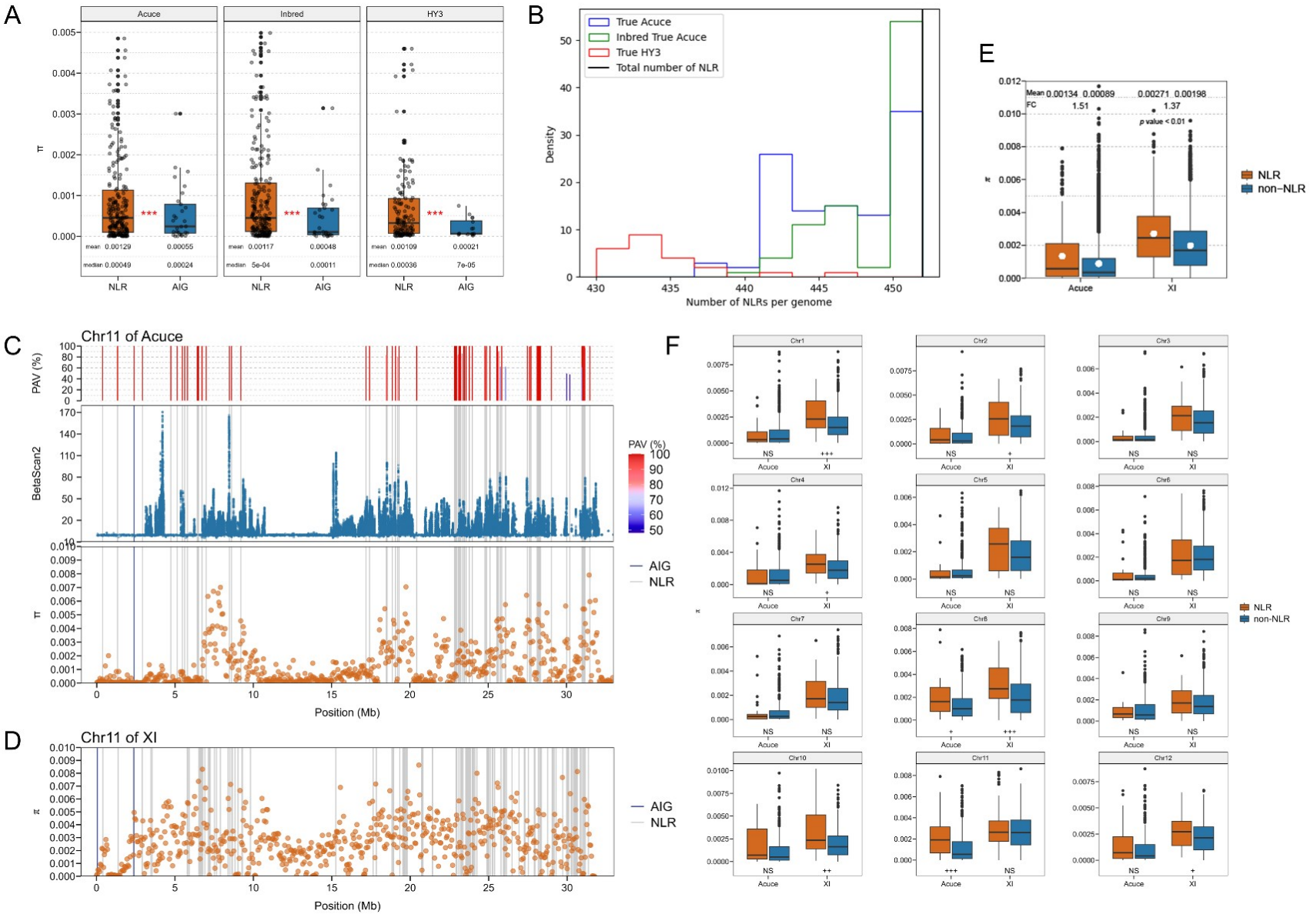


Figure 2. Structure of genetic diversity in Acuce landrace and its relatives. (A) The top panel represents the assignment probability to the clusters of the structure analysis of 494 indica samples (K=9). Cluster assignment was used to define true Acuce and delineate Acuce families 1 to 3. When no assignment was possible, the position in the tree was considered. (B) The distance-based tree contains 510 samples (including 10 confirmed and 5 potential non-indica samples, and the outgroup *Oryza barthii*). Branch lengths are square-root scaled for

better readability. Clades are colored according to the group of origin of samples, and outer markings indicate assigned cluster in the structure analysis. The worldwide XI (XI-1 to XI-3C) and GJ (JAP) sub-species as defined in Wang et al. (2018) are indicated as well as four sub-groups of YYT landraces identified in separate studies (Ali et al., 2023; Huang et al.). Red "+" signs represent probable seed contaminations of Acuce fields by genotypes detected by extreme genetic divergence from the bulk of Acuce accessions. The red dot indicates the Acuce line A15 used for Pacbio sequencing. A15


A
Acuce
Inbred
HY3
π
mean 0.00129 0.00055
median 0.00049 0.00024
mean 0.00117 0.00048
median 5e-04 0.00011
mean 0.00109 0.00021
median 0.00036 7e-05
NLR
AIG
B
True Acuce
Inbred True Acuce
True HY3
Total number of NLR
Density
Number of NLRs per genome
E
Mean 0.00134 0.00089 0.00271 0.00198
FC 1.51 1.37
p value < 0.01
π
Acuce
XI
NLR
non-NLR
C
Chr11 of Acuce
PAV (%)
BetaScan2
π
Position (Mb)
AIG
NLR
D
Chr11 of XI
π
Position (Mb)
AIG
NLR
F
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
NS
+
+++
++
Acuce
XI
NLR
non-NLR

830

Figure 3. Genomic diversity landscape of NLR and agronomically important genes in Acuce landrace. The average diversity levels (π) of NLRs and AIGs (A) and the distribution of NLR genes (B) were analyzed in 108 true Acuce, 87 true inbred Acuce and 23 true HY3 genomes. The red plus signs (+++) signify statistical significance at $p < 0.001$ based on the Wilcoxon Rank Sum and Signed Rank Tests. (C) The average diversity (π) and balancing selection (betascan2) levels (50 kb window) are shown along the chromosomes 11 of the Acuce line 15 genome, as well as the position of NLR and AIGs. Their average presence/absence variation (PAV) in the 108 true Acuce genomes is also displayed. (D) The average diversity (π) levels (50 kb window) are shown along the chromosomes 11 of the R498 genome. (E) For both Acuce and XI, average diversity levels (π) of NLR and non-NLR bins were calculated within 50 kb windows. The term "mean" represents the average value, "fold change" indicates the multiplicative change in the mean values of NLR/non-NLR, and the p-value is computed using the Wilcoxon test. (F) On each chromosome, the average diversity (π) levels of NLR and non-NLR bins were computed within 50 kb windows for both Acuce and XI populations. Statistical analysis was performed using the Wilcoxon test, where "NS" denotes not significant, "+" indicates a p-value < 0.05, "++" signifies a p-value < 0.01, and "+++" represents a p-value < 0.001.