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

Stéphane De Mita[2,†], Shengchang Duan[1,3,†], Xiang Li[1,3,†], 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

[†]: Stéphane De Mita, Shengchang Duan and Xiang Li contributed equally to this work.

[*]: 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

Abstract

Landraces are important resources for breeding programs. However, the diversity harbored at the genomic level by landraces is yet to be analyzed. Here we sequenced more than 200 individual plants from the long-lasting and durably resistant rice landrace Acuce (*Oryza sativa* subsp. indica). We found that Acuce exhibits diversity levels as high as 38% of the entire indica sub-species. We pinpoint signatures of selection in favor of polymorphism associated with resistance, but not other agronomic trait genes. Furthermore, we provide evidences that emerging properties upon mixing genomic diversity also increase Acuce's performances.

Keywords

landrace, diversity, immune receptor

Background

Landraces represent a traditional way of maintaining diversity in cultivated systems [1]. Indeed, while they do not result from deliberate improvement, landraces are both locally adapted and genetically diverse but are still identified and delineated by farmers [2, 3]. In rice, local or regional rice landraces have been analyzed to some level in Borneo or Madagascar islands [4, 5]. However, there is only limited knowledge on how diversity is structured within individual landraces [6], none at the genomic scale and even less on measuring our capacity to reproduce modern landraces.

The Yuanyang terraces (YYT; Yunnan, China) were shaped by the agricultural practices of the Hani ethnic group [7]; they were classified by UNESCO in 2013 as a Cultural Landscape and are known for being home to a wide range of rice landraces[8]. Quite importantly, the available records indicate that the YYT only suffer 1% losses caused by disease [9], a striking feature compared to the well-documented 32% losses at the global level [10]. High genetic diversity of some YYT landraces has been documented in previous reports using a limited number of molecular markers[11-13], suggesting they resemble composite populations. However, these reports did not allow the identification of specific genes or gene families that accounted for this diversity. Acuce (literally "God rice" in the Hani language and "Moon rice" in Mandarin) is a prevalent landrace in YYT and is related to XI-1 indica rice [14, 15]. It has been generated and grown in YYT for more than 130 years (Figure S1A) and appears to have remained immune to

the major fungal pathogen *Pyricularia oryzae* [15]. Another composite population produced by breeders, HongYang 3 (HY3; Figure S1C) was introduced in the mid 2010s in YYT.

The genetic determinants for durability of resistance to pathogens have never been addressed for any 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, what are the molecular bases of stable and robust pathogen resistance, and through which phenotypic and genotypic traits the high intra-landrace diversity is expressed.

Results and Discussion

A total of 227 Acuce lines from 10 villages in three zones of the YYT (Figure S1B) were sampled from individual panicles (Figure S1D; Table S1). Among these, a sub-set of 89 Acuce, collected in one field from the Gingko village, were amplified from single seeds (later called inbred) for the purpose of field phenotyping. In YYT fields, we observed negative correlations between, on the one hand, rice blast and smut disease scores and, on the other hand, thousand-seed weight and yield (Figure S2), suggesting that fungal pathogens are potential drivers of plant fitness. As a whole, major agronomical traits such as plant height, maturity, tiller number and thousand-seeds weight exhibited low variability in Acuce inbred lines (Figure 1A, B). In contrast to these agronomically traits, levels of disease resistance to three major fungal pathogens monitored was more variable. This variability was also observed under controlled conditions for blast (Table S1). Thus, although major traits easily tractable by farmers are quite homogeneous in Acuce, disease resistance traits appeared to be highly variable.

We also tested in field conditions the hypothesis that emerging properties in mixtures of Acuce inbred lines could favor 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 more than other (Figure S3B). As expected, mixtures tended to reduce disease levels: the number of plants showing smut symptoms was reduced by 23% in mixtures compared to pure stands (Figure S3D) and, the average severity was also reduced by 17% (Figure 1C); although this was not significant in general, it was significant in certain pairs of Acuce lines, where severity was reduced by up to 62% (Figure S3C). Thus, the mere mixing of Acuce lines generates emergent properties on traits that are key components for yield. This is

87 consistent with observations made under controlled conditions where disease resistance to
88 pathogens was modulated by plant-plant interactions in Acuce mixtures [16].

89 The genomes of 227 Acuce (including 89 Acuce inbred lines) and 30 HY3 plants were
90 compared to representative genomes of 192 YYT and 54 worldwide indica accessions (Table
91 S1; Figure S1E). A structure analysis based on two different methods (see Methods) showed
92 that Acuce samples are grouped together along with the indica XI group (K=2; Figure S4).
93 Optimization criteria of the two methods used suggested that K=9 is the most appropriate for
94 biological interpretation (Figure 2A) and allowed the identification of three families for Acuce
95 (Acuce 1 to 3; Table S1; Figure S4C). Phylogenetic analysis confirmed that these families are
96 clearly separated while HY3 belonged to a unique, different genomic sub-group (Figure 2B),
97 consistent with its known pedigree (Figure S1C). Using structure analysis and phylogenetic
98 trees, we defined 108 "true" Acuce, 87 "true" inbred Acuce and 23 "true" HY3 (Figure S4D)
99 for all further analyses (see Methods).

100 Nucleotide diversity of Acuce was 38% as compared to XI indica entire sub-species (Table S2),
101 consistent with preliminary analysis showing that diversity is very high within one given
102 landrace in YYT [11]. On average, nucleotide diversity values were 28% higher in Acuce than
103 in Acuce inbred and 71% higher than in modern population HY3. Selection signatures between
104 these sub-groups also confirmed probable different breeding or natural selection histories
105 (Figure S5 and Tables S3 and S4).

106 Quite remarkably, the range of differentiation values (estimated with the D of Jost) between
107 Acuce families overlapped those between world indica sub-groups (Figure S1B and Table S5).
108 Thus, despite strong genomic differences, farmers consider these plants as belonging to one
109 single landrace. There was no strong difference for nucleotide diversity in Acuce families when
110 considering villages or clusters of villages (Table S2). Considering villages with sufficient
111 sample sizes, the differentiation varied from 1.1% to 6.3%, much below differentiation between
112 Acuce families (Table S2 reported in Figure S1B), consistent with a strong social habit to
113 exchange seeds in Yuanyang [17].

114 At the genome wide level, the analysis of balancing selection yielded signatures on gene
115 ontologies in Acuce genomes only, all associated to Nucleotide-binding domain and Leucine-
116 rich Repeat genes (NLRs; Table S6) which are major determinants of pathogen resistance [18].
117 Because NLRs are known to be highly polymorphic, we produced a PacBio genome reference
118 for a specific NLR analysis in Acuce (Table S7 and Figure S7). We identified 452 NLRs in the
119 genome of Acuce (Figure 3A), a number rather small compared to other known rice genomes

120 (e.g. 712 NLRs were identified in the Shuhui498 genome). Thirteen NLRs were identified under
121 balancing selection in Acuce, 4 times more than expected by chance (Table S6). Quite
122 interestingly, the recently introduced HY3 modern population also showed some NLRs under
123 balancing selection. This suggests that this pattern of selection is connected to cultural practices
124 applied in the YYT. We next compared the diversity of NLRs with respect to other
125 agronomically important genes [19]. Agronomically important genes (AIGs; Table S8)
126 involved in traits such as leaf architecture, tillering or flowering were identified by homology
127 in the Acuce genome (Table S8). Overall, NLRs were much more diverse than AIGs in Acuce,
128 inbred Acuce or HY3 (Figure 3A; Table S9). In addition to single polymorphic sites, we looked
129 for presence/absence variations: 17 NLRs were more often absent in modern HY3 than in Acuce
130 landrace (Figure 3B; Table S10). In contrast, Acuce also showed both elevated numbers of NLR
131 and a wider range (from 438 to 452). Typical feature of diversity landscape for the Acuce
132 landrace and XI genomes is shown in Figure 3C and 3D (Figure S8). A significant association
133 of diversity with regions containing NLRs was found for both populations (Figure 3E), with a
134 stronger enrichment in Acuce and some for some but not all chromosomes (Figure 3F).

135 Considering the high diversity of NLR genes specifically found in Acuce, we performed a
136 GWAS analysis based on field phenotyping data. We identified two SNPs on chromosome 2
137 nearby a gene coding for a Leucine-rich repeat associated with brown spot resistance (Figure
138 S6). Whether this gene corresponds to a causal gene for resistance to this disease remains to be
139 demonstrated but this finding demonstrates the potential of landraces as a resource for the
140 identification of determinants of disease resistance.

141 Landraces are considered as untapped, critical reservoirs for future breeding efforts [1]. Our in-
142 depth genomic analysis of Acuce demonstrates that despite visible homogeneity for traits easily
143 tractable by farmers, a very high level of genomic diversity is maintained in this landrace. This
144 genomic diversity is particularly embedded in immune receptors, consistent with phenotypic
145 variation for disease resistance traits. The analysis of HY3 indicates that modern breeding can
146 achieve some levels of diversity, still lower than those observed in the traditional and long-lived
147 Acuce landrace. Why and how farmers maintain such diversity remains elusive. Our results
148 show that in the field, several traits related to yield, which in some respect may be considered
149 as fitness, are improved by simply mixing two genomes. Crop protection against pathogens
150 may also be another trait selected by farmers in Acuce. Indeed, diversity at genotypic level has
151 been shown to reduce disease risks and prevalence in agro-systems [20]. For instance, mixing
152 two varieties has proven to confer resistance to the rice blast fungus [21]. The combination of

two to more resistance genes in separate varieties is classically invoked to explain this property and the sustainability of such agro-systems [22, 23]. To some extent, disease control was also reduced in some pairwise mixtures tested here. Stronger effects may require more complex assemblages to be observed. In addition, the recent release of the HY3 modern composite population, which is depleted from a dozen immune receptors, is an opportunity to challenge the stability of disease control achieved by modern populations in the field.

Conclusions

Acuce landrace exhibits extensive genomic diversity, especially in immune receptors, offering potential advantages in disease resistance and crop protection.

Material and Methods

Sampling of Acuce, HY3 and other rice genomes

Acuce and HY3 plants were sampled in a small sector (~8 km × 13 km) of the Yuanyang county (Yunnan, China; Figure S1B). This area is similar to the one described in [8]. Because of the intense seed exchange and dynamics of farmers' choices [17], it was difficult to sample all villages for each landrace every year. HY3 was introduced in 2014 and results from breeding efforts (Figure S1C). In 2007, 89 Acuce seeds were collected in the Gingko village and single seed descents were amplified for 8 years. All samples were collected in 2015 and 2016 except seeds used a source material for inbred which we collected in 2007 (Figure S1D). The name of the varieties was first considered according to farmers' experience. Other representative 192 genomes of YYT (described in Huang et al, unpublished)[24] and 54 XI genomes representative of rice world diversity [19] were also included in order to position Acuce and HY3 genomes with respect to other major rice groups. In total, 503 genomes were used (Table S1).

Field phenotyping

The 89 inbred Acuce pure stand lines were sown and transplanted in the Gingko village in three separate paddy fields in 2022 (Figure S3A). In Figure 1A, the four parameters—plant height, tiller number, thousand-grain weight, and maturity—are represented using the original value. For the four indicators associated with diseases—blast, neck blast, brown spot, and smut—we employed a hierarchical survey method 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

184 Chiang KS, Liu HI and Bock CH [25]. In Figure 1C, six Acuce inbred lines were chosen to be
185 representative of sub-groups in Acuce 1 (Acuce#15, 36, 49, 69 and 74) or Acuce 3 (Acuce #87)
186 and assembled in 6 binary mixtures (among the 15 possible). They were evaluated for three
187 years. For each focal line and year, the number of plants scored for establishing the phenotypes
188 ranged from 195 to 441. we used linear modeling in R to scrutinize the impact of year, field,
189 pure/mix, and focal plant on the measured metrics using the model equation: lm(log1p(value)
190 ~ year + field + condition * focal, data=data). Subsequently, the R package lsmeans was utilized
191 to compute the least squares means for plotting. The p-values were extracted using lsmeans and
192 the pairs function in R. The R package FactoMineR (version 2.9) [26] was employed for PCA
193 analysis (Figure S2 A and B) with default parameters. For correlation calculations using the
194 Pearson correlation coefficient (Figure S2 C), we utilized the R package WGCNA (version
195 1.72.1) [27]. The *t*-test was applied to examine the significance of differences in data adhering
196 to a normal distribution, while the Wilcoxon test was employed for the remaining non-normally
197 distributed data.

198 **DNA sequencing and mapping on Shuhui**

199 Young leaves were collected from the plants and snap frozen in liquid nitrogen. Total DNA
200 was extracted with the DNAsecure plant kit (TIANGEN, Beijing). For each accession, 2 μg
201 genomic DNA was used to construct a sequencing library following the manufacturer's
202 instructions using NEBNext Ultra DNA Library Prep Kit (NEB Inc., America). Paired-end
203 sequencing libraries with an insert size of approximately 400 bp were sequenced on an Illumina
204 HiSeq 4000 sequencer. The average coverage per genome was 14.5× (Table S11). Paired-end
205 resequencing reads were cleaned using fastp (Version: 0.12.2) [28].

206 Cleaned paired-end reads were mapped to the *indica* rice genome Shuhui498 (R498) reference
207 genome) [29] with BWA (Version: 0.7.10-r789) [30]. SAMtools (Version: 1.3.1) [31] software
208 was used to convert mapping results into the BAM format. Duplicated reads were marked with
209 the Picard package (picard.sourceforge.net, Version: 2.1.1) after sorting the BAM using Picard
210 package. After BWA alignment, the reads around indels were realigned, realignment was
211 performed with Genome Analysis Toolkit (GATK, version 3.3-0-g37228af) [32] in two steps.
212 The first step used the RealignerTargetCreator package to identify regions where realignment
213 was needed, and the second step used IndelRealigner to realign the regions found in the first
214 step, which produced a realigned BAM file for each accession. The variation detection followed
215 the best practice workflow recommended by GATK [32]. The SNP variants were called for
216 each accession by the GATK HaplotypeCaller. A joint genotyping step for comprehensive

217 variations union was performed on the gVCF files. In the filtering step, the SNP filter
218 expression was set as "QD < 5.0 || MQ < 40.0 || FS > 100.0 || SOR > 3.0 || MQRankSum < -5.0
219 || ReadPosRankSum < -5.0 || QUAL < 30". SNPs with none bi-allelic were removed (Table
220 S12). Then the SNPs were cleaned for hidden paralogy as described in Huang et al[24].

221 **Structure analysis**

222 Structure analysis was performed using two complementary methods (Table S12): sNMF
223 version 2.0, as available within the R package LEA version 3.14.0 [33] and Admixture version
224 1.3 [34]. Both programs were used with K ranging from 1 to 15 and 10 replicates. sNMF was
225 run with the parameters: alpha=10, iterations=1000, tolerance=0.000001 and Admixture with
226 cv=10. The analysis was run on 494 samples, excluding 10 japonica rice samples along with
227 three Acuce (B73, 396, and A3) and two worldwide samples (304 and B170) that appear at a
228 basal position of the phylogenetic tree, indicative of a potential taxonomic error. After thinning,
229 27,522 diallelic sites were included in the analysis. Based on correlations of assignment
230 coefficients, we used *ad hoc* Python code to map clusters with each other over the replicates,
231 the two methods, and the successive values of K in order to have a consistent set of clusters.
232 Individuals were assigned to the cluster with the highest coefficient if it was at least 0.75, or if
233 the gap with the second best cluster was larger than 0.20.

234 **Phylogenetic analysis**

235 The pairwise distance matrix was computed between 510 samples (including *O. barthii*) over
236 all SNPs (Table S12), as the rate of pairwise differences at all sites where the given pair of
237 samples have non-missing genotypes. The neighbor-joining tree was computed using the
238 neighbor program of the PHYLIP package version 3.698 [35]. *O. barthii* was used as outgroup
239 to root the tree using EggLib version 3.1 [36] and the tree representation was realized using the
240 online tool iTol v6 [37].

241 **Analysis of nucleotide diversity**

242 Only diallelic sites from SNP Table S12 were considered in the analysis. Diversity statistics
243 were computed using EggLib version 3.1 [36]. For each group, we computed the average
244 number of non-missing samples (nseff), the number of polymorphic sites (S), the total
245 nucleotide diversity (Pi or π), average unbiased heterozygosity (He), and Tajima's D [38] which
246 is an indicator of the bias in allele frequency spectrum. For differentiation between populations,
247 we computed F_{ST} as in Weir BS and Cockerham CC [39], G'_{ST} [40], and Jost's D [41], the latter

248 being designed to be less sensitive than the former two to within-population levels of diversity,
249 and both un-standardized (Dxy) and net (Da) pairwise distance [42].

250 **Acuce genome construction**

251 The Acuce line #15 (see Figure 2B) was used as a reference for PacBio sequencing. A total of
252 50 mg DNA was used to construct the PacBio Sequel sequencing libraries. Then sequencing
253 was performed on PacBio Sequel platform to produce reads. The PacBio reads were assembled
254 using MECAT2 (Version: 2019.3.14) [45], then the yielded contigs were polished using Pilon
255 (Version: 1.21) [46] basing on the Illumina reads of A13 accession used for resequencing. For
256 obtaining a chromosome-scale assembly, Hi-C data were also generated. The Hi-C library was
257 prepared by NowBio Biotechnology Co., Ltd (Yunnan, China), with the standard procedure
258 described as follows. First, the high molecular weight genomic DNA was extracted from cross-
259 linked *in situ* and then digested with a restriction enzyme DpnII. The sticky ends of the digested
260 fragments were biotinylated, diluted, and then ligated to each other randomly. Biotinylated
261 DNA fragments were enriched and sheared to a fragment size of 300–500 bp again for preparing
262 the sequencing library, which was sequenced on an Illumina NovaSeq platform. The polished
263 sequences were further scaffolded using Hi-C data based on 3d-dna pipeline [47]. First, Hi-C
264 reads were mapped to the polished assembly with juicer (Version: 1.5.6) [48], and then a
265 candidate chromosome-length assembly was generated automatically using the 3d-dna pipeline
266 (Version: 180922) [47]. Manual review and refinement of the candidate assembly was
267 performed in Juicebox Assembly Tools (Version 1.9.1) [48] for quality control and interactive
268 correction. Then the genome was reviewed using 3d-dna [47] according to manually adjust.
269 With the modified 3d-dna and Juicebox workflow, 12 chromosomes were anchored. The
270 completeness of assembled genome was assessed using BUSCO (v5.4.4) [49] with
271 embryophyte_odb10 lineage dataset (2020-09-10) (Suppl Figure S7).

272 **RNA-seq**

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

279

Repeat and gene annotation

First, we searched tandem repeats across the genome using Tandem Repeat Finder (TRF, Version: 4.09) [50]. The transposable elements (TEs) in the genome were identified with a combination of homology-based and *de novo* approaches. For homolog based prediction, known repeats were identified using RepeatMasker (Version: open-4.0.9) [51] and RepeatProteinMask [51] against Repbase [52] (Repbase Release 20181026; http://www.girinst.org/repbase/index.html). For *de novo* prediction, LTR FINDER (Version: v1.07) [53] and ltrharvest in GenomeTools (Version: 1.5.10) [54, 55] were used to predict LTR retrotransposons, and the results were integrated using LTR_retriever (Version: v2.8) [56]. RepeatModeler (Version: 2.0; http://repeatmasker.org/) was also used to identify TEs in the genome. The results of LTR_retriever and RepeatModeler were merged as a custom library and fed to RepeatMasker to predict TEs.

We employed EVidence Modeler (EVM) [57] to consolidate *ab initio* gene predictions and homologous method annotation into a final gene set. For *ab initio* gene prediction, Augustus (v.2.5.5) [58] was run with the trained model of rice. Three additional *ab initio* gene prediction software were also used: Genescan (version 2015-10-31) [59], GlimmerHMM (v.3.0.2) [60], and SNAP (version 2006-07-28) [61]. Protein sequences of Nipponbare [62], 93-11 [63], Shuhui498 (R498) [29], *Oryza barthii* [64], *O. glaberrima* [65], and *O. nivara* [64] were obtained and used for the homology-based gene annotation. In brief, tblastn was first performed to map protein sequences of the aforementioned species to the Acuce genome with parameter ‘‘E-value = 1e-5’’. For a single protein, all matching DNA sequences in the reference genome were concatenated by Solar after low-quality records were filtered. A protein-coding region was obtained by extending 2,000 bp both upstream and downstream of the concatenated sequence. GeneWise (v.2.2.0) [66] was then used to predict gene structure within each protein-coding region. Based on the abovementioned all annotation results, a weighted and non-redundant gene set was generated by merging all of the gene models with EVM (v.r2012-06-25). Finally, the full-length transcripts were fed into PASA pipeline (v.2.4.1) [57] together with the EVM result for gene structure refinement and alternative spliced isoform annotation. The genes were filtered if the proportion of repeat sequences in gene sequences greater than 40% or CDS sequence length less than 150 bp.

The completeness of predicted genes was assessed using BUSCO (v5.4.4)[49] with embryophyte_odb10 lineage dataset (2020-09-10). 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) [67]. The pathway to which the gene might belong was derived from the matching genes in KEGG.

We obtained an initial 390.72 Mb Acuce genome assembly from 46.84 Gb PacBio single-molecule sequencing data (Table S13a), with contig N50=5.67 Mb. After polishing the assembly, we were able to anchor 385.15 Mb (98.43%) sequences into 12 pseudo-chromosomes using 60.84 Gb Hi-C data (Figure S7, Table S13b), achieving 98.2% completeness of the BUSCO core embryophyta genes (Table S13c). We predicted ~215.41 Mb (55.05%) repetitive elements in the Acuce genome (Table S13d), among which 212.86% repeats were transposable elements (TEs; Table S13e). We predicted 40,011 protein-coding genes (41,619 transcripts) following a combination of homology and *ab initio* methods (Table S13f), which captured 96.8% complete BUSCOs in embryophyte_obd10 (Table S13g). 67% (26,803) genes were assigned to at least a function in InterPro, GO or KEGG (Table S13h).

AIG and NLR comparisons

The pairwise gene synteny between R498 and Acuce was identified using MCscan (Python version v1.2.7) [68]. Then agronomically important genes were extracted from Wang et al [19] and 57 of them were found unambiguously in Acuce according to the gene synteny and blastp check (Table S8). NLR genes were identified using full-length Homology-based R-gene Prediction (HRP) pipeline [69]. Because NLR genes known to be highly diversified, the SNP calling was performed with Acuce genome as reference according to above method (Table S14). The nucleotide divergence of each SNP were computed using VCFtools software (v0.1.13) [70](Table S14). The nucleotide diversity of a gene was calculated as nucleotide divergency of all SNPs in the gene region divided by gene length. The gene presence/absence variation was characterized using the SGSGeneLoss package (Version: v0.1) [71] 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) [43] was applied to SNP matrix (Table S14 to detect selective sweeps based on the CLR test to detect signatures of artificial selection or natural selection in each rice group with grid number set as (chromosome length / 1000). The dashed lines mark the regions at the top 0.5%. The CLR cut-off values were as follow: Acuce1 (1795.98), Acuce2 (3263.83),

Inbred_Acuce1 (1725.57), Inbred True Acuce (2708.91), True Acuce: (1639.96) and True HY3 (2540.68).

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 (v0.1.13) [70].

The impact of NLR presence on π was assessed using a Wilcoxon test. A 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.

Balancing selection

Balancing selection was determined using Beta2 statistic in BetaScan [44] applied to SNP matrix (Table S14 with a sliding window size of 1000bp, the substitutions and ancestral alleles were called using *O. rufipogon* as root. The mutation rate was calculated as 4*Ne*μ 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 analysis was performed on rice brown spot from field data. BLUP (best linear unbiased predictors) were calculated using a generalized linear mixed model (glmer function from the R package lme4)[72], with a logit link function and a binomial distribution of errors:

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

where Yijk: the severity of the rice brown spot from plot i, line j and repetition k, Pi: the random effect of the plot i, Lj: the random effect of the line j and PLij: the interaction between the plot i and the line j.

Then the GWAS was performed with the BLUP values based on SNP set called with Acuce genome using Multiple Loci Mixed linear Model (MLMM) in GAPIT [73]. The first five PCA values (eigenvectors) were used as fixed effects in the model to correct for stratification. We defined the whole-genome significance cutoff with -log10 (P) = 5.

Supplementary Information

Additional file 1: 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 in true Acuce, Indred and HY3 sub-groups. **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.** Agronomically important genes (AIG) in Acuce genome. **Table S9.** Nucleotide diversity at AGI and NLR genes. **Table S10.** Presence/Absence polymorphism for AIG and NLR genes in Acuce and HY3. **Table S11.** Genome statistics. **Table S12.** SNP Table using R498 reference genome. **Table S13.** Statistics of the Acuce genome assembly. **Table S14.** SNP Table using Acuce reference genome. **Table S15.** Table for average diversity.

Additional file 2: Figure S1. Origins of Acuce and HY3 genomes. **Figure S2.** Field phenotypes of Acuce inbred lines. **Figure S3.** Phenotypes in binary Acuce mixtures. **Figure S4.** Structure analysis of Acuce and HY3 genomes. **Figure S5.** Separate breeding histories between Acuce and HY3. **Figure S6.** GWAS analysis of brown spot resistance in the Acuce landrace. **Figure S7.** Heatmap of Hi-C interactions for Acuce pseudo-chromosomes. **Figure S7.** Pi and betascore analysis of AIG and NLR in Acuce and XI. **Figure S8.** Diversity feature of Acuce and XI genomes.

Declarations

Ethics approval and consent to participate

Ethical approval is not applicable for this manuscript.

Availability of data and materials

The raw reads, genome assembly and annotation are available at the China National Center for Bioinformation under the accession number PRJCA022862.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

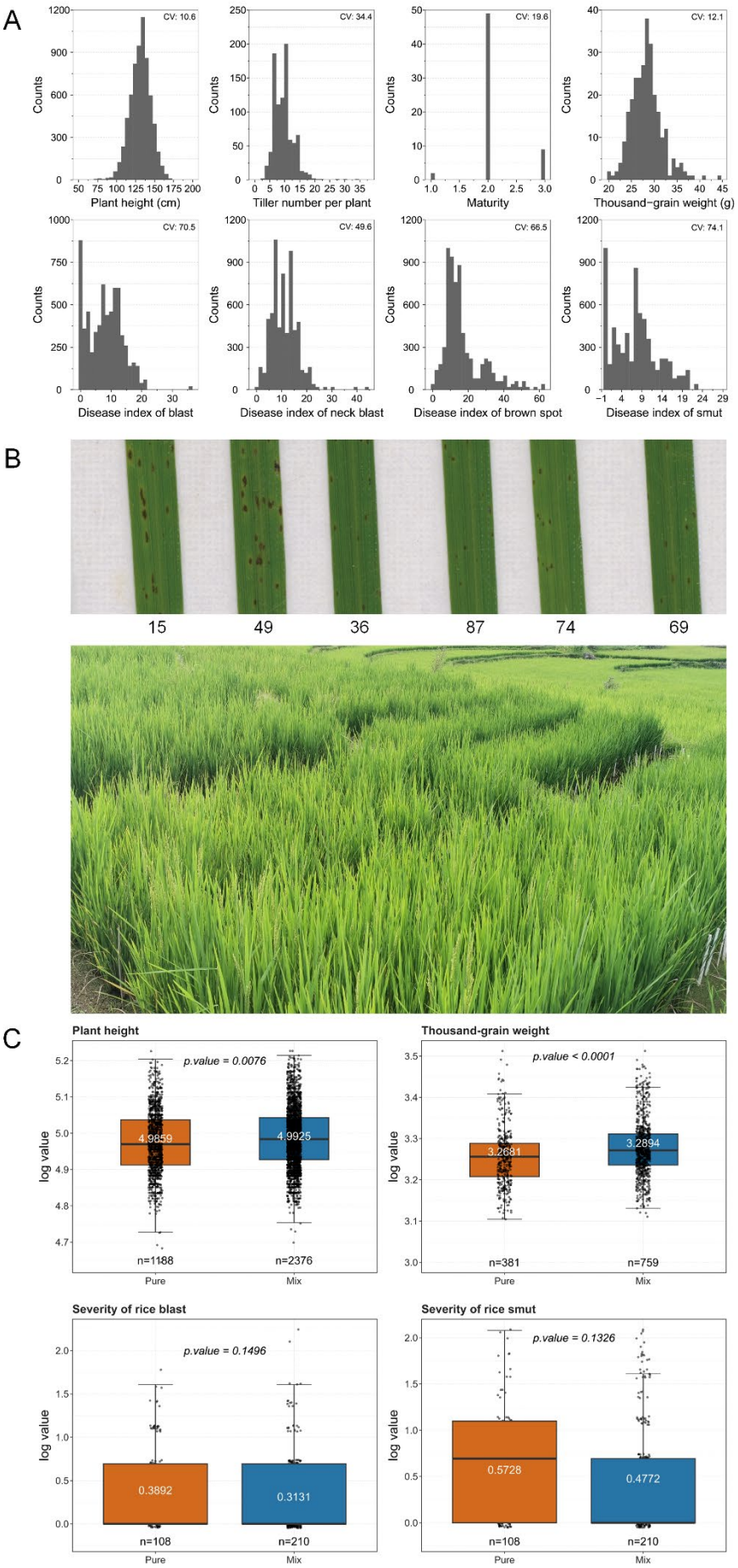
We thank Elisabeth Fournier and Henry Adreit for their support in collecting plants in the field.

Authors' 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.

Funding

This study was supported by National Key R&D Program of China (2023YFE0107500, 2023YED1400800), National Natural Science Foundation of China (31801792, 31960554) and Yunnan Fundamental Research Projects (202301AS070257). 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 LIA PLANTOMIX projects.


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
4.9859
4.9925
n=1188
n=2376
Pure
Mix
log value
Thousand-grain 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
n=108
n=210
Severity of rice smut
p.value = 0.1326
0.5728
0.4772

422

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-grain weight and Wilcox test for incidence of disease.

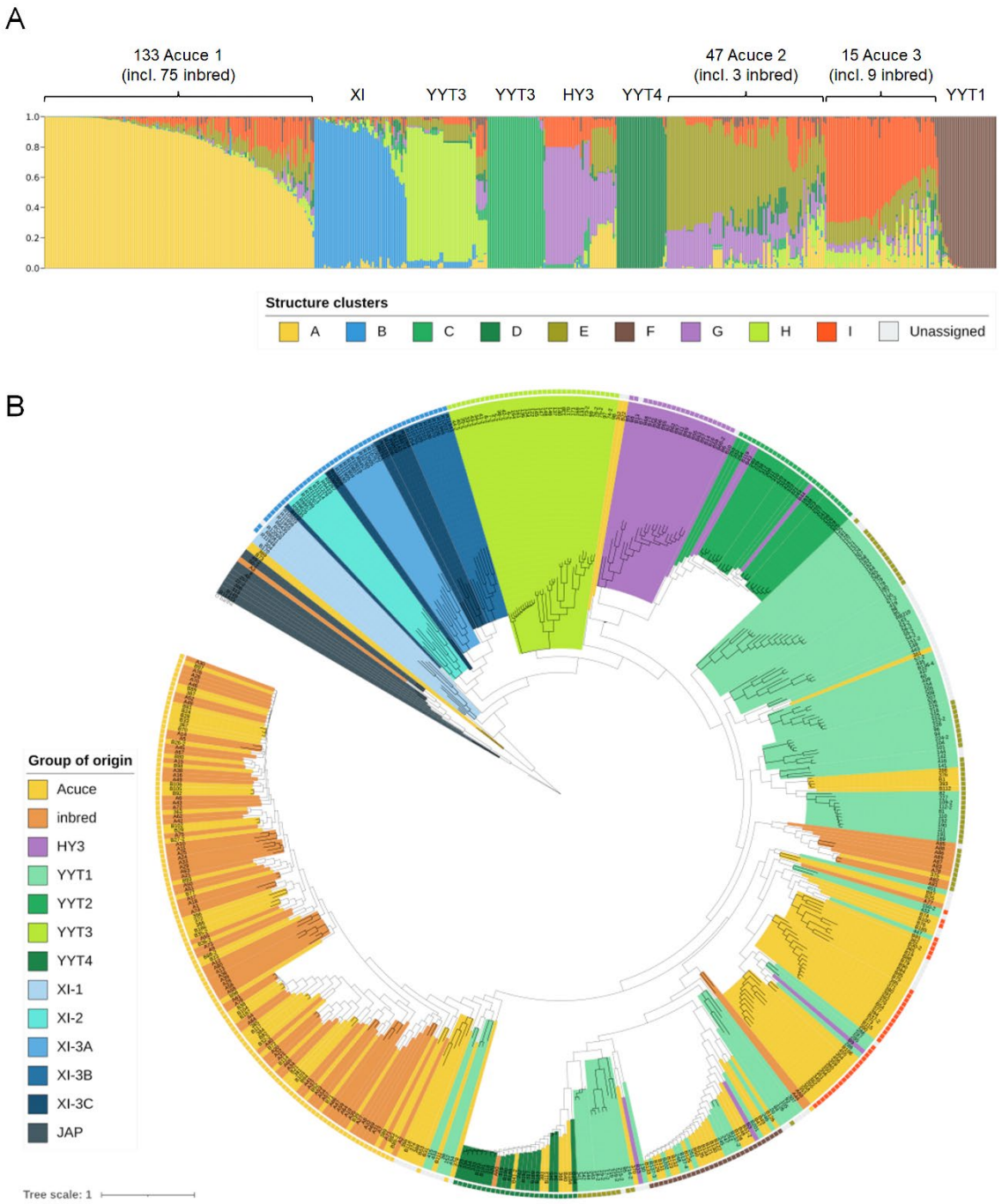
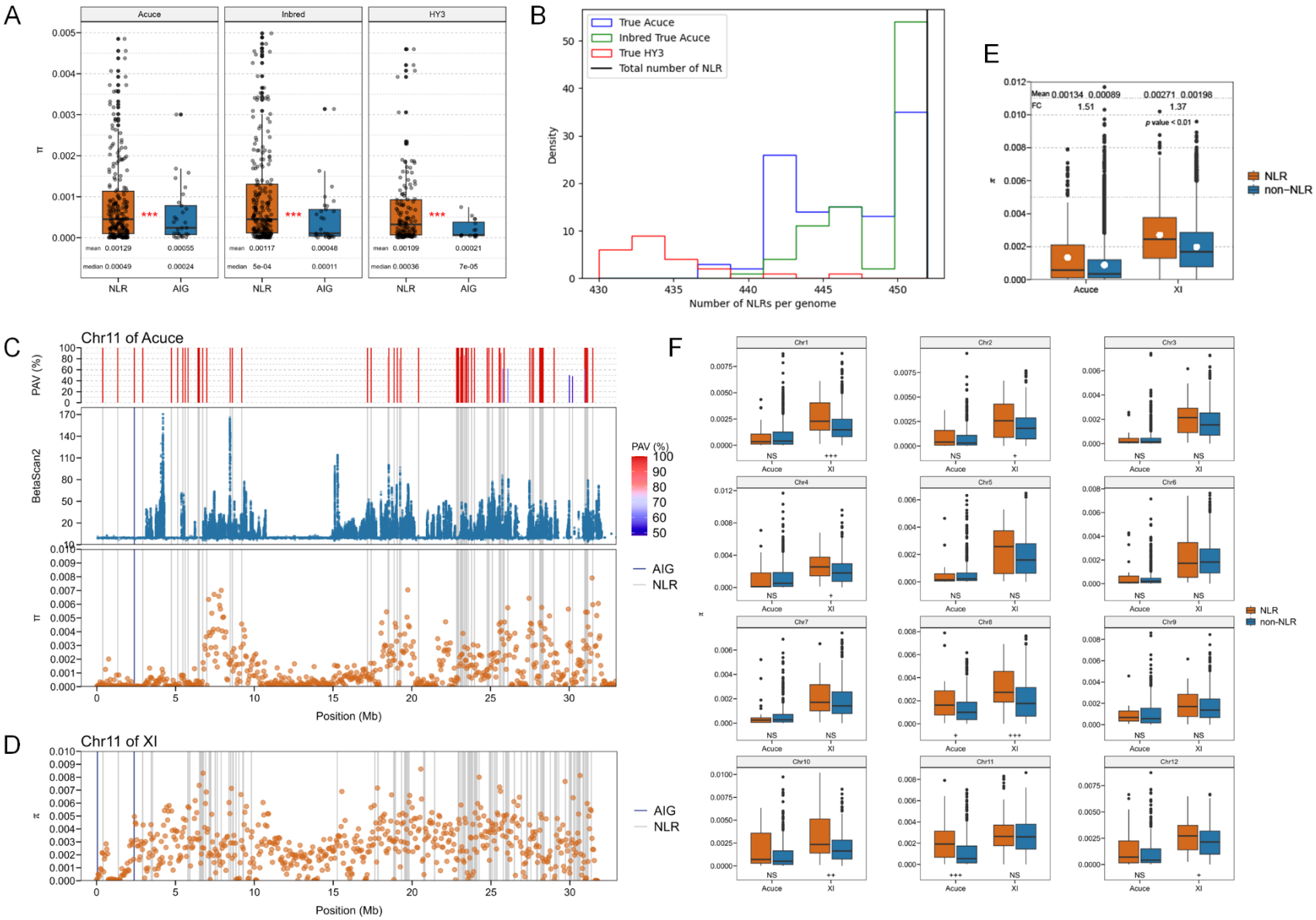


Figure 2. Structuration of genetic diversity in Acuce landrace and its relatives. (A) The top panel represents the assignation probability to the clusters of the structure analysis of 494 indica samples (K=9). Cluster assignation was used to define true Acuce and delineate Acuce families 1 to 3. When no assignation 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 world indica (XI-1 to XI-3C) and japonica (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 (Huang et al, unpublished; Ali et al, 2023). Red "+" signs represent probable seed contaminations of Acuce fields by other genotypes. The red dot indicates the Acuce line #15 used for Pacbio sequencing.


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
Acuce
XI
NLR
non-NLR
C
Chr11 of Acuce
PAV (%)
BetaScan2
AIG
NLR
Position (Mb)
D
Chr11 of XI
F
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12

447

Figure 3. Genomic diversity landscape of NLR and agronomically important genes in Acuce landrace. The average diversity levels (π) of NLRs and AIGs (A)(Table S9) and (B) the distribution of NLR genes was 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.

1. McCouch S, Navabi ZK, Abberton M, Anglin NL, Barbieri RL, Baum M, Bett K, Booker H, Brown GL, Bryan GJ. Mobilizing crop biodiversity. Molecular plant. 2020; 13:1341-1344.
2. Harlan JR. Our Vanishing Genetic Resources. Science. 1975; 188:618-621.
3. Villa TCC, Maxted N, Scholten M, Ford-Lloyd B. Defining and identifying crop landraces. Plant genetic resources. 2005; 3:373-384.
4. Thomson MJ, Polato NR, Prasetiyono J, Trijatmiko KR, Silitonga TS, McCouch SR. Genetic Diversity of Isolated Populations of Indonesian Landraces of Rice (*Oryza sativa* L.) Collected in East Kalimantan on the Island of Borneo. Rice. 2009; 2:80-92.
5. Mather KA, Molina J, Flowers JM, Rubinstein S, Rauh BL, Lawton-Rauh AMY, Caicedo AL, McNally KL, Purugganan MD. Migration, isolation and hybridization in island crop populations: the case of Madagascar rice. Molecular Ecology. 2010; 19:4892-4905.
6. Pusadee T, Jamjod S, Chiang Y-C, Rerkasem B, Schaal BA. Genetic structure and isolation by distance in a landrace of Thai rice. Proceedings of the National Academy of Sciences. 2009; 106:13880-13885.
7. Wang Q. On the culture of the terraced fields. Kunming: Yunnan University Press; 1999.
8. Jiao Y, Li X, Liang L, Takeuchi K, Okuro T, Zhang D, Sun L. Indigenous ecological knowledge and natural resource management in the cultural landscape of China's Hani Terraces. Ecological Research. 2012; 27:247-263.
9. Li Y. Yuanyang county chronicles. Guiyang: Gui Zhou National Press; 1990.
10. Savary S, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. The global burden of pathogens and pests on major food crops. Nature Ecology & Evolution. 2019; 3:430-439.
11. Gao D, Mao R, Zhu Y. Comparative studies on intra-varietal heterogeneity between rice landraces and improved varieties. Rice Genomics and Genetics. 2012; 3.
12. Gao D, Wang Y, He H, Li C, Zhu Y. Intra-varietal heterogeneity and implications of Baijiaolaogeng rice landraces in Yuanyang county. Molecular Plant Breeding. 2009; 7:283-291.
13. Zhang H, Sun J, Wang M, Liao D, Zeng Y, Shen S, Yu P, Mu P, Wang X, Li Z. Genetic structure and phylogeography of rice landraces in Yunnan, China, revealed by SSR. Genome. 2007; 50:72-83.
14. Alonso P, Gladieux P, Moubset O, Shih P-J, Mournet P, Frouin J, Blondin L, Ferdinand R, Fernandez E, Julian C, et al. Emergence of Southern Rice Black-Streaked Dwarf Virus in the Centuries-Old Chinese Yuanyang Agrosystem of Rice Landraces. Viruses. 2019; 11:985.
15. Ali S, Gladieux P, Ravel S, Adreit H, Meusnier I, Milazzo J, Cros-Arteil S, Bonnot F, Jin B, Dumartinet T, et al. Evolution of the rice blast pathogen on spatially structured rice landraces maintains multiple generalist fungal lineages. Molecular Ecology. 2023; 32:2519-2533.
16. Pélissier R, Ballini E, Temple C, Ducasse A, Colombo M, Frouin J, Qin X, Huang H, Jacques D, Florian F, et al. The genetic identity of neighboring plants in intraspecific mixtures modulates disease susceptibility of both wheat and rice. PLOS Biology. 2023; 21:e3002287.
17. Dedeurwaerdere T, Hannachi M. Socio-economic drivers of coexistence of landraces and modern crop varieties in agro-biodiversity rich Yunnan rice fields. Ecological Economics. 2019; 159:177-188.
18. Ngou BPM, Ding P, Jones JDG. Thirty years of resistance: Zig-zag through the plant immune system. The Plant Cell. 2022; 34:1447-1478.
19. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature. 2018; 557:43-49.
20. Mundt CC. USE OF MULTILINE CULTIVARS AND CULTIVAR MIXTURES FOR DISEASE MANAGEMENT. Annual Review of Phytopathology. 2002; 40:381-410.
21. Zhu Y, Chen H, Fan J, Wang Y, Li Y, Chen J, Fan J, Yang S, Hu L, Leung H, et al. Genetic diversity and disease control in rice. Nature. 2000; 406:718-722.
22. Borg J, Kiær LP, Lecarpentier C, Goldringer I, Gauffreteau A, Saint-Jean S, Barot S, Enjalbert J. Unfolding the potential of wheat cultivar mixtures: A meta-analysis perspective and identification of knowledge gaps. Field Crops Research. 2018; 221:298-313.
23. Garrett KA, Zúñiga LN, Roncal E, Forbes GA, Mundt CC, Su Z, Nelson RJ. Intraspecific functional diversity in hosts and its effect on disease risk across a climatic gradient. Ecological Applications. 2009; 19:1868-1883.

24. Huang H, Duan S, Dong Y, Morel J-B, Pierre G. Genome sequencing of rice landraces from the Yuanyang terracs uncovers ancient and diverse lineages of indica rice. unpublished.
25. Chiang KS, Liu HI, Bock CH. A discussion on disease severity index values. Part I: warning on inherent errors and suggestions to maximise accuracy. Annals of Applied Biology. 2017; 171:139-154.
26. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software. 2008; 25:1 - 18.
27. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559.
28. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34:i884-i890.
29. Du H, Yu Y, Ma Y, Gao Q, Cao Y, Chen Z, Ma B, Qi M, Li Y, Zhao X, et al. Sequencing and de novo assembly of a near complete indica rice genome. Nature Communications. 2017; 8:15324.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009; 25:1754-1760.
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078-2079.
32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010; 20:1297-1303.
33. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and Efficient Estimation of Individual Ancestry Coefficients. Genetics. 2014; 196:973-983.
34. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome research. 2009; 19:1655-1664.
35. Felsenstein J. PHYLIP (phylogeny inference package), version 3.5 c. Joseph Felsenstein.; 1993.
36. Siol M, Coudoux T, Ravel S, De Mita S. EggLib 3: A python package for population genetics and genomics. Molecular Ecology Resources. 2022; 22:3176-3187.
37. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research. 2021; 49:W293-W296.
38. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123:585-595.
39. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984; 38:1358-1370.
40. Hedrick PW. A STANDARDIZED GENETIC DIFFERENTIATION MEASURE. Evolution. 2005; 59:1633-1638.
41. Jost LOU. GST and its relatives do not measure differentiation. Molecular Ecology. 2008; 17:4015-4026.
42. Nei M. Molecular evolutionary genetics. Columbia university press; 1987.
43. Pavlidis P, Živković D, Stamatakis A, Alachiotis N. SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. Molecular Biology and Evolution. 2013; 30:2224-2234.
44. Siewert KM, Voight BF. BetaScan2: Standardized Statistics to Detect Balancing Selection Utilizing Substitution Data. Genome Biology and Evolution. 2020; 12:3873-3877.
45. Xiao C-L, Chen Y, Xie S-Q, Chen K-N, Wang Y, Han Y, Luo F, Xie Z. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nature Methods. 2017; 14:1072-1074.
46. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE. 2014; 9:e112963.
47. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 2017; 356:92-95.

48. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Systems. 2016; 3:95-98.
49. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015; 31:3210-3212.
50. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research. 1999; 27:573-580.
51. Chen N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Current Protocols in Bioinformatics. 2004; 5:4.10.11-14.10.14.
52. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research. 2005; 110:462-467.
53. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research. 2007; 35:W265-W268.
54. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 2008; 9:18.
55. Gremme G, Steinbiss S, Kurtz S. GenomeTools: A Comprehensive Software Library for Efficient Processing of Structured Genome Annotations. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2013; 10:645-656.
56. Ou S, Jiang N. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons Plant Physiology. 2018; 176:1410-1422.
57. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biology. 2008; 9:R7.
58. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008; 24:637-644.
59. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA11Edited by F. E. Cohen. Journal of Molecular Biology. 1997; 268:78-94.
60. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004; 20:2878-2879.
61. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004; 5:59.
62. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. japonica). Science. 2002; 296:92-100.
63. Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. A Draft Sequence of the Rice Genome (Oryza sativa L. ssp. indica). Science. 2002; 296:79-92.
64. Jacquemin J, Bhatia D, Singh K, Wing RA. 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. 2013; 16:147-156.
65. Wang M, Yu Y, Haberer G, Marri PR, Fan C, Goicoechea JL, Zuccolo A, Song X, Kudrna D, Ammiraju JSS, et al. The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication. Nature Genetics. 2014; 46:982-988.
66. Birney E, Durbin R. Using GeneWise in the Drosophila annotation experiment. Genome Research. 2000; 10:547-548.
67. Zdobnov EM, Apweiler R. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001; 17:847-848.
68. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and Collinearity in Plant Genomes. Science. 2008; 320:486-488.
69. Andolfo G, Dohm JC, Himmelbauer H. Prediction of NB-LRR resistance genes based on full-length sequence homology. The Plant Journal. 2022; 110:1592-1602.
70. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. GigaScience. 2021; 10:giab008.

71. Golicz AA, Martinez PA, Zander M, Patel DA, Van De Wouw AP, Visendi P, Fitzgerald TL, Edwards D, Batley J. Gene loss in the fungal canola pathogen Leptosphaeria maculans. Functional & Integrative Genomics. 2015; 15:189-196.
72. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823. 2014.
73. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012; 28:2397-2399.
74. Chen T, Chen X, Zhang S, Zhu J, Tang B, Wang A, Dong L, Zhang Z, Yu C, Sun Y, et al. The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. Genomics, Proteomics & Bioinformatics. 2021; 19:578-583.
75. Members C-N, Partners. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2022. Nucleic Acids Research. 2022; 50:D27-D38.