Extensive immune receptor repertoire diversity in disease-resistant rice landraces

Pierre Gladieux, Cock van Oosterhout, Sebastian Fairhead, Agathe Jouet, Diana Ortiz, Sébastien Ravel, Ram-Krishna Shrestha, Julien Frouin, Xiahong He, Youyong Zhu, et al.

Article

Current Biology

Extensive immune receptor repertoire diversity in disease-resistant rice landraces

Highlights

- Rice landraces from the Yuanyang terraces rarely show severe losses to disease
- Landraces display high and ancient sequence diversity at NLR immune receptors
- Variation in NLR number is higher in landraces than modern varieties
- Diversity-maintaining (i.e., balancing) selection is widespread in landrace NLRs

Authors

Pierre Gladieux, Cock van Oosterhout, Sebastian Fairhead, ..., Huichuan Huang, Thomas Kroj, Jonathan D.G. Jones

Correspondence

pierre.gladieux@inrae.fr (P.G.), absklhhc@gmail.com (H.H.), thomas.kroj@inrae.fr (T.K.), jonathan.jones@tsl.ac.uk (J.D.G.J.)

In brief

Gladieux et al. show that rice landraces from Yunnan have high and ancient sequence diversity in NLR immune receptors. The maintenance of elevated standing genetic variation at NLRs may explain why traditional agroecosystems are more resistant to diseases, providing valuable insights for engineering crops that are less prone to epidemics.

Gladieux et al., 2024, Current Biology *34*, 3983–3995
September 9, 2024
https://doi.org/10.1016/j.cub.2024.07.061

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

Extensive immune receptor repertoire diversity in disease-resistant rice landraces

Pierre Gladieux,[1,10,*] Cock van Oosterhout,[2] Sebastian Fairhead,[3] Agathe Jouet,[3] Diana Ortiz,[1] Sebastien Ravel,[1] Ram-Krishna Shrestha,[3] Julien Frouin,[4,5] Xiahong He,[6] Youyong Zhu,[7,8] Jean-Benoit Morel,[1] Huichuan Huang,[7,8,*] Thomas Kroj,[1,9,*] and Jonathan D.G. Jones[3,9,*]

[1]Plant Health Institute Montpellier, University of Montpellier, INRAE, CIRAD, IRD, Institut Agro, 34398 Montpellier, France
[2]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK
[3]The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norwich NR4 7UH, UK
[4]CIRAD, UMR AGAP Institut, 34398 Montpellier, France
[5]UMR AGAP Institut, Université de Montpellier, CIRAD, INRAE, Institut Agro, 34398 Montpellier, France
[6]School of Landscape and Horticulture, Southwest Forestry University, Kunming 650233, China
[7]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming 650201, China
[8]Key Laboratory of Agro-Biodiversity and Pest Management of Education Ministry of China, Yunnan Agricultural University, Kunming 650201, China
[9]These authors contributed equally
[10]Lead contact
*Correspondence: pierre.gladieux@inrae.fr (P.G.), absklhhc@gmail.com (H.H.), thomas.kroj@inrae.fr (T.K.), jonathan.jones@tsl.ac.uk (J.D.G.J.)
https://doi.org/10.1016/j.cub.2024.07.061

SUMMARY

Plants have powerful defense mechanisms and extensive immune receptor repertoires, yet crop monocultures are prone to epidemic diseases. Rice (*Oryza sativa*) is susceptible to many diseases, such as rice blast caused by *Magnaporthe oryzae*. Varietal resistance of rice to blast relies on intracellular nucleotide binding, leucine-rich repeat (NLR) receptors that recognize specific pathogen molecules and trigger immune responses. In the Yuanyang terraces in southwest China, rice landraces rarely show severe losses to disease whereas commercial inbred lines show pronounced field susceptibility. Here, we investigate within-landrace NLR sequence diversity of nine rice landraces and eleven modern varieties using complexity reduction techniques. We find that NLRs display high sequence diversity in landraces, consistent with balancing selection, and that balancing selection at NLRs is more pervasive in landraces than modern varieties. Notably, modern varieties lack many ancient NLR haplotypes that are retained in some landraces. Our study emphasizes the value of standing genetic variation that is maintained in farmer landraces as a resource to make modern crops and agroecosystems less prone to disease. The conservation of landraces is, therefore, crucial for ensuring food security in the face of dynamic biotic and abiotic threats.

INTRODUCTION

Plant immunity requires timely activation of defense mechanisms, based upon detection of pathogen molecules via either cell-surface or intracellular immune receptors. Evasion of detection enables pathogens to proliferate and cause disease. When pathogens encounter large populations of genetically identical and susceptible crop plants, rapid pathogen propagation and crop destruction can occur. *Resistance* (*R*) genes protect crops from disease and frequently encode intracellular nucleotide binding, leucine-rich repeat (NLR) immune receptors that detect specific pathogen effectors (virulence factors) and confer the innate ability to recognize pathogens. Most plants carry hundreds of NLR-encoding genes[1] and display extensive variation at *R* gene loci. In host-parasite coevolution, extensive standing variation at these *R* genes is critical to cope with evolutionary diversity in pathogens, which enables sustainable resistance in natural host populations.[2]

Genetic diversity for pathogen recognition in host populations can block or slow down epidemics. Variation between host genotypes in their resistance to different strains of the same pathogen reduces the risk that the host population is colonized by a single pathogen strain.[3–5] Importantly, population-level resistance can be thought of as an emergent property resulting from diversity in immune receptor repertoires. A single "perfect" genotype cannot capture the variation present in an outbreeding population; it is both the presence as well as the absence of genetic variants that provides the crucial biodiversity. In contrast, standard plant breeding practice in modern agriculture requires varieties to display uniformity and reliable performance over a wide range of environments. Such properties of modern agroecosystems are incompatible with population-level heterogeneity in immune receptor repertoires.

Traditional farming systems tend to rely on genetically heterogeneous mixtures of traditional varieties, referred to as landraces,[6] and they often provide effective and sustainable

disease control.[7] For example, traditional farmers in the Yuanyang terraces in Yunnan (south-west China) cultivate rice (*Oryza sativa*) landraces that rarely show severe losses to infectious diseases.[8,9] About 200 landraces[10] are maintained by a traditional social organization involving sporadic seeds exchange between farmers.[11] Furthermore, farmers instinctively carry out varietal selection by not planting varieties that were heavily impacted by disease in the previous season.[11] This social organization may have contributed to *R* gene heterogeneity in two ways: (1) by enhancing spatiotemporal variation in *R* gene repertoires and intensifying selection through selective planting and (2) by increasing gene flow (through the exchange of seeds). Both processes may have contributed to resistance durability in rice populations grown in the Yuanyang terraces.

Modern farming practices have profound coevolutionary implications, and these can be best understood in the light of population genetic theory.[12–17] During coevolution, adaptations in one species provoke counter-adaptations in the coevolving species. Consequently, the direction and intensity of natural selection constantly change.[18] Assuming that both antagonists possess sufficient genetic variation to fuel these continuous adaptations, none of the interacting species gains a sustained fitness advantage. Balancing selection maintains genetic polymorphisms in host resistance genes due to spatiotemporal variation in selection pressures posed by the pathogens. In other words, different genetic variants (e.g., alleles or haplotypes) are favored in different places and at different times, meaning that genetic polymorphism can be maintained long term. This is known as the trench-warfare model[19] or Red Queen dynamics.[18] Importantly, this also limits the infection incidence (i.e., the number of infected hosts) because the susceptible host genotype is locally and/or temporally continuously replaced by a genotype that is resistant to the prevailing pathogen strain. The composition of the prevailing pathogen strains is itself variable. The variation in the host population and the pathogen population makes antagonistic coevolution a zero-sum game with no knockout winners or losers.

In contrast, if there is insufficient host genetic variation, a pathogen strain that can overcome the defenses of the predominant host genotype is likely to cause damaging reductions in population viability if the host lacks any resistant plants. Even if the host population survives, the susceptible genotype may be lost completely (because of the unrestrained, exponential increase of the winning pathogen strain). In turn, this tends to result in a turnover of sequence variation. This type of host-parasite coevolution is sometimes observed in plant breeding, when new monoculture varieties are released, and it is known as the Red Queen arms race.[18,19] Modern crops consisting of genetically near-uniform host plants are ill-equipped to face the coevolutionary challenges posed by diverse, rapidly evolving pathogens with a trench-warfare model. Rather, to keep pace with their rapidly evolving pathogens, they are forced into an arms race that requires a continuous input of novel resistant varieties carrying individual *R* genes,[20] as well as agrichemical disease-control measures. In other words, the standing variation implied by the trench-warfare model is a "recycle-and-reuse" strategy that is sustainable, whereas the arms-race model uses sequence variation in a disposable fashion, making it less sustainable. In this study, we examine this coevolutionary hypothesis (see Ebert and Fields[2] for an excellent review).

We report here on a study to examine whether *indica* landraces of the Yuanyang terraces might show relatively elevated levels of diversity in their NLR immune gene repertoires, hypothesizing that traditional farming practices are better than modern breeding at conserving variation. We used RenSeq sequence capture[21] to enrich NLR sequences prior to sequencing, and we designed a set of biotinylated bait sequences to capture NLR-encoding immune receptors, based on reference genomes of *O. sativa ssp. japonica* and *O. sativa ssp. indica* (herein referred to as japonica and indica, respectively). All rice genotypes were assessed using Illumina sequencing of captured NLRs. We evaluated RenSeq data from 11 japonica, aus, and indica inbreds, and we compared these data to those from 38 accessions from seven different landraces from the Yuanyang terraces. We evaluated presence/absence variation and sequence diversity. These analyses revealed a marked depletion of NLR polymorphisms in the japonica and indica inbred lines and substantial within-landrace NLR sequence heterogeneity that likely underpins the relatively low incidence of rice blast in Yuanyang terraces.[22] We discuss the demographic events and evolutionary forces that can explain these data.

RESULTS

Data generation

We selected 49 Asian rice (*Oryza sativa*) accessions for short-read RenSeq analysis, representing 7 indica landraces (36 accessions), 2 japonica landraces (2 accessions), and 11 modern varieties of indica, japonica, and aus (11 accessions). We also generated long-read RenSeq data for 15 accessions of indica landraces, modern indica, and modern japonica. Landrace accessions were sampled in 2014 and 2015 in the fields of traditional rice farmers in three villages from the Yuanyang rice terrace region in Yunnan. RenSeq baits were designed to hybridize with 761 NLR-coding sequences from japonica and indica rice. To generate a baseline against which to identify features of polymorphism in genes of interest, 68 accessions were also characterized using genotyping-by-sequencing (GBS), representing nine landraces and 28 modern varieties (Table S1). After standardizing the number of reads, read mapping, SNP calling, and masking of paralogous calls or other SNPs with excess heterozygosity, the RenSeq dataset included 40,530 biallelic SNPs (reference sequence: 596 genes, 2.4 Mb) and the GBS dataset 199,130 biallelic SNPs (reference sequence: 42,031 genes, 99.6 Mb).

Population subdivision

To understand the genetic relationships among rice subspecies and landraces, and to investigate signatures of natural selection at the intraspecific level, we inferred population structure from short-read RenSeq and GBS data using complementary approaches that make no assumption about Hardy-Weinberg equilibrium and are therefore appropriate to analyze structured or inbred populations. Both clustering analyses with the sNMF software[23] (Figure S1) and neighbor-net phylogenetic networks[24] (Figure 1) revealed consistent patterns that split genetic variation, primarily by type of rice: aus, modern temperate japonica,

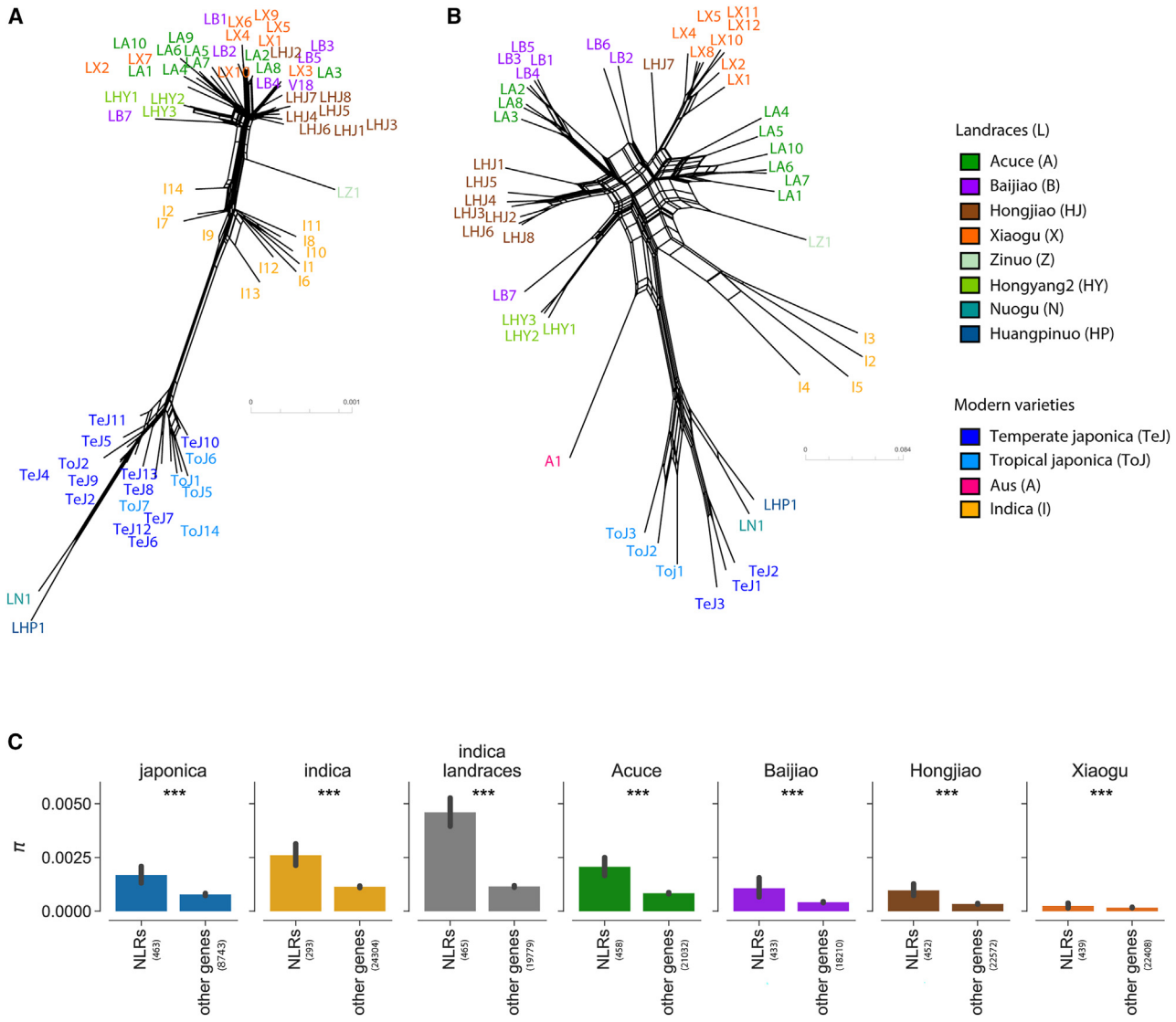

Figure 1. Analysis of nucleotide diversity separates modern varieties and landraces, which reveals high standing genetic diversity in landraces compared to modern varieties

(A and B) Population subdivision was inferred from 49 and 68 accessions, for GBS (A) and RenSeq (B), respectively, representing 13 varieties or landraces shown with different colors. Neighbor-net phylogenetic networks estimated with SPLITSTREE[24] for GBS and RenSeq data. Reticulations in the network indicate phylogenetic conflicts caused by homoplasy. SPLITSTREE analysis was based on 31,770 biallelic SNPs with less than 80% missing data for RenSeq data, and 60,166 biallelic SNPs with less than 50% missing data for GBS data.

(C) Represents bar plots of nucleotide diversity π in RenSeq data (NLRs) and GBS data (other genes). The "indica landraces" group includes one randomly chosen accession per individual landrace; 1 out of 30 resamples of 1 accession per landrace is included in the plot, and summary statistics for the remaining 29 resamples are presented in Table S3. Error bars represent the standard error. The sample size is reported in parentheses alongside the corresponding x-labels. $^{**}p < 0.01$, $^{***}p < 0.001$ (Mann-Whitney U tests).

See also Figures S1 and S2 and Table S3.

modern tropical japonica, modern indica, indica landraces, and japonica landraces. Within indica landraces, accessions from the same variety tended to cluster together, except for three Acuce accessions that clustered with Baijiao accessions and one Baijiao accession that clustered with Hongyang2 (Figure 1; Figure S1). Note that both networks are plotted at very different scales, with the RenSeq network of NLRs being approximately 84 times more extensive than the GBS network of genomic data. This indicates that the level of differentiation is much higher at NLRs compared to the remainder of the genome, consistent with directional (positive) selection on the NLRs and Red Queen coevolution.[18]

Clustering analyses further revealed that Baijiao accessions LB2 and LB6, and Hongjiao accession LHJ7, had mixed ancestry in multiple clusters at most K values (Figure S1), and did not branch with other accessions from the same varieties in the neighbor-net network (Figure 1). These accessions likely represent genetically introgressed (hybrid) lineages.

Demography of modern varieties and landraces

We used GBS sequences in non-NLR genes to explore how population history shaped patterns of genome-wide polymorphism in the different rice populations. Nucleotide diversity (π) differed significantly between populations (Kruskal-Wallis test: H = 11,330.1, df = 6, $p < 0.001$), and most post hoc pairwise comparisons were statistically significant ($p < 0.001$; Mann-Whitney U tests with Bonferroni-Holm correction; Table S3). The high level of variation contained in landrace accessions becomes apparent when this diversity is compared to the total variation possessed by modern varieties. For example, seven indica landrace accessions harbored almost as much nucleotide diversity (from π = 0.00089/bp to π = 0.00108/bp) as eleven modern indica varieties combined (π = 0.00107/bp) (here, the variation in landraces was calculated by resampling and including only one accession per population). Moreover, the seven indica landrace accessions possessed more nucleotide diversity than seventeen modern japonica varieties combined (π = 0.00060/bp; Figure 1C; Table S3; Figure S2A). Nucleotide diversity in individual landraces ranged from π = 0.00018/bp in Xiaogu to π = 0.00075/bp in Acuce (Figure 1C; Table S3; Figure S2A). These patterns of polymorphism indicate that farming practices have maintained a relatively high level of genome-wide standing variation in landraces.

The frequency distribution of polymorphisms, as measured by Tajima's D, indicated that modern landraces and varieties have experienced distinct evolutionary and/or demographic processes. The average D across GBS loci was positive but close to zero, the value expected under mutation-drift equilibrium, across all indica landraces as a group (30 resamples: mean Tajima's D = 0.114; range from D = −0.046 to D = 0.249). In contrast, at the scale of individual landraces, the average D indicated an excess of low-frequency variants in Acuce (D = −0.339), Baijiao (D = −0.242), and Xiaogu (D = −0.095). (D < 0 indicates a recent selective sweep, background selection, or population growth.) Conversely, there was a shift toward higher frequency alleles in Hongjiao (D = 0.686) and in the two populations of modern varieties (indica: D = 0.444; japonica: D = 0.473) (Table S3). (D > 0 indicates balancing selection, population structure, or population decline.) Overall, these patterns indicate that the individual landraces have followed distinct demographic histories and/or selection pressures and that they represent distinct populations with unique evolutionary histories.

To more accurately estimate the demographic history of the different rice groups, we used coalescent simulations within an approximate Bayesian computation (ABC) framework[25] to compare demographic models. The simulations were carried out with a Lambda coalescent with multiple mergers[26] as the demographic models consistently and significantly fitted better under this coalescent model (posterior probabilities: 1) than under a classical coalescent model (posterior probabilities: 0). While demographic models were fitted independently for indica and japonica, the demographic history of indica landraces was modeled by building a model integrating all individual landraces. The best-supported model was without gene flow, with exponential growth for Acuce and Baijiao, population expansion for Xiaogu, and population subdivision for Hongjiao. For indica and japonica, two-epoch models with population expansion had higher posterior probabilities. In the following, we use the demographic histories inferred from GBS loci as baselines to test for selection at NLRs.

Linking NLR diversity to function and presence/absence variation

Before testing the impact of selection at NLRs, we first examined the factors accounting for the molecular variability of NLRs. To test whether variation was evenly distributed among the different protein domains of NLRs, we used INTERPRO to define domains and computed summary statistics at the scale of domains. Nucleotide diversity (π) and the ratio of non-synonymous to synonymous nucleotide diversity (π_N/π_S) differed significantly between protein domains both at the species-wide scale (Kruskal-Wallis test: H = 18.4, df = 3, $p = 0.001$ for π; H = 20.4, df = 3, $p = 0.0004$ for π_N/π_S; Figure 2) and at the scale of individual varieties and landraces (Table S2). Nucleotide diversity at the leucine-reach repeats (LRRs) was significantly higher than nucleotide diversity at the coiled-coil (CC), and nucleotide-binding (NBARC[27]) domains (post hoc Mann-Whitney U tests: p = 0.016 and $p = 0.037$, respectively; Figure 2A), and the π_N/π_S at the LRR was significantly higher than π_N/π_S at the CC domain (post hoc Mann-Whitney U test: $p = 0.0006$; Figure 2A). Nucleotide diversity in coding sequence was most strongly correlated with nucleotide diversity in LRR compared to nucleotide diversity in the CC or NBARC domain (Figure 2B). The same pattern was observed using long-read PacBio data for 15 accessions of indica, japonica, and indica landraces, ruling out misalignments in small repetitive regions as the origin of the high variation observed in the LRR region (Figure S3). We conclude that LRR variation is the best predictor of NLR molecular diversity, consistent with the central role of the LRR domain in recognition and thus in trench-warfare coevolution with cognate ligands.

We used normalized read mapping depth to investigate the impact of presence/absence variation on the molecular variability of NLRs. At the species level, we found significant positive correlations between presence/absence diversity and nucleotide diversity (Spearman's rank correlation coefficient $\rho = 0.25$, $p < 0.001$) (Figure 3A). Species-wide nucleotide diversity π was significantly higher in core NLRs compared to accessory NLRs (π = 0.00455 for core NLRs, π = 0.00338 for accessory NLRs; core NLRs are present in all accessions of all subsamples of two accessions from a given population; Figure 3B), and the same pattern was observed at the population level, except in Xiaogu (Mann-Whitney U tests, $p < 0.0001$; Figure S2C). Although we observed no significant difference in Tajima's D between core NLRs and accessory NLRs (Mann-Whitney U test, $p = 0.779$), the maintenance of greater nucleotide diversity suggests balancing selection could be acting on core NLRs.

NLRs showed remarkable levels of presence/absence variation. Approximately 50% of the NLRs (358 NLRs out of the 596 NLRs used as reference sequences for read mapping) were present in all accessions of all populations of modern varieties and landraces, and these can be considered species-level core NLRs. Of the remainder, ca. 30% were present in less than 90% of all accessions (Figure 3C). At the population level, the number of core NLRs was similar in modern varieties (451 in japonica and 465 in indica) and in landraces (460 in Acuce, 482 in Xiaogu, 473 in Baijiao, and 459 in Hongjiao; Figure 3D), and most NLRs that were core in a given population were core in

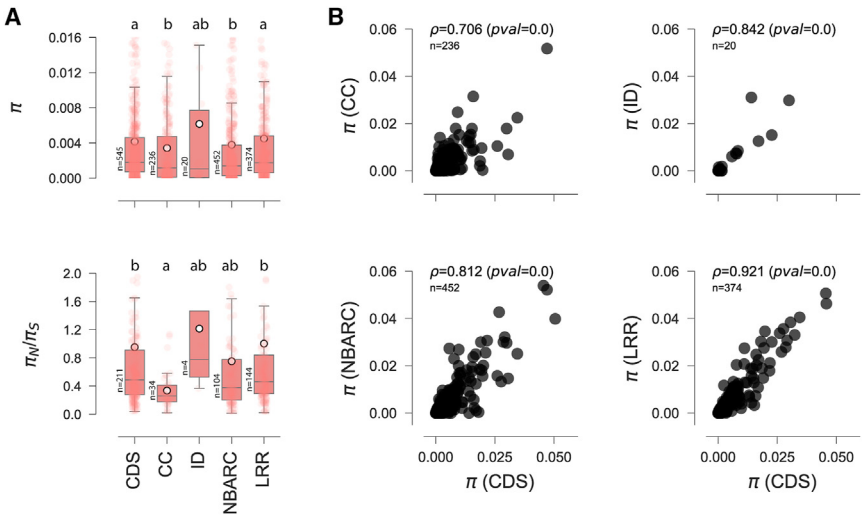

Figure 2. Patterns of nucleotide variation in NLR genes

(A) Species-wide nucleotide diversity (π) and ratio of non-synonymous to synonymous nucleotide diversity (π_N/π_S) in full coding sequence (CDS) and functional domains (CC, coiled-coil; ID, integrated domain[28]; NBARC, nucleotide-binding domain[27]; LRR, leucine-rich repeats).

(B) Correlation between nucleotide diversity in domains and full coding sequences. In (A), a number of data points were cropped from the nucleotide diversity plot for visually optimal presentation, but all data are included in statistical tests. In boxplots, the black circles represent the mean and the solid gray horizontal lines represent the median. The sample size n is reported alongside the corresponding boxplots. The letters a and b above the boxplots indicate whether the distributions are similar (when sharing the same letter), or significantly ($p < 0.05$) different, based on a Mann-Whitney U test. In (B), ρ is Spearman's rank correlation coefficient (***$p < 0.0001$) and n is the sample size.

See also Figures S2 and S3 and Table S2.

all populations (Figure S5A). Interestingly, though, the variation in the number of NLRs per population was higher in landraces than in modern varieties (Figure 3E). Presence/absence variation in NLR repertoires was significantly higher for landraces from different populations (median = 73) than for modern varieties of different japonica types (median = 65), (Mann-Whitney, W = 513434.0, $p = 0.0104$). In other words, two randomly picked plants from two landraces differ more in their NLR repertoire than two randomly picked plants from temperate or tropical rice populations. All NLRs from the chromosome 7 of the indica reference genome were missing in all accessions of Acuce and Xiaogu (Figure S4). Population-level presence frequency distributions followed the same reversed L-shaped distribution at the species-wide level (Figures S5B–S5G).

Impact of balancing selection on overall NLR variation

Comparisons of nucleotide diversity (π) at NLRs between landraces and varieties further revealed statistically significant differences between indica modern varieties and individual landraces, as well as between indica modern varieties and indica landraces as a group (i.e., measured using only one seed per bag of seeds; Kruskal-Wallis test: H = 814.1, df = 6, $p < 0.0001$; Mann-Whitney post hoc tests in Figure S2D). Individual landraces (a single "bag of seeds") harbored 7% (in Xiaogu) to 56% (in Acuce) of the total nucleotide diversity measured in modern indica. Even more remarkably, a single bag of seeds of Xiaogu and Acuce contained 17%–134%, respectively, of the nucleotide diversity measured in all modern japonica. Seven indica landraces displayed similar or significantly higher nucleotide diversity (from $\pi = 0.00343$/bp to $\pi = 0.00436$/bp; 30 resamples) than 4 modern indica ($\pi = 0.00343$/bp) and 6 modern japonica ($\pi = 0.00144$/bp; Figure 1; Table S3; Figure S2D). Levels of nucleotide diversity at NLRs that function as *R* genes against rice blast and other pathogens or pests were comparable between indica landraces and japonica varieties, but only half of the diversity observed in indica (Figure S2A; list of *R* genes in the figure's legend). The observed differences in nucleotide diversity indicate that traditional breeding of landraces maintained higher molecular diversity across the full complement of immune receptors, while breeding and improvement of modern varieties—for indica more specifically—instead injected diversity into a dozen widely used *R* NLR genes. Given that the census population size is likely to be considerably larger for the modern varieties, the difference in genomic erosion at immune receptors is likely to result from differences in selection pressures. In particular, modern varieties may have experienced more intense directional selection (potentially resulting in selective sweeps) and/or, conversely, landraces may have experienced more (human-mediated) balancing selection that maintained diversity at their NLRs. In other words, traditional farming practices may have generated spatiotemporal variation by avoiding planting seeds from previously susceptible plants, resulting in balancing selection on NLR variation.

To test for balancing selection at NLRs as a group, we corrected for the deviation from the standard demographic equilibrium by simulating datasets of a similar number of sequences and loci as the observed NLR datasets according to the best-supported demographic models estimated from GBS data. Nucleotide diversity π observed in RenSeq data was significantly higher than expected under neutrality in all populations ($p < 0.032$) except Xiaogu ($p = 0.074$), and Tajima's D was significantly higher than expected in all populations ($p < 0.01$) except Xiaogu and Hongjiao ($p = 0.946$ and $p = 0.164$, respectively). These signatures, too, are consistent with balancing selection.

Higher nucleotide diversity at NLRs was also supported by comparisons between GBS sequences from NLR and non-NLR genes (Table S4). The synonymous substitution rate was not significantly different between NLRs and non-NLRs (NLRs: dS = 0.25 [SD 0.31]; non-NLRs: dS = 0.34 [SD 0.30]; Mann-Whitney U test $p > 0.05$), which indicates that the mutation rate is not elevated at NLRs. The elevated nucleotide diversity at NLRs thus seems to be maintained by balancing selection, both in the modern varieties and landraces. Importantly, given

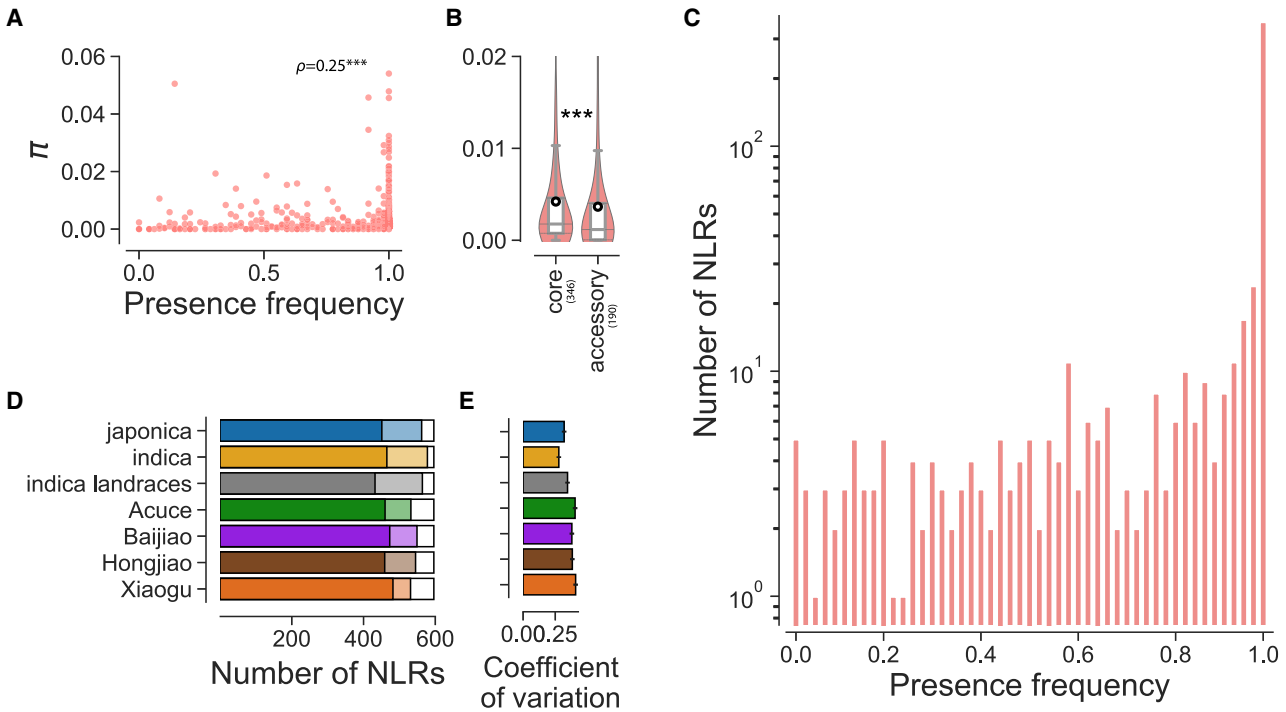

Figure 3. Presence-absence variation of 596 NLRs in 49 rice accessions

(A) Species-wide nucleotide diversity (π per bp) vs. presence frequency of NLRs; ρ is Spearman's rank correlation coefficient (***$p < 0.001$).

(B) Species-wide nucleotide diversity (π per bp) in core and accessory NLRs, ***$p < 0.001$; Mann-Whitney post hoc test with Holm-Bonferroni correction; in boxplots, dashed black line is mean, solid black line is median.

(C) Distribution of NLR presence frequency.

(D) Numbers of core (dark), accessory (light), and missing (white) NLRs, a core (missing) NLRs being present (absent) in all accessions of all subsamples of two accessions from a given population.

(E) Jackknife estimates of coefficient of variation in number of NLRs present, with error bars representing confidence intervals. In (D) and (E), the indica landraces group includes one randomly chosen accession per individual landrace; 1 out of 30 resamples of 1 accession per landrace is included in the plot. In (B), the black circles represent the mean and the solid gray horizontal lines represent the median; the sample size is reported in parentheses alongside the corresponding x-labels.

See also Figures S2, S4, and S5.

that the signature of elevated nucleotide diversity remains present long after balancing selection has ceased, we cannot rule out that balancing selection may have been historic rather than contemporary. However, our data do rule out that directional selection (or selective sweeps) has resulted in more severe genomic erosion of NLRs in modern varieties. Next, we examined whether landraces may have experienced more balancing selection than modern varieties, which could be expected if balancing selection ceased to operate in the modern varieties in present day.

Search for NLRs under balancing selection

To identify NLRs under balancing selection, we mapped the observed values of nucleotide diversity π and Tajima's D of each NLR on the joint density of (π, D) expected under neutrality while accounting for the demographic history of each group. NLRs were identified as under balancing selection if falling in the top 5% of π and D values calculated on datasets simulated for each individual NLR under the best-supported demographic models. Twenty-one and 19 NLRs were identified as being under balancing selection in modern indica and japonica. Eighty-eight NLRs were identified to be under balancing selection across all individual landraces, and 130.1 NLRs on average were identified as under balancing selection in 30 resamplings of one accession per landrace (standard deviation: 8.0; Figure 4A). Fifty-three NLRs were in the top 5% of π and D in Acuce, 23 NLRs in Baijiao,14 NLRs in Hongjiao, and none in Xiaogu and Hongyang2 (Figure 4A). Chromosome 11 harbored most of the NLRs in the top 5% of π and D in indica and indica landraces, with 16% of NLRs in indica, 19% in indica landraces, 16% in Acuce, and 44% in Baijiao. This analysis conclusively shows that, overall, NLRs in landraces appear to be under stronger balancing selection compared with NLRs in modern varieties, which could also explain why the NLR diversity of landraces is markedly elevated.

Heterozygote advantage and negative frequency-dependent selection are two types of balancing selection that maintain haplotypes in high frequencies in different populations. Such balancing selection reduces population genetic differentiation (F_{ST}). In contrast, spatiotemporal variation in selection pressures can result in a rapid turnover of alleles within a population.[29] This, too, is a form of balancing selection that helps to maintain the polymorphism across the entire metapopulation (i.e., all landraces), but it increases genetic differentiation between

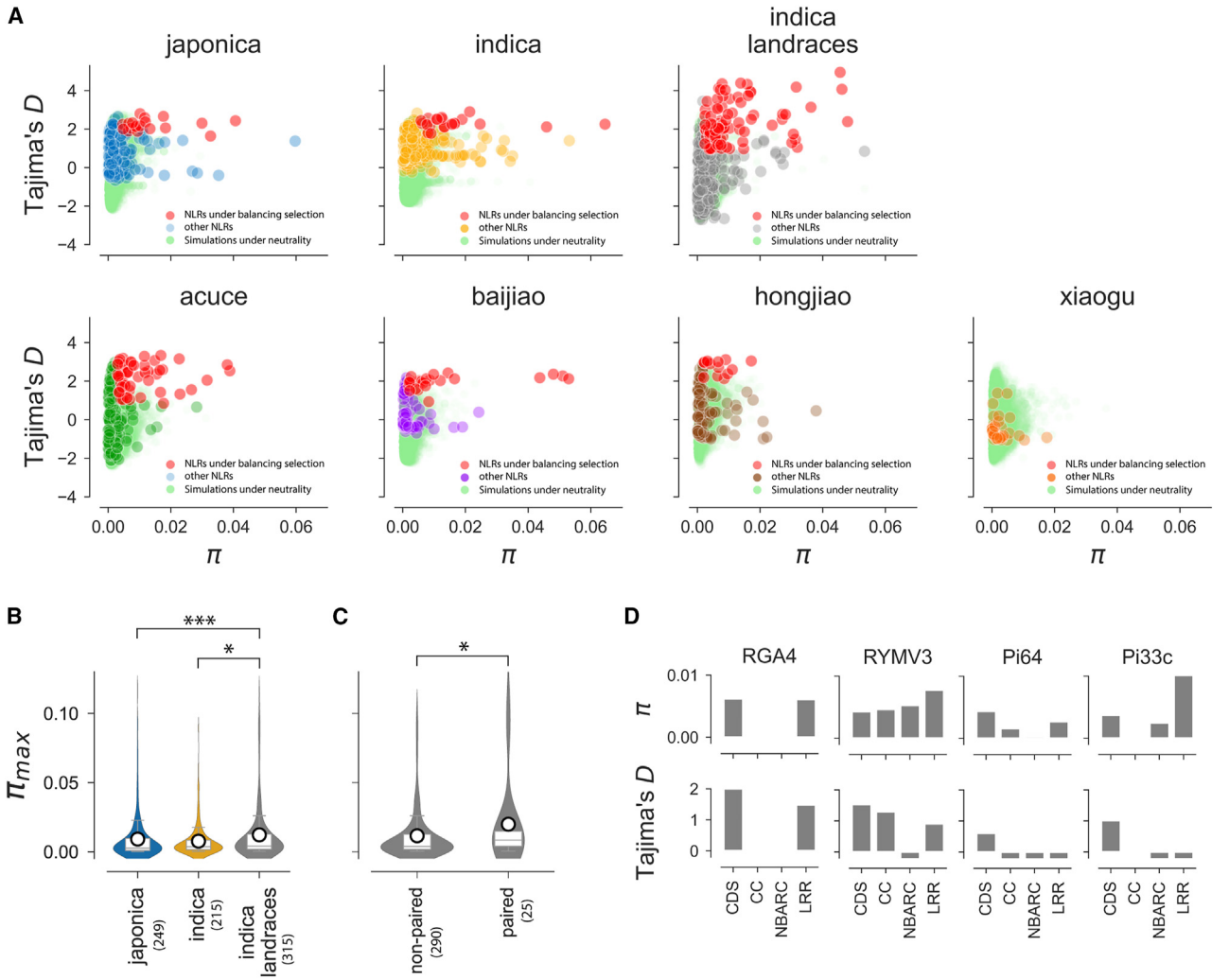

Figure 4. NLRs in indica landraces display signatures of balancing selection and enrichment in long-lived alleles

(A) Tajima's D and nucleotide diversity π in NLRs from modern japonica varieties, modern indica varieties, individual indica landraces, and indica landraces as a group; light green dots represent summary statistics computed on 10,000 datasets simulated for each NLR, with datasets being of the same sample size and sequence length as NLR sequences; simulations were carried out by sampling parameters from Gaussian kernel density estimates fitted to parameter posterior distributions of the best-supported demographic models; (B) π_{max} in modern indica varieties and indica landraces, computed for all NLRs, π_{max} being the maximum number of pairwise differences and measures the maximum depth of gene trees; (C) π_{max} in indica landraces, computed for non-paired and paired NLRs; (D) nucleotide diversity π and Tajima's D in full coding sequence (CDS) and functional domains for four *R* genes found to be under balancing selection in indica landraces. In (B) and (C), the black circles represent the mean and the solid gray horizontal lines represent the median; the sample size is reported in parentheses alongside the corresponding x-labels. Gaussian kernel density estimates were obtained using the PYTHON package SEABORN 0.11.2. *$p < 0.05$, ***$p < 0.001$, two-sided Mann-Whitney U tests with Bonferroni-Holm correction. The indica landraces group includes one randomly chosen accession per individual landrace; 1 out of 30 resamples of 1 accession per landrace is included in the plot.

See also Tables S4 and S5.

subpopulations (i.e., between landraces). We found that genetic differentiation (F_{ST}) among individual landraces was significantly higher at NLRs than other genes (on average, NLRs: $F_{ST} = 0.48$; other genes: $F_{ST} = 0.24$; Mann-Whitney U test $p < 0.001$). This is thus consistent with a type of balancing selection resulting from spatiotemporal variation in selection pressures, resulting in the maintenance of different sets of haplotypes in different landraces.

π_{max}, which represents the maximum number of pairwise differences and measures the maximum depth of gene genealogies, was significantly higher in indica landraces than modern indica and japonica varieties (one-sided Mann-Whitney U tests; $p < 0.05$; Figure 4B), indicating that indica landraces have kept older NLR alleles than modern varieties. For instance, unlike landraces, modern indica varieties lacked NLRs with π_{max} in the range 0.089–0.145, which corresponds to a minimal allelic divergence of T = 6.8 million years (assuming $\pi_{max} = 2\mu T$, with $\mu = 6.5e{-}9/bp$[30]). NLRs had deeper genealogies in indica landraces (T = 2.4 My; average π_{max}: 0.0317–0.0319, median π_{max}: 0.0204–0.0225, max π_{max}: 0.103) than in modern indica (T = 1.0 My; average π_{max}: 0.013, median π_{max}: 0.014; max π_{max}: 0.019) and modern japonica (T = 2.2 My; average π_{max}: 0.029,

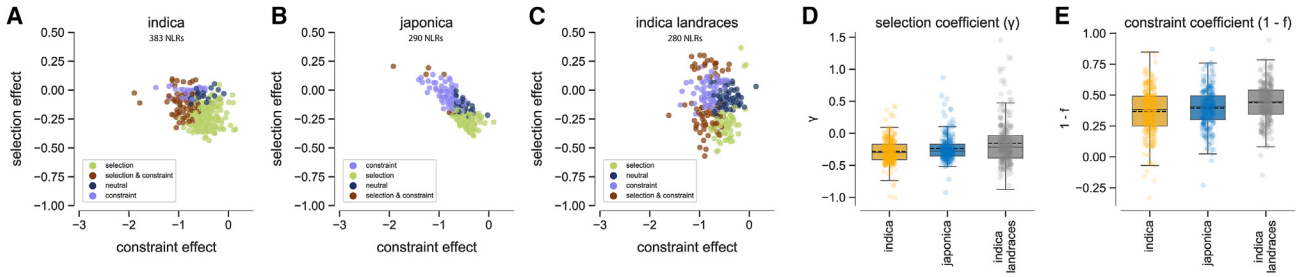

Figure 5. SNIPRE estimates of recurrent directional selection in 285 NLRs with outgroup data

(A) Selection and constraint effects in indica.

(B) Selection and constraint effects in japonica.

(C) Selection and constraint effects in indica landraces.

(D) Selection coefficients in indica, japonica, and indica landraces.

(E) Constraint coefficients in indica, japonica, and indica landraces. The selection effects reflect the selection coefficients (γ), with $\gamma > 0$ indicating positive selection and $\gamma < 0$ negative selection. The constraint (or non-synonymous) effects reflect mutational constraint ($1 - f$, f being the proportion of non-synonymous mutations that are not lethal). The indica landraces group includes one randomly chosen accession per individual landrace; 1 out of 30 resamples of 1 accession per landrace is included in the plot. The number of NLRs included in analyses is indicated alongside the names of the groups in (A)–(C).

median π_{max}: 0.020, max π_{max}: 0.078), (Figure 4C). These patterns indicate that, compared with landraces, ancient NLR polymorphisms have been lost due to more severe genomic erosion in modern varieties.

Examples of NLRs under balancing selection

Among NLRs under balancing selection in indica landraces featured gene *RGA4* (*BGIOSGA034263*), which is involved in resistance to rice blast. *RGA4* displayed relatively high values of nucleotide diversity π and Tajima's D in indica landraces as a group (π: 0.0064, percentile [π]: 82.9%, D: 3.569, percentile [D]: 96.5%). The high molecular diversity detected at *RGA4* was mostly driven by the LRR domain in indica landraces (Figure 4D). Amino acid changes in the LRR of RGA4 did not exhibit obvious clustering; they were primarily situated in the inner, unexposed residues of the LRR (data not shown). In addition to *RGA4*, 5 other NLRs out of the 32 NLRs with a signature of balancing selection in indica landraces, were involved in head-to-head pairs of NLRs, which is more than expected by chance (Fisher's exact test, p = 0.0014). More generally, paired NLRs harbored significantly more nucleotide diversity and more anciently diverged alleles than other NLRs in indica landraces (Figure S2E; one-sided Mann-Whitney U tests with Bonferroni-Holm corrections; π: H = 3,063.0, p = 0.030; π_{max}: H = 4,727.0, p = 0.006). Nucleotide diversity values within head-to-head pairs were also correlated (π: Spearman's ρ = 0.76, p = 0.005) (Figure S2E). Such signatures of coevolution were not observed for the *RGA4/RGA5* pair. RGA5, the NLR that binds to effectors AVR-Pia and AVR-CO39, was polymorphic at the species level but monomorphic in indica landraces and modern japonica and indica. *RGA5* was also less frequent than *RGA4*: although *RGA4* was detected in all 49 accessions, *RGA5* was present in only 22. The remaining accessions presumably carry the Pias-2 gene, which is allelic to RGA5 and functionally interacts with RGA4.[31]

Other *R* genes with signatures of balancing selection in resamples and individual landraces included *RYMV3* (resistance to rice yellow mottle virus[32]), *Pi64*, and *Pi33c* (resistance to rice blast[33–35]). All three *R* genes, as well as *RGA4*, displayed haplotypes exclusive to the landraces (data not shown).

Impact of recurrent directional selection on NLR variation

The observed deep gene genealogies of NLRs are indicative of ancient, pathogen-mediated balancing selection. To examine this further, we identified NLRs with ancient signatures of adaptation present in both modern varieties and landraces. We therefore used a Bayesian extension of the McDonald-Kreitman test implemented in the SNIPRE program.[36] In particular, we compared polymorphism and divergence at synonymous and non-synonymous sites in NLRs for which an orthologous sequence could be identified in the outgroup *O. barthii*. In both indica landraces and modern indica varieties, we detected widespread purifying selection against strongly deleterious mutations in almost all NLRs. Purifying selection has thus reduced the genetic load in almost all NLRs, indicating that their nucleotide sequence is functionally constrained. Nevertheless, between 1 and 4 NLRs from the 285 polymorphic NLRs with outgroup data showed evidence of directional selection across the 30 resampled groups of indica landraces. Three NLRs (*BGIOSGA027982*, *BGIOSGA040540*, and *BGIOSGA024574*) showed a consistent signature of directional selection, being flagged up in >14 of the 30 resampled groups of indica landraces (Figure 5). In contrast, none of the NLRs displayed significant directional selection in modern indica and japonica (Figure 5). Apparently, some NLRs show evidence of adaptive evolutionary change, possibly in response to changes in pathogen pressures, but this signature is only observed in indica landraces, not in modern indica and japonica.

Differences between the landraces and modern varieties were also revealed when analyzing and comparing the selection coefficient γ and the constraint coefficient $1 - f$ using the SNIPRE program. Both statistics were significantly different between indica landraces and modern indica varieties in 28 and 30 resampled indica landraces groups, respectively (Mann-Whitney U tests, $p < 0.05$). Furthermore, the mean values of both coefficients were greater in indica landraces than in modern indica varieties for all 30 resampled groups (Figure 5). Again, these analyses suggest that more adaptive evolutionary changes have been occurring in the indica landraces than in the modern

varieties or, alternatively, that more of these signatures have been retained in the landraces. Our data are consistent with the hypothesis that NLR diversity plays a role in population-level resistance against rice pathogens. We have also shown that a considerable amount of the diversity has been lost from the modern varieties, which shows that landraces provide a rich source of additional recognition capacities that could be recruited into modern varieties.

DISCUSSION

We used a combination of RenSeq sequence capture and Illumina sequencing to provide a comprehensive overview of nucleotide polymorphism at NLR-encoding loci in 49 rice accessions (*Oryza sativa*). In all modern and landrace populations, nucleotide diversity at NLRs was consistently higher than at other loci—and significantly higher than predicted from models of neutral evolution. NLRs thus appear to be highly variable in plant genomes at the intraspecific level, similar to other types of immune receptors outside the kingdom Plantae.[37–40] The high diversity of NLRs is consistent with their involvement in coevolutionary interactions with pathogen-derived ligands that impose strong selection on NLRs.[19,41] Pathogen-mediated selection can result in balancing selection (which maintains diversity) and/or directional selection (which results in changes in variation).[18] If directional selection changes across time and space, for example, due to changes in the composition of local pathogen communities, it can help to maintain diversity and act like balancing selection.[18] In this study, we carefully evaluate the (indirect) evolutionary genetic evidence for balancing selection against the alternative explanations that could potentially also result in similar signatures in genomic data.

Not all NLRs are hypervariable, and the observed range in diversity patterns included ~20% loci without polymorphism. Lack of diversity might reflect the fact that some NLRs can contribute to downstream signaling,[42] which may impose strong purifying selection to maintain the function. However, no sequenced-conserved helper NLRs was functionally defined in grasses. Lack of diversity may also stem from the fact that numerous NLRs detect pathogen signals indirectly and via host molecules modified by the pathogen termed guardees or decoys.[41] Such NLRs are therefore not in direct coevolution with the pathogen but are strongly constrained by the host molecules they guard. A well-studied example is ZAR1, which guards specific classes of receptor-like kinases and receptor-like cytoplasmic kinases and is broadly conserved in seed plants.[43] Strong directional selection (e.g., by a ubiquitous, dominant pathogen) can also reduce variation by fixing a (likely temporarily) selectively favored allele. The correlation between the diversity of the coding sequence of NLRs with the diversity of their LRR domains suggests that the main driver of the diversification of NLRs is the selective pressure exerted on this domain and, therefore, on the recognitional capacity of the NLR.

Another important observation is that indica landraces carry significantly more nucleotide diversity at NLRs than modern indica and japonica varieties. Similarly, the presence/absence variation in number of NLRs per population was also higher in landraces than in modern varieties. The high diversity is not only observed at the scale of the agroecosystem but also within landraces; there is as much nucleotide diversity in nine Acuce individuals as in six japonica modern varieties. In other words, nine individual plants of a landrace may possess as much NLR diversity as what could be found among billions of individuals of modern japonica. The lack of diversity in modern varieties is evidence of substantial genomic erosion, and it is likely to have negative coevolutionary consequences for the long-term sustainability of disease-resistant rice. In landraces, the high genetic diversity at immune receptors, combined with certain characteristics of traditional agrosystems (such as low levels of nitrogen inputs), could contribute to reducing the disease burden by promoting the emergence or maintenance of a generalist lifestyle in associated pathogens.[44]

Our findings suggest that balancing selection contributes significantly to the high diversity in NLRs of landraces. Both in modern varieties and landraces, nucleotide diversity was higher at the NLRs relative to other genes. However, after controlling for differences in demographic histories, we found a greater number of genes with signatures of balancing selection across indica landraces than in the modern varieties. We also found that, compared with landraces, modern varieties are depleted in ancient polymorphisms. This observation is consistent with the evidence of stronger balancing selection (i.e., higher selection coefficient γ) and with the fact that several NLRs were only found to be under balancing selection in indica landraces. Various ecological, demographic, and evolutionary processes lead to deviations from neutrality[45] and mimic the signatures of balancing selection. For example, relatively elevated levels of nucleotide diversity and positive Tajima's D could be caused by genetic introgression from a wild source population with high genetic diversity. However, the higher rate of non-synonymous diversity is discordant with introgression from a diverse source population (with large effective population size), as more efficient natural selection would reduce the π_N/π_S in such a large population. With three lines of evidence (nucleotide diversity, Tajima's D, and π_N/π_S), we argue that balancing selection is the most plausible explanation for our results.

In addition to genes with signatures of balancing selection (i.e., genes with an excess of nucleotide diversity) being exclusively found in indica landraces, we also identified three genes with signatures of recurrent directional selection (i.e., genes with an excess of non-synonymous substitutions) that were unique to this group. The number of genes under strong directional selection is likely underestimated as only 285 NLRs displayed outgroup data and could thus be included in the analysis. Regardless of this limitation, we were surprised to find fewer signals of directional than balancing selection in NLRs in general, as modeling of finite populations also suggests that signatures of directional selection are more likely to be observable than signatures of balancing selection.[46] We also did not expect to detect evidence for directional selection only in landraces and not in modern varieties. Naively, one might expect arms-race coevolution to be more widespread in modern agroecosystems, where NLRs with new resistance specificities are deployed and quickly overcome (boom-and-bust dynamics). However, such coevolution could also play a role in traditional agroecosystems, where farmers have used the same landraces for centuries, being able to select from a multitude of landraces. The resistance breaking of a single NLR allele is then of little concern because

there are other landraces with different alleles that confer resistance. The susceptible allele may then be replaced, or at least reduced in prevalence, leaving a signature of directional selection but without risk to rice harvest. Although this is a plausible hypothesis, it may also be the low diversity in modern varieties that hampers the detection of NLRs under positive selection in these lines.

An interesting case of an NLR under balancing selection in the landraces is RGA4. RGA4 is a helper NLR that interacts functionally with the sensor NLR RGA5, which recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39, or with the sensor NLR Pias-2 that detects the sequence-unrelated *M. oryzae* effector AVR-Pias.[31,47–49] RGA5 carries a non-canonical heavy metal-associated (HMA) domain after its LRR that directly binds AVR-Pia and AVR1-CO39, which is crucial for their detection.[50,51] Pias-2 harbors a completely different integrated domain[28] of unknown function (DUF761), whose role in effector recognition remains unknown.[31] The canonical NLR domains of RGA5 and Pias-2 have limited sequence similarity (59.8% identity), while the *RGA4* alleles coupled either with *RGA5* or with *Pias2* are highly identical (96.6%) and functionally interchangeable. The *RGA4*/*RGA5* and *Pias1*/*Pias2* haplotypes also occur in wild rice species together with four additional *RGA5* alleles that have even other integrated domains.[31] Population genetics and comparative genomics analyses indicate that balancing selection maintains these multiple *RGA4*/*RGA5* alleles with contrasting recognition specificities across speciation in multiple species of the *Oryza* genus.[31,48] Our study shows that such balancing selection occurs also at the population level, within landraces, thereby potentially providing complementary protection against isolates with different virulence effectors.

Interestingly, previous work has shown that, in the Yuanyang terraces, the effector AVR-Pia is absent from most *M. oryzae* isolates collected on indica landraces, while it occurs at high frequency in isolates from japonica rice on which it confers a gain in virulence.[9] Here, we report that RGA4 is present at high frequency and under strong balancing selection in indica landraces, with high diversity in the LRR domain. Under the interaction model described above, it is probably RGA5 or Pias-2 that is the main target of coevolutionary interactions with fungal effectors, and the signatures of balancing selection detected in RGA4 are a byproduct that results from compensatory changes in the helper induced by coevolution-driven changes in the sensors.

Rice NLRs show not just SNP variation, but also presence/absence variation. Similar to the SNP variation, the presence/absence variation in number of NLRs within landraces was higher than in japonica or indica varieties. When all accessions are compared, a slight majority of NLRs (~350/~600) are present in all investigated accessions (Figures 3A and 3C—note the log scale in Figure 3C). A few NLRs are present in a minority of accessions. Other researchers have classified the conserved and presence/absence variable NLRs as "core" and "dispensable,"[52] though we would argue that because polymorphism for recognition capacity underpins resistance and selects against specialist races of *Magnaporthe oryzae*,[53,54] the term dispensable conveys a misleading impression of lack of utility. Indeed, it is the absence of certain allelic variants in accessions that underpins the variation needed in Red Queen coevolution. As so eloquently noted by Lewontin: "a genotype is its own worst enemy, its fitness will decrease as it becomes more common."[55] Hence, it may be the rarity of dispensable NLRs that gives them their adaptive coevolutionary advantage.

In aggregate, our work shows that rice NLRs represent a highly variable gene family and that this variability is particularly high in landraces from the Yuanyang terraces. Our data mirrors the patterns found across other crops as well as livestock in that there is a worrying disparity between total biomass and genetic diversity, which makes the species we most rely on for our food security vulnerable to emerging infectious diseases.[18] We found hints in the data for positive selection, but indications of balancing selection were more evident and pervasive than indications of directional selection. Therefore, the data tend to provide more support to the trench-warfare hypothesis over the arms-race hypothesis as a general coevolutionary model for this class of genes. Integrated domains and LRRs seem to be the preferred target of balancing selection, consistent with their role in the recognition of pathogen-derived ligands. The effect of trench warfare is visible in the maintenance of high values of the π_{max} statistic in the landraces, which indicates the maintenance of ancient NLR alleles in these populations. Understanding how elevated NLR diversity and enrichment in older alleles reduces the burden of disease in traditional agroecosystems gives guidance for re-engineering modern crops and agroecosystems to make them less conducive to extant and emerging diseases.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Design of a rice RenSeq bait-library and application to explore NLR diversity
  - Empirical distribution of genome-wide polymorphism
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Presence/absence variation
  - Population subdivision
  - Polymorphism and divergence
  - Assessing the impact of sampling effort on measures of molecular variability
  - Demographic modeling
  - Directional selection

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2024.07.061.

ACKNOWLEDGMENTS

We thank Peter Balint-Kurti for critical reading of the manuscript, Stephane de Mita for help with random forest ABC and EGGLIB, and TSL's Dan MacLean and Robert Heal for useful suggestions. This work was supported by a grant from the Gordon and Betty Moore Foundation to the 2Blades Foundation (GBMF4725) (J.D.G.J., A.J., and S.F.), by the Gatsby Charitable Foundation (A.J., S.F., and J.D.G.J.), by BBSRC grant BB/P021646/1 (S.F. and

J.D.G.J.), by the Earth and Life Systems Alliance (ELSA) (A.J. and C.v.O.), and by Yunnan Fundamental Research Projects (202301AS070257) (H.H.).

AUTHOR CONTRIBUTIONS

J.-B.M., H.H., T.K., and J.D.G.J. designed the study; P.G. conducted population genetic analyses with help from C.v.O.; P.G. and S.F. conducted presence/absence variation analyses with help from C.v.O.; S.F., A.J., J.F., and D.O. generated the data; S.R., J.-B.M., and T.K. designed sequence capture baits; R.-K.S. conducted SNP calling; X.H., Y.Z., J.-B.M., and H.H. collected the biological materials; P.G., C.v.O., S.F., T.K., and J.D.G.J. interpreted the results; P.G., C.v.O., T.K., and J.D.G.J. wrote the manuscript, with final approval from all the co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests

Received: February 16, 2023
Revised: April 19, 2024
Accepted: July 16, 2024
Published: August 14, 2024

REFERENCES

1. Tamborski, J., and Krasileva, K.V. (2020). Evolution of plant NLRs: from natural history to precise modifications. Annu. Rev. Plant Biol. *71*, 355–378.
2. Ebert, D., and Fields, P.D. (2020). Host-parasite co-evolution and its genomic signature. Nat. Rev. Genet. *21*, 754–768.
3. Karasov, T.L., Shirsekar, G., Schwab, R., and Weigel, D. (2020). What natural variation can teach us about resistance durability. Curr. Opin. Plant Biol. *56*, 89–98.
4. Browning, J.A., and Frey, K.J. (1969). Multiline cultivars as a means of disease control. Annu. Rev. Phytopathol. *7*, 355–382.
5. Wolfe, M.S. (1985). The current status and prospects of multiline cultivars and variety mixtures for disease resistance. Annu. Rev. Phytopathol. *23*, 251–273.
6. Villa, T.C.C., Maxted, N., Scholten, M., and Ford-Lloyd, B. (2005). Defining and identifying crop landraces. Plant Genet. Resour. *3*, 373–384.
7. Thurston, H.D. (1990). Plant disease management practices of traditional farmers. Plant Dis. *74*, 96–102.
8. He, X., Sun, Y., Gao, D., Wei, F., Pan, L., Guo, C., Mao, R., Xie, Y., Li, C., and Zhu, Y. (2011). Comparison of Agronomic Traits between Rice Landraces and Modern Varieties at Different Altitudes in the Paddy Fields of Yuanyang Terrace, Yunnan Province. J. Resour. Ecol. *2*, 46–50.
9. Liao, J., Huang, H., Meusnier, I., Adreit, H., Ducasse, A., Bonnot, F., Pan, L., He, X., Kroj, T., Fournier, E., et al. (2016). Pathogen effectors and plant immunity determine specialization of the blast fungus to rice subspecies. eLife *5*, e19377.
10. 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.
11. Hannachi, M., and Dedeurwaerdere, T. (2018). Des semences en commun pour gérer les maladies. Étude comparative de rizières dans le Yuanyuang (Chine). Etudes rurales *202*, 76–97.
12. May, R.M., and Anderson, R.M. (1983). Epidemiology and genetics in the coevolution of parasites and hosts. Proc. R. Soc. Lond. B Biol. Sci. *219*, 281–313.
13. Thrall, P.H., Oakeshott, J.G., Fitt, G., Southerton, S., Burdon, J.J., Sheppard, A., Russell, R.J., Zalucki, M., Heino, M., and Ford Denison, R. (2011). Evolution in agriculture: the application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. Evol. Appl. *4*, 200–215.
14. Williams, P.D. (2010). Darwinian interventions: taming pathogens through evolutionary ecology. Trends Parasitol. *26*, 83–92.
15. Van Oosterhout, C. (2021). Mitigating the Threat of Emerging Infectious Diseases; a Coevolutionary Perspective (Taylor & Francis).
16. Karasov, T.L., Horton, M.W., and Bergelson, J. (2014). Genomic variability as a driver of plant–pathogen coevolution? Curr. Opin. Plant Biol. *18*, 24–30.
17. Märkle, H., Saur, I.M.L., and Stam, R. (2022). Evolution of resistance (R) gene specificity. Essays Biochem. *66*, 551–560.
18. Lighten, J., Papadopulos, A.S.T., Mohammed, R.S., Ward, B.J., Paterson, G., I., Baillie, L., Bradbury, I.R., Hendry, A.P., Bentzen, P., et al. (2017). Evolutionary genetics of immunological supertypes reveals two faces of the Red Queen. Nat. Commun. *8*, 1.
19. Bakker, E.G., Toomajian, C., Kreitman, M., and Bergelson, J. (2006). A genome-wide survey of R gene polymorphisms in Arabidopsis. Plant Cell *18*, 1803–1818.
20. Witek, K., Lin, X., Karki, H.S., Jupe, F., Witek, A.I., Steuernagel, B., Stam, R., Van Oosterhout, C., Fairhead, S., Heal, R., et al. (2021). A complex resistance locus in Solanum americanum recognizes a conserved Phytophthora effector. Nat. Plants *7*, 198–208.
21. Witek, K., Jupe, F., Witek, A.I., Baker, D., Clark, M.D., and Jones, J.D.G. (2016). Accelerated cloning of a potato late blight-resistance gene using RenSeq and SMRT sequencing. Nat. Biotechnol. *34*, 656–660.
22. Sheng, G. (1990). Yuanyang county chronicles. in Chinese (Gui Zhou National Press), pp. 94–127.
23. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. Genetics *196*, 973–983.
24. Huson, D.H., and Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. Mol. Biol. Evol. *23*, 254–267.
25. Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., and François, O. (2010). Approximate Bayesian computation (ABC) in practice. Trends Ecol. Evol. *25*, 410–418.
26. Tellier, A., and Lemaire, C. (2014). Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. Mol. Ecol. *23*, 2637–2652.
27. van der Biezen, E.A., and Jones, J.D.G. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr. Biol. *8*, R226–R227.
28. Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P.N. (2014). A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. Front. Plant Sci. *5*, 606.
29. McMullan, M., and Van Oosterhout, C. (2012). Inference of selection based on temporal genetic differentiation in the study of highly polymorphic multigene families. PLoS One *7*, e42119.
30. Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. Proc. Natl. Acad. Sci. USA *93*, 10274–10279.
31. Shimizu, M., Hirabuchi, A., Sugihara, Y., Abe, A., Takeda, T., Kobayashi, M., Hiraka, Y., Kanzaki, E., Oikawa, K., Saitoh, H., et al. (2022). A genetically linked pair of NLR immune receptors shows contrasting patterns of evolution. Proc. Natl. Acad. Sci. USA *119*, e2116896119.
32. Bonnamy, M., Pinel-Galzi, A., Gorgues, L., Chalvon, V., Hébrard, E., Chéron, S., Nguyen, T.H., Poulicard, N., Sabot, F., Pidon, H., et al. (2023). Rapid evolution of an RNA virus to escape recognition by a rice nucleotide-binding and leucine-rich repeat domain immune receptor. New Phytol. *237*, 900–913.
33. Ma, J., Lei, C., Xu, X., Hao, K., Wang, J., Cheng, Z., Ma, X., Ma, J., Zhou, K., Zhang, X., et al. (2015). Pi64, encoding a novel CC-NBS-LRR protein, confers resistance to leaf and neck blast in rice. Mol. Plant Microbe Interact. *28*, 558–568.
34. Ballini, E., Berruyer, R., Morel, J.B., Lebrun, M.H., Nottéghem, J.L., and Tharreau, D. (2007). Modern elite rice varieties of the 'Green Revolution'

have retained a large introgression from wild rice around the Pi33 rice blast resistance locus. New Phytol. *175*, 340–350.

35. Zhou, B., Qu, S., Liu, G., Dolan, M., Sakai, H., Lu, G., Bellizzi, M., and Wang, G.-L. (2006). The eight amino acid differences within three leucine-rich repeats between Pi2 and Piz-t resistance proteins determine the resistance specificity to Magnaporthe grisea. Mol. Plant Microbe Interact. *19*, 1216–1228.
36. Eilertson, K.E., Booth, J.G., and Bustamante, C.D. (2012). SnIPRE: selection inference using a Poisson random effects model. PLoS Comput. Biol. *8*, e1002806.
37. Russell, R.M., Bibollet-Ruche, F., Liu, W., Sherrill-Mix, S., Li, Y., Connell, J., Loy, D.E., Trimboli, S., Smith, A.G., Avitto, A.N., et al. (2021). CD4 receptor diversity represents an ancient protection mechanism against primate lentiviruses. Proc. Natl. Acad. Sci. USA *118*, e2025914118.
38. Chapman, J.R., Hill, T., and Unckless, R.L. (2019). Balancing Selection Drives the Maintenance of Genetic Variation in Drosophila Antimicrobial Peptides. Genome Biol. Evol. *11*, 2691–2701.
39. Zhao, J., Gladieux, P., Hutchison, E., Bueche, J., Hall, C., Perraudeau, F., and Glass, N.L. (2015). Identification of Allorecognition Loci in Neurospora crassa by Genomics and Evolutionary Approaches. Mol. Biol. Evol. *32*, 2417–2432.
40. Heller, J., Clavé, C., Gladieux, P., Saupe, S.J., and Glass, N.L. (2018). NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. Proc. Natl. Acad. Sci. USA *115*, E2292–E2301.
41. Cesari, S. (2018). Multiple strategies for pathogen perception by plant immune receptors. New Phytol. *219*, 17–24.
42. Feehan, J.M., Castel, B., Bentham, A.R., and Jones, J.D. (2020). Plant NLRs get by with a little help from their friends. Curr. Opin. Plant Biol. *56*, 99–108.
43. Adachi, H., Sakai, T., Kourelis, J., Pai, H., Gonzalez Hernandez, J.L., Utsumi, Y., Seki, M., Maqbool, A., and Kamoun, S. (2023). Jurassic NLR: conserved and dynamic evolutionary features of the atypically ancient immune receptor ZAR1. Plant Cell *35*, 3662–3685.
44. 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.
45. de Jong, M.J., van Oosterhout, C., Hoelzel, A.R., and Janke, A. (2024). Moderating the neutralist–selectionist debate: exactly which propositions are we debating, and which arguments are valid? Biol. Rev. Camb. Philos. Soc. *99*, 23–55.
46. Tellier, A., Moreno-Gámez, S., and Stephan, W. (2014). Speed of adaptation and genomic footprints of host–parasite coevolution under arms race and trench warfare dynamics. Evolution *68*, 2211–2224.
47. Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., et al. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding. Plant Cell *25*, 1463–1481.
48. Okuyama, Y., Kanzaki, H., Abe, A., Yoshida, K., Tamiru, M., Saitoh, H., Fujibe, T., Matsumura, H., Shenton, M., Galam, D.C., et al. (2011). A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. Plant J. *66*, 467–479.
49. Césari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., and Kroj, T. (2014). The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. EMBO J. *33*, 1941–1959.
50. Ortiz, D., De Guillen, K., Cesari, S., Chalvon, V., Gracy, J., Padilla, A., and Kroj, T. (2017). Recognition of the Magnaporthe oryzae effector AVR-Pia by the decoy domain of the rice NLR immune receptor RGA5. Plant Cell *29*, 156–168.
51. Guo, L., Cesari, S., de Guillen, K., Chalvon, V., Mammri, L., Ma, M., Meusnier, I., Bonnot, F., Padilla, A., Peng, Y.-L., et al. (2018). Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. Proc. Natl. Acad. Sci. USA *115*, 11637–11642.
52. Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. *50*, 278–284.
53. Couch, B.C., Fudal, I., Lebrun, M.-H., Tharreau, D., Valent, B., van Kim, P., Nottéghem, J.-L., and Kohn, L.M. (2005). Origins of Host-Specific Populations of the Blast Pathogen *Magnaporthe oryzae* in Crop Domestication With Subsequent Expansion of Pandemic Clones on Rice and Weeds of Rice. Genetics *170*, 613–630.
54. Tosa, Y., Osue, J., Eto, Y., Oh, H.-S., Nakayashiki, H., Mayama, S., and Leong, S.A. (2005). Evolution of an avirulence gene, AVR1-CO39, concomitant with the evolution and differentiation of Magnaporthe oryzae. Mol. Plant Microbe Interact. *18*, 1148–1160.
55. Lewontin, R.C. (1974). The Genetic Basis of Evolutionary Change (Columbia University Press).
56. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. Nucleic Acids Res. *47*, D427–D432.
57. Steuernagel, B., Jupe, F., Witek, K., Jones, J.D.G., and Wulff, B.B.H. (2015). NLR-parser: rapid annotation of plant NLR complements. Bioinformatics *31*, 1665–1667.
58. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
59. Bonfield, J.K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., and Davies, R.M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data. GigaScience *10*, giab007. https://doi.org/10.1093/gigascience/giab007.
60. 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.
61. Tsagkogeorga, G., Cahais, V., and Galtier, N. (2012). The population genomics of a fast evolver: high levels of diversity, functional constraint, and molecular adaptation in the tunicate Ciona intestinalis. Genome Biol. Evol. *4*, 740–749.
62. Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., Lourenco, J.M., Alves, P.C., Ballenghien, M., Faivre, N., et al. (2013). Reference-free population genomics from next-generation transcriptome data and the vertebrate–invertebrate gap. PLoS Genet. *9*, e1003457.
63. Nabholz, B., Sarah, G., Sabot, F., Ruiz, M., Adam, H., Nidelet, S., Ghesquière, A., Santoni, S., David, J., and Glémin, S. (2014). Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (Oryza glaberrima). Mol. Ecol. *23*, 2210–2227.
64. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31.
65. Miles, A., pyup.io bot, Murillo, R., Ralph, P., Harding, N., Pisupati, R., Rae, S., and Millar, T. (2021). Cggh/Scikit-Allel: V1. 3.3. Version v1. 3.3. Zenodo. https://doi.org/10.5281/zenodo.4759368.
66. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722–736.
67. Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. *33*, W465–W467.
68. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.

69. Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. *9*, 286–298.

70. Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. *33*, 511–518.

71. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

72. Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*, W7–W13.

73. Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. *15*, 1179–1191.

74. Siol, M., Coudoux, T., Ravel, S., and De Mita, S. (2022). EggLib 3: A python package for population genetics and genomics. Mol. Ecol. Resour. *22*, 3176–3187.

75. Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. *13*, 555–556.

76. Collin, F.-D., Estoup, A., Marin, J.-M., and Raynal, L. (2020). Bringing ABC inference to the machine learning realm: AbcRanger, an optimized random forests library for ABC. JOBIM 2020, 66. https://hal.science/hal-02910067v1/file/jobim_proceedings.pdf.

77. Collin, F.D., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., Marin, J.M., and Estoup, A. (2021). Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. Mol. Ecol. Resour. *21*, 2598–2613.

78. Pickrell, J., and Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.

79. Nelson, D., Kelleher, J., Ragsdale, A.P., Moreau, C., McVean, G., and Gravel, S. (2020). Accounting for long-range correlations in genome-wide simulations of large cohorts. PLOS Genet. *16*, e1008619.

80. Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A.P., Tsambos, G., Zhu, S., Eldon, B., Ellerman, E.C., Galloway, J.G., et al. (2022). Efficient ancestry and mutation simulation with msprime 1.0. Genetics *220*, iyab229.

81. Kelleher, J., Etheridge, A.M., and McVean, G. (2016). Efficient coalescent simulation and genealogical analysis for large sample sizes. PLoS Comput. Biol. *12*, e1004842.

82. Glaszmann, J.C., Mew, T., Hibino, H., Kim, C.K., Vergel de Dios-Mew, T.I., Vera Cruz, C.M., Nottéghem, J.L., and Bonman, J.M. (1996). Molecular variation as a diverse source of disease resistance in cultivated rice. In Rice Genetics III: (In 2 Parts) (World Scientific), pp. 460–465.

83. Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J., and Zhou, S. (2013). Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice *6*, 4.

84. Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., and Zhang, X. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science *296*, 79–92.

85. Luo, S., Zhang, Y., Hu, Q., Chen, J., Li, K., Lu, C., Liu, H., Wang, W., and Kuang, H. (2012). Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. Plant Physiol. *159*, 197–210.

86. Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. *8*, 4321–4325.

87. 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.

88. Ali, S., Gladieux, P., Ravel, S., Adreit, H., Meusnier, I., Milazzo, J., Cross-Arteil, S., Bonnot, F., Jin, B., and Dumartinet, T. (2021). Coevolution with Spatially Structured Rice Landraces Maintains Multiple Generalist Lineages in the Rice Blast Pathogen. Preprint at bioRxiv.

89. Rakotoson, T., Dusserre, J., Letourmy, P., Frouin, J., Ratsimiala, I.R., Rakotoarisoa, N.V., cao, T.-V., Vom Brocke, K., Ramanantsoanirina, A., Ahmadi, N., et al. (2021). Genome-wide association study of nitrogen use efficiency and agronomic traits in upland rice. Rice Sci. *28*, 379–390.

90. Frouin, J., Languillaume, A., Mas, J., Mieulet, D., Boisnard, A., Labeyrie, A., Bettembourg, M., Bureau, C., Lorenzini, E., Portefaix, M., et al. (2018). Tolerance to mild salinity stress in japonica rice: A genome-wide association mapping study highlights calcium signaling and metabolism genes. PLoS One *13*, e0190964.

91. Van de Weyer, A.-L., Monteiro, F., Furzer, O.J., Nishimura, M.T., Cevik, V., Witek, K., Jones, J.D.G., Dangl, J.L., Weigel, D., and Bemm, F. (2019). A species-wide inventory of NLR genes and alleles in Arabidopsis thaliana. Cell *178*, 1260–1272.e14.

92. Marth, G.T., Czabarka, E., Murvai, J., and Sherry, S.T. (2004). The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. Genetics *166*, 351–372.

93. Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., and Robert, C.P. (2016). Reliable ABC model choice via random forests. Bioinformatics *32*, 859–866.

94. Csilléry, K., François, O., and Blum, M.G.B. (2012). abc: an R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. *3*, 475–479.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Tris-EDTA	Sigma-Aldrich	93283
Chloroform-isoamyl alcohol 24:1	Sigma-Aldrich	C0549
CTAB	Sigma-Aldrich	H6269
Deposited data		
Raw reads (Illumina sequencing)	This study	GenBank: PRJEB23459
Raw reads (PacBio sequencing)	This study	GenBank: PRJEB29200
Experimental models: Organisms/strains		
Landrace and modern rice accessions	Table S1	N/A
Software and algorithms		
PfamScan	El-Gebali et al.[56]	https://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/
NLR Parser v1	Steuernagel et al.[57]	https://github.com/steuernb/NLR-Parser
bowtie v2.3.5	Langmead and Salzberg[58]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools v1.9	Bonfield et al.,[59] Danecek et al.[60]	http://www.htslib.org/
Reads2snp 2.0	Tsagkogeorga et al.,[61] Gayral et al.,[62] Nabholz et al.[63]	https://kimura.univ-montp2.fr/PopPhyl/index.php?section=tools
Exonerate	Slater and Birney[64]	https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
R Project for statistical computing v. 4.3.2	R Core Team, 2023	https://www.r-project.org/
scikit-allel v. 1.3.3 (Python package)	Miles et al.[65]	https://github.com/cggh/scikit-allel
Canu v2.0	Koren et al.[66]	https://github.com/marbl/canu
Augustus 3.5	Stanke and Morgenstern[67]	https://github.com/Gaius-Augustus/Augustus
Interproscan 5.60-92.0	Jones et al.[68]	https://interproscan-docs.readthedocs.io/en/latest/
MAFFT	Katoh and Toh,[69] Katoh et al.[70,71]	https://mafft.cbrc.jp/alignment/software/
TranslatorX	Abascal et al.[72]	https://translatorx.co.uk/
astropy.stats (Python package)	N/A	https://docs.astropy.org/en/stable/stats/
sNMF	Frichot et al.[23]	https://www.rdocumentation.org/packages/LEA/versions/1.4.0
Clumpak	Kopelman et al.[73]	http://clumpak.tau.ac.il/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Splitstree 5	Huson and Bryant[24]	https://software-ab.cs.uni-tuebingen.de/download/splitstree5/welcome.html
Egglib 3 (Python package)	Siol et al.[74]	https://www.egglib.org/
scikit_posthocs 0.6.6 (Python package)	N/A	https://pypi.org/project/scikit-posthocs/
scipy 1.8.0 (Python package)	N/A	https://docs.scipy.org/doc/scipy/release/1.8.0-notes.html
Codeml	Yang[75]	http://abacus.gene.ucl.ac.uk/software/paml.html
pyabcranger 0.0.70 (Python package)	Collin et al.[76,77]	https://pypi.org/project/pyabcranger/
Treemix	Pickrell et al.[78]	https://bitbucket.org/nygcresearch/treemix/wiki/Home
msprime (Python package)	Nelson et al.,[79] Baumdicker et al.,[80] Kelleher et al.[81]	https://tskit.dev/software/msprime.html
SnIPRE	Eilertson et al.[36]	N/A
Other		
Single-nucleotide polymorphism datasets, and code for their analysis	This study	Zenodo Data: https://doi.org/10.5281/zenodo.12666118

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Pierre Gladieux (pierre.gladieux@inrae.fr)

Materials availability

This study did not generate new unique reagents.

Data and code availability

RenSeq Illumina and PacBio sequencing data have been deposited at NCBI - Short Sequence Archive and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Single-nucleotide polymorphism datasets have been deposited at Zenodo and are publicly available as of the date of publication. DOI is listed in the key resources table.

All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOI is listed in the key resources table.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

RenSeq sequence capture and Illumina sequencing was performed on 49 rice accessions (Table S1), representing seven indica landraces (36 accessions), two japonica landraces (two accessions) and eleven modern varieties of indica, japonica, and aus. The nine landraces were represented by thirty-eight accessions, which are part of a rice diversity collection (single panicle descendants) established in 2014 and 2015 by sampling individual plants in the fields of traditional rice farmers in three villages (Xiaoshuijing, Xinjie, Qingkou) in the Yuanyang rice terrace region in Yunnan province (China). Thirty-one landrace accessions were selected as representatives of the genetic diversity of four major landraces cultivated in this region: Acuce (7 accessions), Baijiao (9 accessions), Hongjiao (8 accessions) and Xiaogu (8 accessions). Three accessions correspond to the Hongyang2 variety, a true breeding line bred from landrace germplasm and widely cultivated in the Yuanyang terraces in recent years. Three accessions are glutinous rice: Zinuo (indica), Huangpinuo (japonica) and Nuogu (japonica). Japonica rice is cultivated on limited surfaces in the Yuanyang terraces, ca. 5% of total rice acreage. The eleven modern varieties were selected in a collection[82] representative of the world-wide rice phenotypic and genetic diversity (temperate japonica: four varieties, tropical japonica: three varieties, indica: four varieties, aus: one variety).

METHOD DETAILS

Design of a rice RenSeq bait-library and application to explore NLR diversity

We designed a bait library capable of hybridizing to a wide variety of Asian rice NLRs. We characterized the NLR complements in the genomes of the japonica rice reference variety Nipponbare (MSU Rice Genome Annotation Project Release 7[83]) and the indica rice reference variety 93-11[84] by three different approaches: (1) searching NBARC domain-coding sequences (containing Pfam|PF00931 motif) in the CDS of both genomes with PFAMSCAN[56]; (2) identifying NLRs among CDS of both genomes with NLR PARSER v1[57] using default parameters followed by removing those with an NBARC domain coding sequence shorter than 500 or longer than 1100 nucleotides; (3) recovering the NLR repertoires identified by Luo et al.[85] in the Nipponbare and 93-11 genomes and filtering them for presence the NBARC domain (Pfam|PF00931). Redundancy within the japonica and the indica NLR gene sets was removed by positional information of the corresponding genes. In addition, to further remove redundancy in the NLR repertoire, NLRs whose NBARC -coding sequences were more than 95% identical between japonica and indica NLRs or among indica NLRs were removed by keeping the homolog with the longest NBARC domain. From the resulting set of 761 NLR sequences, 21,000 baits of 120 nucleotides and with 20 bp overlap were designed using a proprietary script from Arbor Bioscience (https://arborbiosci.com/). These oligos were aligned to the Nipponbare and 93-11 genomes with BLAST-N and oligonucleotides with more than 10 perfect matches per genome were excluded.

Genomic DNA was extracted from two weeks-old rice seedling using a CTAB method.[86] Enrichment and library preparation were carried out as described in Witek et al.[20,21] Fourty-nine post-enrichment samples were sequenced using Illumina HiSeq 2500. We mapped RenSeq reads against a reference set of NLR sequences identified in Ensembl Plants Genes database (*O. sativa* indica ASM465v1 version 43, *O. sativa* japonica ASM465v1 version 45) using the BioMart utility to filter gene IDs with Interpro entry IPR002182 (NBARC domain). To avoid redundancy among sequences caused by orthology between *O. sativa* indica and *O. sativa* japonica, and because 36/38 of the landraces included in our dataset are of indica type, we determined orthology relationships between *O. sativa* indica and *O. sativa* japonica sequences, and retained a final set combining all *O. sativa* indica sequences with *O. sativa* japonica sequences having no ortholog in *O. sativa* indica. BLAST-N analysis revealed that 593 of the 596 reference sequences had a minimum identity of 69% with sequences used as baits.

Raw Illumina paired-end reads from 49 rice accessions were aligned to the FASTA file of 596 NLR gene sequences using BOWTIE v2.3.5[58] with the option –very-sensitive and the rest all defaults to produce 49 BAM files that were then sorted using Samtools (v1.9) tool.[59,60]

Illumina sequencing of captured RenSeq sequences provided 3,702,690 pairs of paired-end reads per accession on average (standard deviation: 4,462,281; min: 937,145; max: 27,528,626). Variability in the number of reads, in particular between landraces and domesticates, had an impact on the proportion of sites covered and the coverage (data not shown). In order to reduce the impact of the heterogeneity in the number of reads in our analysis of presence/absence and nucleotide diversity, we standardized our dataset to the same level of average sequencing depth, by randomly subsampling 937,145 pairs of reads for landraces and 2,342,863 pairs of reads for domesticates (937,145 is the number of pairs of reads of the less deeply sequenced accession V11; 2,342,863 is 937,145*2.5). This procedure reduced the coupling between the number of reads and sequencing depth statistics, as observed by computing the standard deviation of sequencing depth across NLRs, which decreased from 219.3 to 10.8.

NLRs present copy number variation, so a substantial fraction of heterozygous calls is expected to result from hidden paralogy. Alleles at the same NLR locus can also vary in their affinity to sequencing baits, which can also influence the detection of heterozygosity. To identify and remove erroneous calls caused by hidden paralogy while controlling for allele imbalance, we used a SNP caller that explicitly models these two features. SNP calling was carried out using READS2SNP 2.0[61–63] using 2592 combinations of the following parameters: min (minimal number of reads required to call a genotype; values: 10 or 30), th1 (minimal posterior probability required to call a genotype; values: 0.95, 0.99 or 0.999), par (filtering for SNPs caused by hidden paralogy; values: 0 or 1), th2 (maximal p-value required to reject a paralogous SNP; values: 0.001, 0.01 or 0.05), aeb (accounting for allelic expression bias; values: True or False), fis (inbreeding coefficient; values: 0.8, 0.9, 0.95 or 0.99), bqt (filtering out positions of quality below threshold; values: 20, 30 or 40), rqt (filtering out reads of mapping quality below threshold; values: 20, 30 or 40). To select the best combination of SNP calling parameters, we computed the number of segregating sites (S) and inbreeding coefficient (Fis) at NLRs for the group of four indica varieties in each of the 2592 SNP sets and compared with the values obtained for six random samples of four indica accessions from the three thousand rice genome dataset, referred to as 3KRG indica reference datasets.[87] We computed the Euclidian distance between (Fis,S) estimated for the indica group in the 2592 SNP sets and for the three thousand rice genomes (3KRG) dataset, and by minimizing this distance, we selected the "best" (or most comparable) SNP dataset with the lowest deviation from the 3KRG dataset. SNPs occurring in NLRs in Wang et al.[87] were identified by mapping protein sequences of our reference set of NLR against their older version of the 93-11 genomic sequence using EXONERATE.[64] Four hundred eighty-eight of our reference set of 519 indica NLRs could be identified. Summary statistics S and Fis were computed using the Python package SCIKIT-ALLEL v. 1.3.3.[65] On average, the inbreeding coefficient in the 3KRG indica reference datasets was Fis=0.62 (standard deviation: 0.02) and the number of segregating sites was S=14707 (standard deviation: 1059). The closest of the 2592 READS2SNP datasets was dataset obtained with the following parameters: min=10, th1=0.95, par=1, th2=0.001, aeb=False, Fis=0.99, bqt=40, rqt=20; summary statistics estimated for this dataset were Fis=0.63 and S=15016.

Rice is a selfing species and accessions were subjected to single seed descent before sequencing, so the number of heterozygous calls per SNP was expected to be much lower than postulated by Hardy-Weinberg equilibrium. Although the -par option in

READS2SNP removed most SNP calls caused by hidden paralogy, SNPs with excess heterozygosity within populations remained in the dataset. In particular, plotting observed heterozygosity (Hobs) against the minor allele frequency (p) revealed SNPs with no or very few homozygous alternate calls, distributed along the Hobs=2p line, likely caused by gene duplications present in certain accessions (Figure S6). SNPs with excess heterozygosity were removed using the same criterion as in Wang et al.,[87] by filtering out, in each landrace and rice subspecies, sites where observed heterozygosity was more than ten times the most likely value for a given frequency and inbreeding rate (Figure S6). After filtering, summary statistics computed across the six 3KRG indica reference datasets (average [standard deviation]: Fis=1 [0], S=10200 [898]) remained very close to those computed on READS2SNP dataset 286 (Fis=1, S=10051). Reference NLR sequence used for SNP calling represented 2,423,478 bp. After masking 56,894 paralogous calls and SNPs with excess heterozygosity, the selected SNP set included 41,422 SNPs, of which 40,530 were biallelic.

A subset of 15 post-enrichment libraries was sequenced using PacBio RSII, including five indica landrace accessions, five indica varieties, and five japonica varieties (Table S1). Sequencing reads were assembled using CANU v2.0[66] with options genomesize=430m minOverlapLength=500 minOverlapLength=100 useGrid=true rawErrorRate=0.6 correctedErrorRate=0.105. The number of contigs per accession was 822 on average (min: 566 in C101A51; max: 1196 in Zhenshan2). Genes were predicted using AUGUSTUS 3.5[67] with option –SPECIES=rice, and functional annotation was performed using INTERPROSCAN 5.60-92.0.[68] The number of predicted genes per accession was 729 on average (min: 485 in C101A51; max: 1152 in NSF-TV116). Orthology relationships between predicted genic sequences and the 596 reference NLR sequences used for SNP calling were identified with similarity searches with BLAST-N using the following criteria: nucleotide identity greater than 90%, 2000 bp overlap, and at least 30% of the matching reference NLR sequence covered. The number of predicted genes that could be assigned to a reference NLR was 445 on average (min: 305 in C101A51; max: 646 in Zhenshan2), and 510 of the 596 reference NLRs had an orthologous sequence in at least one accession. For each of the 510 groups of NLR orthologs, coding sequences were aligned using TRANSLATORX,[72] with MAFFT[69–71] as the aligner and default settings, to keep sequences in-frame.

Empirical distribution of genome-wide polymorphism

To generate a baseline against which to identify features of polymorphism in genes of interest, we used previously published data for 68 accessions previously characterized[88–90] using genotyping-by-sequencing (GBS), representing nine landraces and 28 modern varieties (Table S1). GBS reads were mapped using BOWTIE v2.3.5[58] (option –very-sensitive) against genic sequences predicted in the reference *O. sativa* indica genome 93-11 (Ensembl Genomes 45). SNP calling was carried out using READS2SNP 2.0[61–63] with the same set of parameters as selected for RenSeq, but relaxing constraints on sequencing depth and mapping quality: min=3, th1=0.95, par=1, th2=0.001, aeb=False, Fis=0.99, bqt=10, rqt=10. Of the 99,576,191 bp of genic reference sequence, 30,918,075 bp were covered by at least three GBS reads passing quality filters. After masking 121,672 paralogous calls and SNPs with excess heterozygosity, the selected SNP set included 200,098 SNPs, of which 199,128 were biallelic. The number of non-NLR genes characterized was 25,102 on average, and ranged from 8,968 (temperate japonica) to 30,875 (Hongjiao).

QUANTIFICATION AND STATISTICAL ANALYSIS

Presence/absence variation

The depth in uniquely mapped reads at each position of a NLR gene in each rice cultivars was obtained from the sorted BAM files in SAMTOOLS. This depth values at all positions in a NLR gene is used to calculate a mean depth across the NLR gene. This gave a mean depth for each NLR gene in each rice cultivar. Each NLR mean depth was then normalized by taking the overall mean from all NLR gene mean depths in each rice cultivar, dividing each NLR gene mean depth by the overall mean depth. Jack-knife estimates of the coefficient of variation were obtained using the ASTROPY.STATS package in PYTHON. For each individual landrace, and each group of modern varieties, we found no significant relationship between mean depth per NLR and the proportion of paralogous calls (data not shown), confirming that there is less of a direct link between depth and copy number in targeted capture data, as the number of reads also depends on the affinity between reads and baits.

Population subdivision

Clustering and phylogenetic network analyses were performed on biallelic SNPs. Clustering in K ancestral populations was performed using sNMF.[23] The K value ranged from 2 to 15 and each sNMF run was repeated 10 times. CLUMPAK[73] was used to process sNMF output. Neighbor-net networks were built using SPLITSTREE 5.[24]

Polymorphism and divergence

Summary statistics of variation were computed using Python package EGGLIB 3[74] (https://www.egglib.org/) after generating pseudo-alignments using the table of SNPs (in VCF format) and reference sequences. To limit the impact of differences in sample sizes between modern varieties and landraces, summary statistics were estimated in landraces using 30 independent subsamples of 7 accessions per population. Fisher exact, Kruskal-Wallis and Mann-Whitney tests were computed with SCIKIT_POSTHOCS 0.6.6 and SCIPY 1.8.0 in Python 3.6. To estimate summary statistics at functional domains, the coordinates of functional domains were obtained using InterPro, as implemented in Ensembl's Biomart.

The synonymous substitution rate dS was estimated in CODEML[75] (runmode = -2, CodonFreq = 2), in pairwise comparisons of protein-coding sequences (*i.e.*, without using a phylogeny). For each gene, we aligned an *O. sativa* indica sequence and an *O. barthii* sequence (Ensembl Genomes 45) using TRANSLATORX,[72] and ran CODEML on the in-frame alignment.

Assessing the impact of sampling effort on measures of molecular variability

To assess the capacity of RenSeq to measure genetic diversity at NLRs, we compared read mapping statistics and measures of sequence variability estimated from RenSeq with estimates obtained from GBS, using a rarefaction approach to overcome potential biases related to differences in sample size. Average nucleotide diversity reached 90% of its maximum value with 23 randomly selected accessions (data not shown), indicating that the majority of nucleotide diversity at NLRs has been uncovered with RenSeq data. Haplotype richness, in contrast, reached 90% of its maximum value with 39 accessions (data not shown), suggesting that the molecular diversification of NLRs occurs not only by mutation but also by recombination and gene conversion.[91] Rarefaction analysis of GBS data revealed that our dataset is sufficient to reliably characterize genomewide levels of polymorphism (data not shown).

Demographic modeling

To determine if patterns of variation at NLRs departed from neutrality, we performed coalescent simulations to correct for deviation from demographic equilibrium (i.e., constant population size). We used an approximate Bayesian computation (ABC) framework[25] with random forest methodology, as implemented in Python package PYABCRANGER 0.0.70[76,77] to identify the historical demographic model accounting for most features of the data at GBS loci without invoking selection. The most supported model served as a null hypothesis to test for neutrality at NLRs. ABC relies on the comparison between summary statistics calculated from observed data and the same statistics calculated from coalescent simulations under different demographic models. We used the number of segregating sites S, Tajima's D, nucleotide diversity π, the variance in Tajima's D, and the variance in nucleotide diversity, computed for each population, as the summary statistics for the three datasets analyzed (i.e., modern indica, modern japonica, and indica landraces). For the indica landraces dataset, summary statistics also included nucleotide divergence (dxy) between populations, and the dataset-wise estimates of Tajima's D and nucleotide diversity π. Models fitted to individual populations were the following: (i) a constant size model determined by a single parameter N1, the effective population size; (ii) an instantaneous bottleneck model,[92] parameterized by the initial effective population size N1, the start of the bottleneck T and the strength of the bottleneck ST, which corresponds to the time period during which coalescence events are collapsed, and the final effective population size N2; (iii) a two-epochs exponential growth model parameterized by the initial effective population size N1, the final effective population size N2, the start of population growth T, and the growth rate G, defined such as at time t in the past, the population size is N2*exp(-Gt); (iv) a two epochs population model parameterized by the initial effective population size N1, the final effective population size N2, the time of population change T; (v) a population subdivision model parametrized by the initial effective population size N1, the time of population splitting T, and the respective effective population size of the two derived populations N2 and N3. For the indica landrace dataset, N was the size of the population ancestral to all landraces, Tz was the splitting time between Zinuo and other landraces, T_bhh2_ax characterized the time of trifurcation between Acuce, Xiaogu, and the combination of Baijiao, Hongjiao and Hongyang2, T_bh2_h was the splitting time between Hongjiao and the combination of Baijiao and Hongyang2, and T_b_h2 was the splitting time between Baijiao and Hongyang2. Migration was modeled as the rate at which lineages are exchanged between populations (i.e., a symmetric migration rate), with one rate Mij per pair of populations (i,j). The boundaries of prior distributions are reported in Table S5. Prior boundaries were set following preliminary simulations to ensure that for each proposed model prior combinations were able to produce some simulated datasets in the vicinity of the observed dataset.[93] The appropriateness of prior combinations was assessed using principal component analyses, by projecting observed summary statistics onto simulated summary statistics. To limit the number of models to compare, we used a population tree inferred with TREEMIX[78] to determine the splitting order of the different landrace clusters, and we first carried out analyses on each cluster independently to determine the best-fitting demographic model. Simulations were performed using MSPRIME[79–81] assuming a recombination rate of 1e-8/generation/bp, and under a Lambda coalescent with multiple mergers (BETACOALESCENT function) or under a classical coalescent (STANDARDCOALESCENT function). Model selection and parameter estimation[94] were performed using the PYABCRANGER package in Python. We simulated 500,000 multilocus datasets for models including all clusters of indica landraces, and 50,000 multilocus datasets for models fitted to modern indica or japonica varieties. Posterior probabilities of demographic scenarios and posterior probabilities for parameters under the best-supported model were estimated using 1000 trees (NTREE option in PYABCRANGER). Posterior predictive checks[94] indicated that models with the highest posterior probabilities provided a good fit to the data for all groups and populations (Table S5). For the best-supported model of each dataset, posterior predictive checks were carried out using Gaussian kernel density estimation with the GAUSSIAN_KDE function in SCIPY, and sampling random values from the fitted density distribution. Ten thousand datasets of the same sample size and sequence length as GBS sequences were simulated per population/group in MSPRIME using the sampled multivariate parameters. For all populations/groups, the best-supported models were able to reproduce the observed values of S, π, and Tajima's D, confirming their goodness-of-fit (data not shown).

Neutrality at NLR genes was tested by simulating null distributions using the most supported demographic models inferred from GBS data. To generate null distributions, 10000 datasets of the same sample size and sequence length as NLR sequences were simulated in MSPRIME by sampling multivariate parameters from posterior distributions using the same procedure as for posterior predictive checks.

Directional selection

The SNIPRE framework uses a generalized linear mixed model to estimate the influence of mutation rate, species divergence time, constraint, and selection effects on polymorphism and divergence. Genome-wide effects are incorporated into the analysis as fixed effects, while individual gene effects are incorporated as random effects, which allows to combine information across genes and increases power to detect the effects of selection on a gene-by-gene basis. We focused our analyses on the selection effects, which reflect the selection coefficients (γ), and the constraint (or non-synonymous) effects, which reflect mutational constraint (1-f, f being the proportion of non-synonymous mutations that are not lethal).