

# Genomic evidence of human selection on Vavilovian mimicry

Chu-Yu Ye<sup>1,6</sup>, Wei Tang<sup>2,6</sup>, Dongya Wu<sup>1,6</sup>, Lei Jia<sup>1</sup>, Jie Qiu<sup>1</sup>, Meihong Chen<sup>1</sup>, Lingfeng Mao<sup>1</sup>, Feng Lin<sup>1</sup>, Haiming Xu<sup>1</sup>, Xiaoyue Yu<sup>2</sup>, Yongliang Lu<sup>2</sup>, Yonghong Wang<sup>3</sup>, Kenneth M. Olsen<sup>4</sup>, Michael P. Timko<sup>5</sup> and Longjiang Fan<sup>1\*</sup>

**Vavilovian mimicry is an evolutionary process by which weeds evolve to resemble domesticated crop plants and is thought to be the result of unintentional selection by humans. Unravelling its molecular mechanisms will extend our knowledge of mimicry and contribute to our understanding of the origin and evolution of agricultural weeds, an important component of crop biology. To this end, we compared mimetic and non-mimetic populations of *Echinochloa crus-galli* from the Yangtze River basin phenotypically and by genome resequencing, and we show that this weed in rice paddies has evolved a small tiller angle, allowing it to phenocopy cultivated rice at the seedling stage. We demonstrate that mimetic lines evolved from the non-mimetic population as recently as 1,000 yr ago and were subject to a genetic bottleneck, and that genomic regions containing 87 putative plant architecture-related genes (including *LAZY1*, a key gene controlling plant tiller angle) were under selection during the mimicry process. Our data provide genome-level evidence for the action of human selection on Vavilovian mimicry.**

Named after the noted Russian botanist and geneticist Nikolai Vavilov, Vavilovian mimicry (or crop mimicry) describes an adaptive phenomenon in which weeds evolve to resemble domesticated crops at specific stages in their life history to avoid eradication<sup>1–3</sup>. In early agriculture, the physical removal of weeds by humans required the ability to discriminate between desirable crops and harmful weeds. Vavilovian mimicry is a phenotypic adaptation that makes it difficult for farmers to distinguish between crop plants and weeds, resulting in the ability of the weeds to evade eradication<sup>1,4</sup>. Similar to most other mimicry systems, Vavilovian mimicry involves three players: (1) the model, the crop that is imitated; (2) the mimic, the weed that imitates the model and (3) the dupe, the farmer that identifies and removes the weeds.

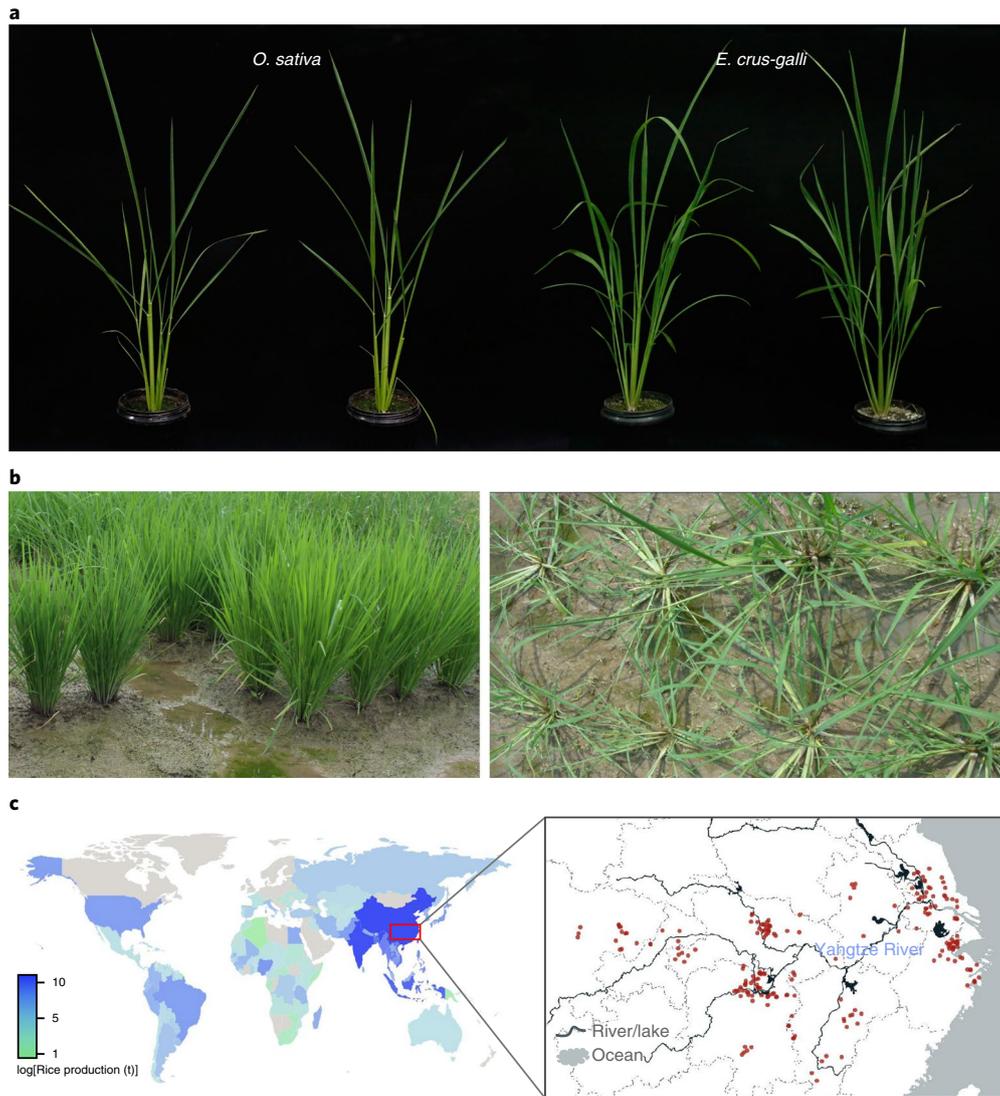
The most widely recognized illustrations of mimicry are those seen in the animal kingdom, including Batesian mimicry and Müllerian mimicry. The molecular mechanisms underlying these two types of mimicry have been investigated for wing colour patterns in *Papilio* and *Heliconius* butterflies, respectively<sup>5–15</sup>. Cases of mimicry in plants are less well documented<sup>16–18</sup>. Floral mimicry is common in orchids<sup>19</sup>, and leaf mimicry has been reported in several plants<sup>16,20,21</sup>. Vavilovian mimicry is a unique form of plant mimicry that occurs in crop ecosystems and results from unintentional selection by humans. A well-known example of Vavilovian mimicry is the evolution of *Echinochloa* weeds in paddy fields to resemble cultivated rice<sup>1,4</sup>. The genus *Echinochloa* (Poaceae) belongs to the subfamily Panicoideae and includes numerous problematic weeds. Among these is *E. crus-galli*, which is dominant in paddy fields with highly diverse morphological phenotypes that are both mimetic and non-mimetic of rice<sup>4</sup>. Barrett<sup>1</sup> compared 15 morphological and growth characters of mimetic and non-mimetic *Echinochloa* weeds and cultivated rice. His results showed that cultivated rice and a

tetraploid rice mimic form of *Echinochloa* (*E. crus-galli* var. *oryzicola*, also called *E. oryzicola* or *E. phyllopopogon*) clustered together, whereas the non-mimetic hexaploid *E. crus-galli* var. *crus-galli* was separated by a considerable phenotypic distance. The mimetic *Echinochloa* had upright tillers and leaves, whereas non-mimetic *E. crus-galli* possessed lax, drooping leaves and frequently exhibited a decumbent growth form. The Barrett lab also compared life-history traits in several *Echinochloa* species/varieties, including time to flower, dry weight allocation, reproductive effort and seed traits<sup>22,23</sup>. Among their findings was the observation that seeds of *E. crus-galli* var. *oryzicola* are two to three times heavier than those of *E. crus-galli* var. *crus-galli*. Comparisons of isozyme variation revealed that *E. crus-galli* var. *crus-galli* maintained significantly more intrapopulation genetic variation than the two rice weed specialists (*E. oryzoides* (hexaploid) and *E. phyllopopogon*)<sup>24,25</sup>.

Seed mimicry is also found in weeds—a classic example is *Camelina sativa*, an annual weed of flax, whose seeds display winnowing properties that resemble those of flax seed and make it difficult to separate the two easily<sup>1</sup>. Crop mimicry is also found in some species including rye (*Secale* spp.), oat (*Avena* spp.) and darnel (*Lolium temulentum*). For example, the morphology of awned darnel grains resembles that of emmer wheat grains, while awnless darnel grains resemble free-threshing wheat grains<sup>26,27</sup>. These mimetic weeds, which share many adaptive features with domesticates, have also been termed as parasitic domesticoids by Fuller and Stevens<sup>28</sup>.

Mimicry is an interesting phenomenon in nature, and understanding its mechanisms is important in the field of evolutionary biology. Vavilovian mimicry occurs in agroecosystems involving human action; therefore, unravelling its molecular mechanisms will contribute to our understanding of mimicry adaptation and the origin and evolution of agricultural weeds. Although the phenomenon

<sup>1</sup>Institute of Crop Sciences and Institute of Bioinformatics, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China. <sup>2</sup>State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, China. <sup>3</sup>State Key Laboratory of Plant Genomics and National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. <sup>4</sup>Department of Biology, Washington University in St. Louis, St. Louis, MO, USA. <sup>5</sup>Department of Biology, University of Virginia, Charlottesville, VA, USA. <sup>6</sup>These authors contributed equally: Chu-Yu Ye, Wei Tang, Dongya Wu. \*e-mail: fanlj@zju.edu.cn



**Fig. 1 | Morphology associated with Vavilovian mimicry in *E. crus-galli* and sampling.** **a**, Cultivated rice (left) and mimetic *E. crus-galli* (right) planted in a greenhouse. **b**, Typical mimetic (left) and non-mimetic (right) *E. crus-galli* in paddy fields. **c**, Sampling in this study. Left, sampling location (marked by a box) of *E. crus-galli* in this study. The colour scale indicates global rice production. Right, geographic distribution of the phenotyped and genotyped *E. crus-galli* lines used in this study (red dots;  $n = 328$ ).

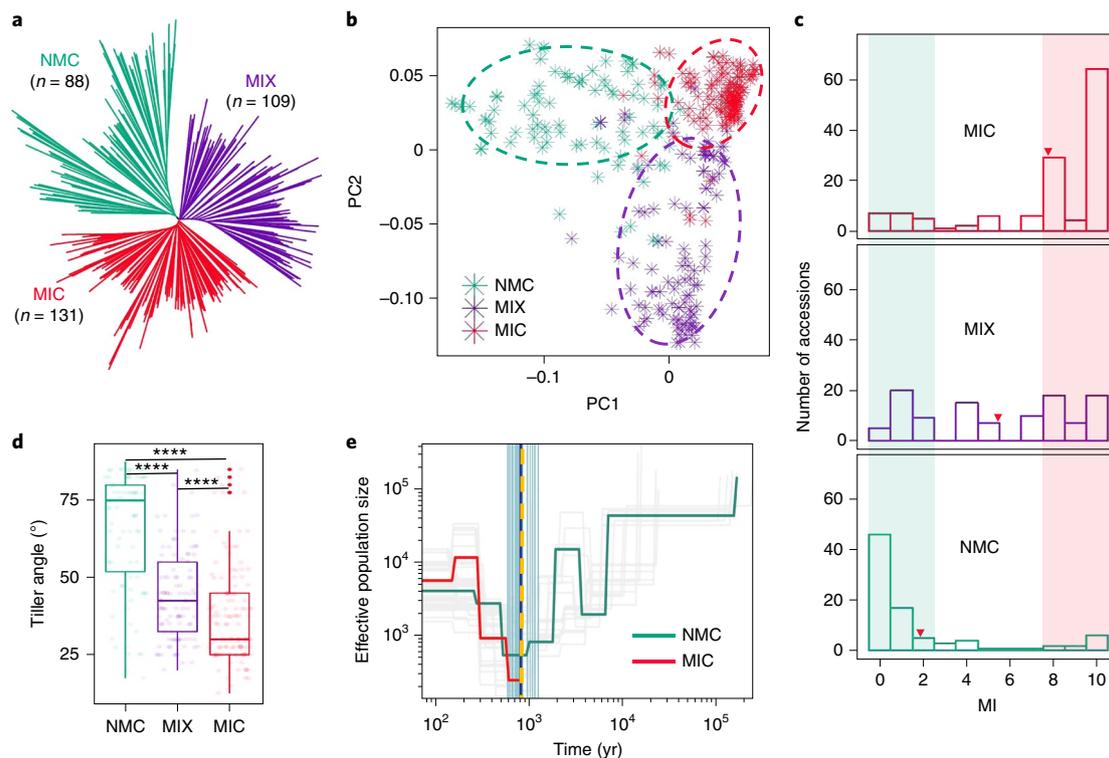
of rice mimicry in *E. crus-galli* has been documented, the molecular mechanisms and genomic changes underlying this example of Vavilovian mimicry remain largely unknown. The *E. crus-galli* genome sequence was previously reported by us<sup>29</sup> and provides an opportunity to reveal the genomic mechanisms underlying this unique mimicry system. Rice mimicry in *E. crus-galli* is thought to be caused by continuous hand weeding, and thus we assume that selection driven by human force would be detected in the *E. crus-galli* genome accompanying the mimicry process. To test this hypothesis, we resequenced the genomes of 328 *E. crus-galli* accessions from the Yangtze River basin in China, a typical rice production area. Our analyses uncovered genomic footprints associated with Vavilovian mimicry under human selection in *E. crus-galli*.

## Results

**Rice mimicry in *E. crus-galli* at the seedling stage.** In the paddy field, morphological similarity between cultivated rice and *E. crus-galli* at the seedling stage makes it hard for farmers to distinguish between the two and, therefore, to remove the weed from the paddy. The basis for the seedling mimicry involves multiple traits. Previous

findings<sup>1</sup> and our own observations indicate that the analogous phenotypes between mimetic *E. crus-galli* and cultivated rice at the seedling stage include a small tiller angle, a straight stem node, a green stem base and compact leaves (a small leaf angle). In contrast, non-mimetic *E. crus-galli* individuals show a loose or prostrate plant architecture, usually accompanied by a crooked stem node, red or purple stem bases and loose leaves (a large leaf angle) (Fig. 1a,b and Supplementary Fig. 1). We therefore used these four representative morphological traits to define Vavilovian mimicry at the seedling stage. We calculated a weighted score for each trait to assess the mimicry level, termed the mimicry index (MI) (for details, see Methods and Supplementary Fig. 2). The MI ranged from 10 to 0, indicating that the mimicry level changed from mimicry to non-mimicry. Empirically, *E. crus-galli* accessions with an MI of  $\geq 8$  were defined as typical mimetic types, while those with an MI of  $\leq 2$  were regarded as non-mimetic types.

**Sampling, phenotyping and sequencing of *E. crus-galli*.** We collected 328 *E. crus-galli* accessions from within paddy fields and adjacent regions in diverse locations in the Yangtze River basin in



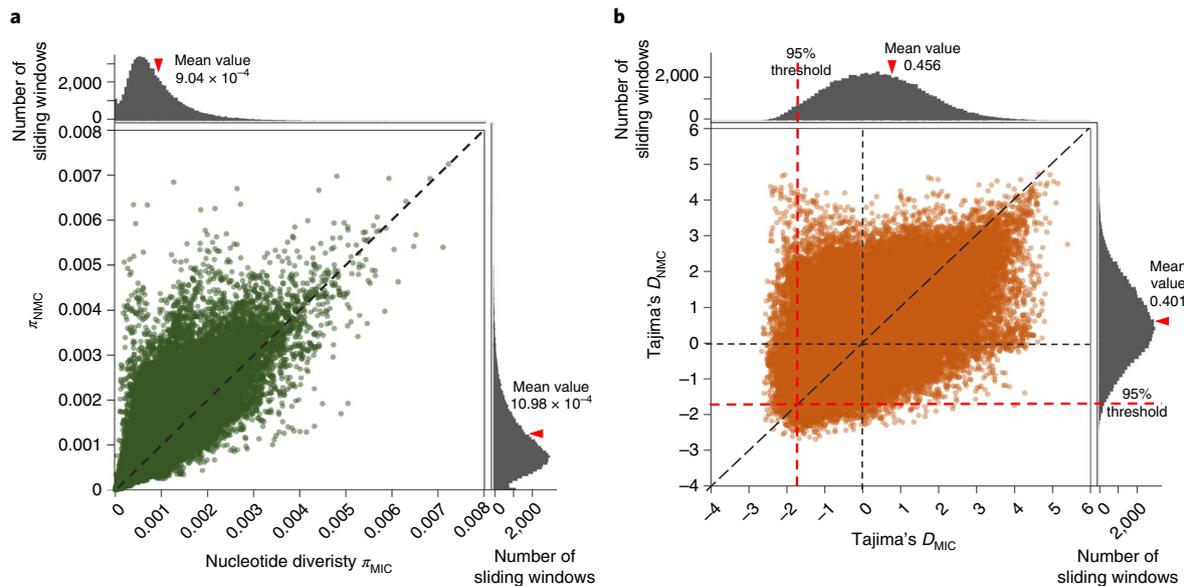
**Fig. 2 | Phylogenetic and phenotypic differentiation of *E. crus-galli* in the Yangtze River basin.** **a**, An approximate maximum-likelihood tree of 328 *E. crus-galli* accessions. **b**, PCA plot of the first two principal components based on whole-genome SNPs. PC1 and PC2 account for 18.9% and 15.3% of the genetic variations, respectively. Three groups were separated indicated by the red (MIC), green (NMC) and purple (MIX) dashed outlines. **c**, Mimicry phenotypic differentiation among the three groups quantified with the MI. The average MIs of the three groups are marked by red triangles. The red and green shading indicate typically mimetic and non-mimetic MI values, respectively. **d**, The tiller angle differentiation among the three groups. In the box plots, the horizontal line shows the median value, and the whiskers show the 25% and 75% quartile values of angles. The red dots indicate outlier values. \*\*\*\* $P < 0.0001$ . **e**, The divergence time between mimetic and non-mimetic *E. crus-galli* determined with SMC++. The dashed vertical line indicates the time of divergence (nearly 1,000 yr ago).

China (Fig. 1c), where rice is thought to have been domesticated and where it has been cultivated for a long time<sup>30</sup>. All accessions were planted together in paddy fields and phenotyped at the seedling stage over two seasons (2017 and 2018; Supplementary Table 1). Consistent phenotypes were observed, and 150 (45.7%) and 121 (36.9%) accessions were classified as typically mimetic and non-mimetic types, respectively, according to the MI (Supplementary Fig. 3). Significant phenotypic differences between mimetic and non-mimetic materials planted in greenhouses under two different photoperiods were consistent with those observed in the paddy fields (Fig. 1b and Supplementary Fig. 4).

With whole-genome sequencing of the 328 phenotyped *E. crus-galli* accessions at a 15× average depth of coverage, we generated a total of 7.17 Tb of base pairs (Supplementary Table 1). Cleaned paired-end reads were mapped against the *E. crus-galli* reference genome (STB08)<sup>29</sup>, and approximately 96.27% of the reads were found to cover >92% of the STB08 reference genome for each accession (Supplementary Table 1). This result, together with the genome size estimation (Supplementary Table 1), indicates that all accessions in this study belong to the species *E. crus-galli*. Finally, we called a set of 9.03 million genome-wide, high-quality single nucleotide polymorphisms (SNPs) (with a minor allele frequency (MAF) greater than 0.01 and integrity rate greater than 0.8) and 2.40 million small insertions/deletions (indels), with an average of 9.07 variations per kilobase. A total of 13,101 SNPs were predicted to have important impacts on function, resulting in start codon losses, transcript elongation or premature stop codon gains (Supplementary Table 2).

**Genomic differentiation between mimetic and non-mimetic *E. crus-galli*.** A phylogenetic tree was constructed based on the genome-wide SNPs, which divided the 328 accessions into three major groups (Fig. 2a). A principal component analysis (PCA) plot showing the first two eigenvectors separated all accessions into three groups, consistent with the results of the phylogenetic analysis (Fig. 2b). Two of the groups could be easily defined as mimetic (MIC) and non-mimetic (NMC) populations, since most of the accessions in the respective groups were mimetic or non-mimetic (Fig. 2c and Supplementary Fig. 5). The third group was defined as a mixed group (MIX) and exhibited a symmetrical distribution of the MI and an average MI of 5.4 (Fig. 2c). MIC ( $n = 131$ ) exhibited an average MI of 7.7, and 97 (74.1%) accessions showed typical mimicry of rice (that is, MI of  $\geq 8$ ). In NMC ( $n = 88$ ), the average MI was 1.9, and there were 68 (77.3%) accessions with an MI of  $\leq 2$  (Fig. 2c). Tiller angles were also significantly differentiated, with averages of 27°, 42° and 75° in MIC, MIX and NMC, respectively (Fig. 2d). Our further analyses focused on the accessions with extreme MI values; that is, mimetic accessions from MIC ( $n = 97$ ) and non-mimetic accessions from NMC ( $n = 68$ ).

Logically, MIC should have emerged and evolved from NMC following their appearance in paddy fields. To test this hypothesis, we constructed a phylogenetic tree using the tetraploid *E. oryzicola*, which is the paternal donor of the hexaploid *E. crus-galli*<sup>31</sup>, as an outgroup. NMC was located at the basal position on the tree and close to *E. oryzicola* accession, and MIC was derived from NMC, supporting our hypothesis that NMC was ancestral to MIC (Supplementary Fig. 6). We also found that 92.5% of the variations



**Fig. 3 | Genomic signatures of positive selection during the mimicry process in *E. crus-galli*.** **a**, The distribution of nucleotide diversity in 50 kb sliding windows with a 20 kb step, showing higher nucleotide diversity in NMC than in MIC. **b**, The distribution of Tajima's  $D$  in 20 kb sliding windows across the whole genome, showing genomic regions in MIC with significant negative Tajima's  $D$  values. Red dashed lines represent the 95% significance threshold for the corresponding population size. Average values are marked by red triangles.

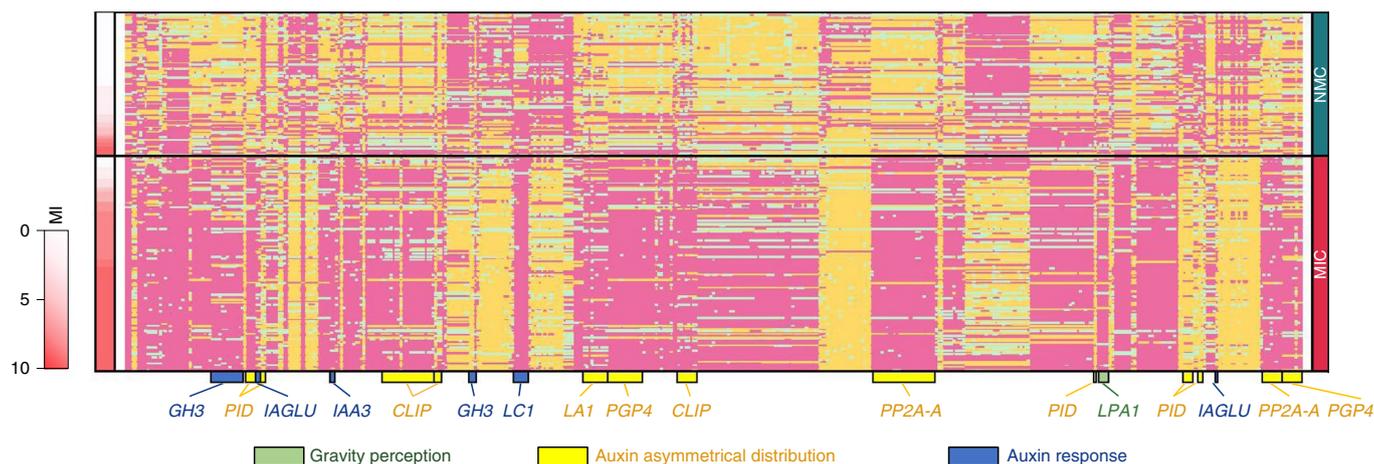
(SNPs and indels) present in MIC were a subset of those in NMC, while NMC contained more variations, of which 17.1% were absent in MIC. These results further supported the hypothesis that MIC evolved from NMC. We then estimated the divergence time of the two groups using SMC++<sup>32</sup>, and the results indicated that MIC differentiated from NMC approximately 1,000 yr ago (Fig. 2e), a period during the Song Dynasty of China (AD 960–1279). Pairwise sequentially Markovian coalescent analysis showed a consistent demographic history for MIC and NMC before 10,000 yr ago, which supports the recent differentiation between these two groups (Supplementary Fig. 7).

We employed the fixation index ( $F_{ST}$ ) to measure genomic differentiation between MIC and NMC. The results showed a relatively low value ( $F_{ST}=0.062$ ) at the whole-genome level, which was consistent with the estimate of a short divergence time between MIC and NMC. A total of 255 genomic regions longer than 100 kb were detected as having an  $F_{ST}$  value greater than 0.147 (top 5% threshold), including seven Mb-scale regions (Supplementary Table 3). The results suggest significant differentiation at some local genomic regions between MIC and NMC.

**Positive selection signals in mimicry.** We used genetic diversity and Tajima's  $D$  to test whether genomic regions were under selection signals during the mimicry process. A decrease in genetic diversity (in terms of nucleotide diversity  $\pi$ ) at the whole-genome level was observed in MIC relative to NMC, with a reduction of diversity ( $ROD=\pi_{NMC}/\pi_{MIC}$ ) value of 1.324, suggesting that MIC suffered a genetic bottleneck during the mimicry process (Fig. 3a). An analysis of genetic diversity using fourfold synonymous third-codon transversion sites, synonymous mutation sites and intragenic variations showed that the level of reduction in whole-genome genetic diversity was higher than that at fourfold synonymous third-codon transversion neutral ( $ROD=1.113$ ,  $P<0.0001$ ) and synonymous sites ( $ROD=1.150$ ,  $P<0.0001$ ) but significantly lower than that in intragenic regions ( $ROD=1.650$ ,  $P<0.0001$ , analysis of variance and Tukey's honest significant difference test). These results suggest that positive selection played an important role in diversity reduction during mimicry evolution, despite the effects of genetic drift

and other factors (Supplementary Fig. 8). Additionally, 21.0% of medium-frequency ( $0.4<MAF\leq 0.5$ ) SNPs ( $n=130,213$ ) in NMC were almost fixed in MIC with a frequency of  $\leq 0.2$  or  $\geq 0.8$ , suggesting genomic selection during the mimicry process (Supplementary Fig. 9). We found 280 regions longer than 100 kb with a significant reduction in genetic diversity ( $ROD<2.504$ ; top 5% threshold) in MIC; that is, putative regions under selection (Supplementary Table 4). Tajima's  $D$  showed similar average values between MIC (Tajima's  $D=0.456$ ) and NMC (Tajima's  $D=0.401$ ; Fig. 3b). However, regions totalling 39.9 Mb in size, containing 127 segments longer than 100 kb, were found to have significant negative Tajima's  $D$  values in MIC (less than  $-1.774$ ; 95% confidence limit for MIC population size; Fig. 3b and Supplementary Table 5). In contrast, only 15.6 Mb of regions were found to have a Tajima's  $D$  value less than  $-1.784$  (95% confidence limit for NMC population size) in NMC, which was partly due to there being more rare alleles in NMC than in MIC. Taken together, these results indicate that some genomic regions experienced positive selection during the mimicry process in *E. crus-galli*. As examples, two large genomic regions in scaffold 10 and scaffold 16 (lengths of 2.21 Mb and 1.59 Mb, respectively) showed selection signals with decreased genetic diversity and a significantly lower Tajima's  $D$  in MIC than in NMC (Supplementary Fig. 10).

**Plant architecture-related genes were enriched as having been selected during mimicry.** On the basis of the results of the genomic differentiation, genetic diversity and neutral tests presented above, a total of 7,596 genes with 114,652 variations (93,063 SNPs and 21,589 indels) were found in putative selected regions. After we applied additional filtering based on significant differences in allele frequency ( $P<1\times 10^{-20}$ ) between MIC and NMC and the impact of variations (excluding synonymous mutations), 1,986 genes (1.83% of the whole-genome annotated genes) with 8,373 variations (7,167 SNPs and 1,206 indels) were identified as candidate loci under selection during mimicry. Among these genes, 87 were annotated as homologues of known plant architecture-related genes in rice and were thus considered a high-confidence gene set related to mimicry traits in *E. crus-galli* (Supplementary Table 6). A total of 455 variations, including 398 SNPs and 57 indels, within the 87 genes were



**Fig. 4 | Haplotype diversity of 455 variations within the 87 plant architecture-related genes in NMC and MIC.** Pink, yellow and green boxes represent the homozygous haplotype of the reference allele (STB08), the homozygous haplotype of the alternative allele and heterozygous sites, respectively. The MI for each individual is indicated by the bar on the left. Genes involved in plant gravitropism are marked at the bottom; colours indicate involved pathways of the genes.

found to be significantly differentiated between MIC and NMC (Fig. 4 and Supplementary Table 7). Furthermore, among all 1,986 candidate genes, the plant architecture-related genes were significantly enriched ( $P < 0.001$ , Fisher's exact test) compared with the whole-genome level in *E. crus-galli*.

We also observed that genes involved in plant gravitropism were enriched ( $P < 0.05$ , Fisher's exact test; Fig. 4 and Supplementary Fig. 11) in the high-confidence gene set for mimicry. Previous studies have established that gravitropism plays a critical role in controlling tiller angle, probably through asymmetric auxin distribution and auxin signal response, leading to differential cell elongation between the abaxial and adaxial sides of tillers<sup>33–39</sup>. At least one homologue involved in plant gravity perception, 14 homologues involved in asymmetric auxin distribution and 7 homologues involved in auxin signal response were found in the 87-gene set (Supplementary Fig. 11 and Supplementary Table 7). Moreover, we detected 22 homologues of five genes in the brassinolide signalling pathway, which has been suggested to play important roles in leaf angle regulation<sup>40–43</sup> (Supplementary Fig. 12 and Supplementary Table 7).

*LAZY1 (LA1)* is one of the most important genes identified thus far; it controls tiller angle and is involved in redistributing auxin in response to gravity stimulus, and its functions are conserved in rice, maize and *Arabidopsis*<sup>33,39,44–46</sup>. During the mimicry process in *E. crus-galli*, the genomic region of *LA1* (*EC\_v6.g079654*) was apparently differentiated between MIC and NMC ( $F_{ST} = 0.476$ ), and its genetic diversity decreased dramatically in MIC relative to NMC ( $ROD = 4.956$ ; Fig. 5a and Supplementary Table 7). Tajima's  $D$  showed that this region had a significant positive value in NMC but dramatically decreased to a negative value in MIC (Fig. 5b and Supplementary Table 1). High linkage disequilibrium also implied that the region was a potential target of selection during mimicry (Fig. 5b). Five homozygous haplotypes (hap01 to hap05) of *LA1* composed of two coding variations and eight intron variations showed dissimilar distributions in MIC and NMC (Fig. 5c,d and Supplementary Table 9). The haplotype hap01 accounted for 70% of the haplotypes in MIC but only 15% in NMC. Correspondingly, accessions with hap01 showed much smaller tiller angles (mean angle: 39.8°) than those with hap02 (mean angle: 51.1°,  $P < 0.0001$ ) and hap03 (mean angle: 61.5°,  $P < 0.0001$ , analysis of variance and Tukey's honest significant difference test; Fig. 5d). Specifically, 76% of the accessions in MIC shared the CT haplotype of two missense sites, while 57% in NMC shared the GC haplotype (Supplementary Fig. 13). Furthermore, the relative expression level of *LA1* in the

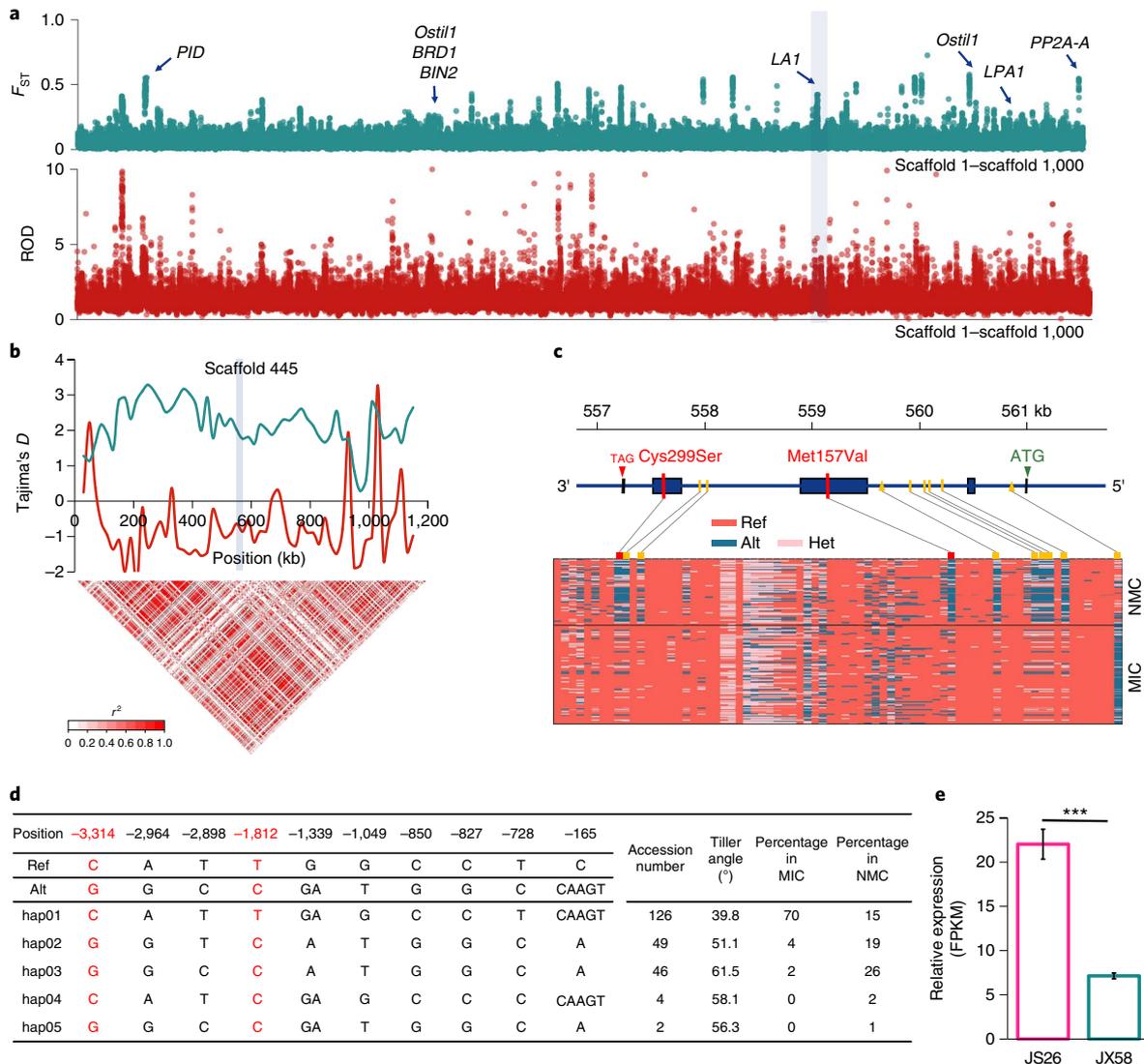
MIC accession (JS26) was much higher than in the NMC individual at the seedling stage (JX58; Fig. 5e). Taken together, these results suggest that *LA1* was under human-driven selection and differentiated during mimicry evolution in *E. crus-galli*. *LA1* was also under selection during rice domestication, showing a significant negative Tajima's  $D$  value and dramatically decreased genetic diversity in the cultivated (*Oryza sativa* ssp. *japonica*) population (Tajima's  $D_{\text{cultivar}} = -1.871$ ;  $ROD = 6.333$ ; Supplementary Fig. 14a). Haplotype analysis revealed that several alleles in *LA1* were highly differentiated between wild and cultivated rice, and rare alleles in the wild population were fixed in cultivated rice (Supplementary Fig. 14b). The results suggest that *LA1* has undergone parallel selection during mimicry evolution in *E. crus-galli* and during rice domestication.

## Discussion

In this study, we provide genomic evidence of human selection on Vavilovian mimicry. Using genome-level analysis, we uncovered significant genomic differentiation and positive selection signals in some local genomic regions between mimetic and non-mimetic *E. crus-galli* populations. These findings extend our core knowledge of mimicry and improve our understanding of the origin and evolution of paddy weeds.

Our analyses revealed genomic regions with selection signals in MIC that were supported by significantly decreased genetic diversity and significant negative Tajima's  $D$  values. In particular, 87 plant architecture-related homologous genes were shown to have been selected in MIC, and parallel selection was observed during mimicry evolution in *E. crus-galli* and during rice domestication, which suggests that the genetic mechanisms of rice mimicry in *E. crus-galli* are partly similar to those of rice domestication. Our findings are consistent with the hypothesis that Vavilovian mimicry is caused by unintentional selection by humans, although such selection is not the same as artificial selection in crops, which is a kind of intentional and directional selection.

Mimicry systems with relatively clear molecular mechanisms usually involve simple traits such as colour, whereas Vavilovian mimicry traits are more complex. It is likely that Vavilovian mimicry is controlled by multiple genes and that the molecular mechanisms underlying this type of mimicry are more complicated than those underlying other types of mimicry. We found hundreds of genomic regions and dozens of high-confidence mimicry-trait-related genes with selection signals in *E. crus-galli*.



**Fig. 5 | *LA1* was putatively under positive selection during mimicry evolution in *E. crus-galli*.** **a**,  $F_{ST}$  (top) and reduction in nucleotide diversity measured by ROD (bottom) in 50 kb sliding windows with a 20 kb step between MIC and NMC along the 1,000 largest scaffolds. Some of the known plant architecture-related genes are marked by arrows, and the genomic region (scaffold 445) harbouring *LA1* is highlighted by grey shading. **b**, Tajima's *D* distribution (top) and a heatmap of linkage disequilibrium measured by the squared Pearson's correlation coefficient ( $r^2$ ) (bottom) in scaffold 445. **c**, Gene structure, variations and haplotype diversity of *LA1*. Top: the gene structure of *LA1* in *E. crus-galli*. Blue boxes represent exons, and lines between boxes represent introns. Highly associated variations are marked by red and yellow vertical lines, corresponding to missense and modificatory effects, respectively. Bottom: haplotype diversity in NMC and MIC. ref, reference allele (STB08); alt, alternative allele; het, heterozygous allele. **d**, Haplotype analysis of *LA1* in 328 *E. crus-galli* accessions. Two non-synonymous SNPs are highlighted in red. Homozygous haplotypes are shown. **e**, Relative expression levels of *LA1* in MIC (JS26) and NMC (JX58) individuals are shown as means  $\pm$  s.d. FPKM, fragments per kilobase of transcript per million fragments mapped. \*\*\* $P < 0.001$ .

The time of divergence of mimetic from non-mimetic samples of *E. crus-galli* was estimated to be  $\sim 1,000$  yr ago in this study. According to Chinese historical records, during the Song Dynasty the economic centre changed from the Yellow River basin to the Yangtze River basin, the human population underwent rapid growth and rice replaced wheat as the staple grain<sup>47</sup>. Mimetic *E. crus-galli* may have originated from the stronger selection pressure by humans accompanied by the requirement for more rice production due to the rapid population growth during that period. Our phylogenetic analysis and PCA data show that the mimetic *E. crus-galli* are derived from non-mimetic forms and support a single origin of mimicry in *E. crus-galli* in the Yangtze River region (Fig. 2 and Supplementary Fig. 6). If multiple origins occurred in this area, then one would expect to observe multiple phylogenetic

groups with each containing non-mimetic *E. crus-galli* and their mimetic derivatives rather than two separated branches. This is not the case. For a few lines with outlier phenotypes in MIC and NMC, a reasonable explanation would be gene introgression. We cannot, however, exclude the possibility of multiple origins of mimicry in *E. crus-galli*. More materials from other regions are needed to resolve this issue. In addition, MIX includes many accessions with traits intermediate between mimetic and non-mimetic phenotypes. Hybridization between the two groups (MIC and NMC) may cause this phenomenon. Alternatively, these lines might be at an intermediate stage in the evolution of mimicry (semi-mimicry) or mimic older rice landraces that were not yet improved to an ideal (compact) plant structure. This work sets the stage for future studies aimed at providing broader insight into process of Vavilovian mimicry.

## Methods

**Plant material sampling and genome resequencing.** A total of 328 *E. crus-galli* accessions were collected from paddy fields, or nearby, in the Yangtze River basin (Supplementary Table 1). The seeds of each accession were germinated and planted in paddy fields at the China National Rice Research Institute, Fuyang, China, with eight individuals per accession. In 2017, we started to record their phenotypes including the tiller angle, stem node shape (straight or crooked), colour of the stem base (green or red/purple) and leaf type (compact or loose; Supplementary Fig. 1); the average value of three individuals was adopted for the tiller angle of each accession. The phenotypes were recorded again in 2018, and the results were consistent with those in 2017. All phenotypes were investigated at the tillering stage 3 weeks after the three-leaf-stage seedlings were transplanted into the paddy fields. The mimicry level was indicated by the MI, with a sum of values for the four traits of  $\geq 8$  and  $\leq 2$  indicating typical mimetic and non-mimetic types, respectively; tiller angles of  $< 30^\circ$ ,  $30\text{--}50^\circ$  and  $> 50^\circ$  were assigned values of 5, 3 and 0, respectively; straight and crooked node shapes were assigned values of 3 and 0, respectively; green and red/purple stem bases were assigned values of 1 and 0, respectively; and compact and loose leaf types were assigned values of 1 and 0, respectively. We also randomly selected four accessions from MIC (CQ1, JS26, HB17 and SC22) and NMC (HB10, ZJ102, JX67 and JX58) to plant in the greenhouse under two photoperiods (14h/10h and 16h/8h) for phenotyping. The phenotypes were consistent with those observed in the paddy fields (Supplementary Fig. 4). DNA was extracted from green leaves using a routine protocol. A total of more than 7.7 Tb of paired-end sequence data was generated by Illumina HiSeq 4000, with an average depth of coverage of approximately 15 $\times$  for each sample. The information about world rice production in Fig. 1b was downloaded from the Food and Agriculture Organization (<http://www.fao.org/faostat/>) and illustrated with the R package recharts (<https://madlogos.github.io/recharts/>).

**Variation calling.** Raw paired-end reads were first cleaned by NGS QC Toolkit v2.3.3<sup>48</sup>, with criteria of a Phred quality score greater than 20 and a percentage of read lengths that met the given quality greater than 70. Clean paired-end reads of each accession were mapped to the *E. crus-galli* reference genome (STB08)<sup>39</sup> using BOWTIE2 v2.2.1<sup>49</sup> with the default settings. Consecutive pipelines, mainly using SAMtools v0.1.19<sup>50</sup> and GATK v2.3<sup>51</sup>, were applied to detect whole-genome variations (SNPs and indels). To meet the criteria for variation discovery, the SNP and indel calls were filtered according to the following parameters with the customized scripts: QUAL  $> 30$ , DP  $< 5$ , QD  $< 2$ , MQ  $< 20$ , FS  $> 60$ , HaplotypeScore  $> 13$ , ReadPosRankSum  $< -8$ , MAF  $> 0.01$  and integrity rate  $> 0.8$ . The variations were annotated by SnpEff v3.6<sup>52</sup> and summarized by customized scripts.

**Phylogenetic analysis.** A total of 2.19 million whole-genome SNPs with filtering standards of an MAF greater than 0.05, an integrity rate greater than 0.9 and a heterozygous site ratio less than 0.1 were used for phylogenetic analysis. An approximate maximum-likelihood phylogenetic tree of all 328 samples was constructed using FastTree v2.1<sup>53</sup> with 1,000 replicates for bootstrap confidence analysis. To test whether MIC evolved from NMC, we constructed a phylogenetic tree using the paternal donor *E. oryzicola* (ZJU2) as an outgroup<sup>29,31</sup>. We used iTOL ([itol.embl.de](http://itol.embl.de)) to show and modify the constructed trees<sup>54</sup> and conducted PCA with the smartPCA script of EIGENSOFT (v6.1.3) with the default settings<sup>55</sup>.

**Divergence time estimation.** The divergence time of MIC from NMC was estimated using SMC++ v1.13.1<sup>52</sup>, which can infer the effective population size history based on whole-genome sequence data and is powerful for recovering history for short timescales. As the SMC++ split model requires no gene flow after the split of two populations, only samples with typical mimetic and non-mimetic types in MIC and NMC, respectively, were selected. Those with outlier phenotypes in each population, which may be caused by gene flow, were excluded from this analysis. Seven and six lineages in MIC and NMC, respectively, were then randomly chosen as the distinguished individuals to improve the estimation of effective population size with SMC++ estimate, and the divergence time was inferred with SMC++ split. Pairwise sequentially Markovian coalescent analysis (v0.6.5-r67)<sup>56</sup> was also used in the demographic history inference of MIC and NMC by selecting individuals with high sequencing depth (greater than 20 $\times$ ). Consensus sequences were generated by SAMtools<sup>50</sup>. To improve the accuracy, sites with read depth less than 10 or greater than 50 were filtered out. Parameter-*p* (which specifies the atomic time intervals) was set twice as '4 + 30  $\times$  2 + 4 + 6 + 10' and '4 + 25  $\times$  2 + 4 + 6' and other parameters were set as default. The mutation rate was assumed as  $\mu = 6.5 \times 10^{-9}$  mutations  $\times bp^{-1} \times generation^{-1}$  as rice<sup>57</sup> and 1 yr per generation were adopted for both SMC++ and pairwise sequentially Markovian coalescent analysis.

**Detection of selection signals.** VCFtools v0.1.15<sup>58</sup> was used to calculate genetic statistics across the whole genome with a 50 kb sliding window and a step size of 20 kb for  $\pi$  and  $F_{ST}$  and a 20 kb sliding window for Tajima's *D* using the SNP dataset with an integrity ratio of  $> 0.8$ . Windows with fewer than 10 SNPs were eliminated. The ROD was then calculated with a customized script. An empirical threshold of 5% was chosen to find outlier windows of ROD and  $F_{ST}$  values. Windows with a Tajima's *D* value less than the 95% confidence limit for the

corresponding population size were considered to be significant<sup>59</sup>. Windows showing significant negative Tajima's *D* values in both MIC and NMC were ruled out if the  $F_{ST}$  values of the windows between the two groups were very low (less than the empirical threshold of 5%). Outlier windows with interval lengths less than 50 kb were merged for each detection. Only genes with significant differences in variation between MIC and NMC ( $P < 1 \times 10^{-20}$ , Fisher's exact test) and potential impacts (MODIFIER, MODERATE and HIGH in the SnpEff annotation results) were considered candidate genes under selection during the mimicry process. Homologues of plant architecture-related genes in *E. crus-galli* were annotated against known plant architecture-related genes by BLASTP v2.6.0 with a cut-off value less than  $1 \times 10^{-20}$  (Supplementary Table 6).

**Haplotype analysis of LA1.** SNPs and indels in scaffold 445 containing the LA1 genetic region were phased using Beagle v5.0 with the default settings<sup>60</sup>. Single-gene genetic statistics were computed by the R package PopGenome v2.2.4<sup>61</sup>;  $r^2$  was calculated for every pair of SNPs given a window size of 5 Mb by Plink v1.9<sup>62</sup> and was illustrated by the R package LDheatmap v0.99-5<sup>63</sup>. To analyse the selection pattern of LA1 in rice, whole-genome resequencing data of 100 *Oryza sativa* ssp. *japonica* accessions and 43 *O. rufipogon* accessions were downloaded from the National Center for Biotechnology Information<sup>64-66</sup> (Supplementary Table 10). Cleaned reads were mapped to the rice reference genome (MSU6.1). SNP calling, scanning of selection signals and LA1 haplotype analysis were then performed.

**RNA-seq analysis.** Accessions JS26 (MIC) and JX58 (NMC), with typical mimetic and non-mimetic characters, respectively, were chosen for transcriptome sequencing. RNA was extracted from collected tiller bases (~1 cm, tillering stage, 3 weeks after three-leaf-stage seedlings were transplanted into the paddy fields) and sequenced with a routine protocol by Illumina HiSeq 4000. Adaptor sequences, low-quality sequences and empty tags were removed before mapping reads to the reference genome of *E. crus-galli*. Reads were mapped to the reference genome (STB08)<sup>39</sup> by TopHat v2.1.1<sup>67</sup>. The number of tags mapped to genes was normalized to FPKM values, and differentially expressed gene analysis was performed by Cufflinks v2.2.1<sup>68</sup>. Differentially expressed genes, which were required to exhibit a minimum fold change of 2, were identified on the basis of an adjusted false discovery rate of 0.05.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The genomic resequencing and RNA-seq data included in this study were deposited into the BIG data centre (<https://bigd.big.ac.cn/>) under accession number PRJCA001519.

## Code availability

The custom scripts and pipelines used in this study have been deposited in Github ([https://github.com/bioinplant/Vavilovian\\_mimicry](https://github.com/bioinplant/Vavilovian_mimicry)).

Received: 10 April 2019; Accepted: 5 August 2019;

Published online: 16 September 2019

## References

- Barrett, S. C. H. Crop mimicry in weeds. *Econ. Bot.* **37**, 255–282 (1983).
- Pasteur, G. A classificatory review of mimicry systems. *Annu. Rev. Ecol. Syst.* **13**, 169–199 (1982).
- Vavilov, N. I. The origin, variation, immunity and breeding of cultivated plants (translation by K. S. Chester). *Chron. Bot.* **13**, 1–366 (1951).
- McElroy, J. S. Vavilovian mimicry: Nikolai Vavilov and his little-known impact on weed science. *Weed Sci.* **62**, 207–216 (2014).
- Dasmahapatra, K. K. et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
- Iijima, T. et al. Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. *Sci. Adv.* **4**, ea05416 (2018).
- Jay, P. et al. Supergene evolution triggered by the introgression of a chromosomal inversion. *Curr. Biol.* **28**, 1839–1845 (2018).
- Kunte, K. et al. *doublesex* is a mimicry supergene. *Nature* **507**, 229–232 (2014).
- Martin, A. et al. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a *Wnt* ligand. *Proc. Natl Acad. Sci. USA* **109**, 12632–12637 (2012).
- Nadeau, N. J. Genes controlling mimetic colour pattern variation in butterflies. *Curr. Opin. Insect Sci.* **17**, 24–31 (2016).
- Nadeau, N. J. et al. The gene *cortex* controls mimicry and crypsis in butterflies and moths. *Nature* **534**, 106–110 (2016).
- Nishikawa, H. et al. A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat. Genet.* **47**, 405–409 (2015).

13. Reed, R. D. et al. *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**, 1137–1141 (2011).
14. Timmermans, M. J. T. N. et al. Comparative genomics of the mimicry switch in *Papilio dardanus*. *Proc. R. Soc. B* **281**, 20140465 (2014).
15. Zhang, W., Dasmahapatra, K. K., Mallet, J., Moreira, G. R. P. & Kronforst, M. R. Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biol.* **17**, 25 (2016).
16. Gianoli, E. & Carrasco-Urra, F. Leaf mimicry in a climbing plant protects against herbivory. *Curr. Biol.* **24**, 984–987 (2014).
17. Pannell, J. R. & Farmer, E. E. Mimicry in plants. *Curr. Biol.* **26**, R784–R785 (2016).
18. Schaefer, H. M. & Ruxton, G. D. Deception in plants: mimicry or perceptual exploitation? *Trends Ecol. Evol.* **24**, 676–685 (2009).
19. Schluter, P. M. & Schiestl, F. P. Molecular mechanisms of floral mimicry in orchids. *Trends Plant Sci.* **13**, 228–235 (2008).
20. Barlow, B. A. & Wiens, D. Host-parasite resemblance in Australian mistletoes: case for cryptic mimicry. *Evolution* **31**, 69–84 (1977).
21. Kellner, A., Ritz, C. M., Schlittenhardt, P. & Hellwig, F. H. Genetic differentiation in the genus *Lithops* L. (Ruschioideae, Aizoaceae) reveals a high level of convergent evolution and reflects geographic distribution. *Plant Biol.* **13**, 368–380 (2011).
22. Barrett, S. C. H. & Wilson, B. F. Colonizing ability in the *Echinochloa crus-galli* complex (barnyard grass). 2. Seed biology. *Can. J. Bot.* **61**, 556–562 (1983).
23. Barrett, S. C. H. & Wilson, B. F. Colonizing ability in the *Echinochloa crus-galli* complex (barnyard grass). 1. Variation in life history. *Can. J. Bot.* **59**, 1844–1860 (1981).
24. Barrett, S. C. H. in *Applied Population Biology* (eds Jain, S. K. & Botsford, L.) 91–120 (Kluwer Academic, 1992).
25. Barrett, S. C. H. in *Weed Management in Agroecosystems: Ecological Approaches* (eds Altieri, M. & Liebman, M. Z.) 57–75 (CRC, 1988).
26. Tominaga, T. & Fujimoto, T. Awn of darnel (*Lolium temulentum* L.) as an anthropogenic dispersal organ: a case study in Malo, south-western Ethiopia. *Weed Biol. Manage.* **4**, 218–221 (2004).
27. Senda, T. & Tominaga, T. Inheritance mode of the awnlessness of darnel (*Lolium temulentum* L.). *Weed Biol. Manage.* **3**, 46–48 (2003).
28. Fuller, D. Q. & Stevens, C. J. Open for competition: domesticated, parasitic domesticoids and the agricultural niche. *Archaeol. Int.* **20**, 110–121 (2017).
29. Guo, L. B. et al. *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat. Commun.* **8**, 1031 (2017).
30. Fuller, D. Q. & Qin, L. Water management and labour in the origins and dispersal of Asian rice. *World Archaeol.* **41**, 88–111 (2009).
31. Aoki, D. & Yamaguchi, H. Genetic relationship between *Echinochloa crus-galli* and *Echinochloa oryzicola* accessions inferred from internal transcribed spacer and chloroplast DNA sequences. *Weed Biol. Manage.* **8**, 233–242 (2008).
32. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
33. Li, P. J. et al. *LAZY1* controls rice shoot gravitropism through regulating polar auxin transport. *Cell Res.* **17**, 402–410 (2007).
34. Wang, Y. D., Zhang, T., Wang, R. C. & Zhao, Y. D. Recent advances in auxin research in rice and their implications for crop improvement. *J. Exp. Bot.* **69**, 255–263 (2018).
35. Strohm, A. K., Baldwin, K. L. & Masson, P. H. Multiple roles for membrane-associated protein trafficking and signaling in gravitropism. *Front. Plant Sci.* **3**, 274 (2012).
36. Wang, B., Smith, S. M. & Li, J. Y. Genetic regulation of shoot architecture. *Annu. Rev. Plant Biol.* **69**, 437–468 (2018).
37. Wang, Y. & Li, J. Molecular basis of plant architecture. *Annu. Rev. Plant Biol.* **59**, 253–279 (2008).
38. Wu, X., Tang, D., Li, M., Wang, K. & Cheng, Z. Loose plant Architecture1, an INDETERMINATE DOMAIN protein involved in shoot gravitropism, regulates plant architecture in rice. *Plant Physiol.* **161**, 317–329 (2013).
39. Zhang, N. et al. A core regulatory pathway controlling rice tiller angle mediated by the *LAZY1*-dependent asymmetric distribution of auxin. *Plant Cell* **30**, 1461–1475 (2018).
40. Hu, X. M. et al. The U-Box E3 ubiquitin ligase TUD1 functions with a heterotrimeric G alpha subunit to regulate brassinosteroid-mediated growth in rice. *PLoS Genet.* **9**, e1003391 (2013).
41. Qiao, S. L. et al. The RLA1/SMOS1 transcription factor functions with *OsBZR1* to regulate brassinosteroid signaling and rice architecture. *Plant Cell* **29**, 292–309 (2017).
42. Tong, H. N. et al. DWARF AND LOW-TILLERING acts as a direct downstream target of a GSK3/SHAGGY-like kinase to mediate brassinosteroid responses in rice. *Plant Cell* **24**, 2562–2577 (2012).
43. Yamamuro, C. et al. Loss of function of a rice brassinosteroid insensitive1 homolog prevents internode elongation and bending of the lamina joint. *Plant Cell* **12**, 1591–1605 (2000).
44. Dong, Z. B. et al. Maize *LAZY1* mediates shoot gravitropism and inflorescence development through regulating auxin transport, auxin signaling, and light response. *Plant Physiol.* **163**, 1306–1322 (2013).
45. Yoshihara, T. & Iino, M. Identification of the gravitropism-related rice gene *LAZY1* and elucidation of *LAZY1*-dependent and -independent gravity signaling pathways. *Plant Cell Physiol.* **48**, 678–688 (2007).
46. Yoshihara, T., Spalding, E. P. & Iino, M. *AtLAZY1* is a signaling component required for gravitropism of the *Arabidopsis thaliana* inflorescence. *Plant J.* **74**, 267–279 (2013).
47. You, X. The rice production and farming technique in the Song Dynasty. *Chin. J. Rice Sci.* **1**, 35–41 (1986).
48. Patel, R. K. & Jain, M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* **7**, e30619 (2012).
49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
50. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
51. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
52. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strainw1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
53. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
54. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
55. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
56. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
57. Molina, J. et al. Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl Acad. Sci. USA* **108**, 8351–8356 (2011).
58. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
59. Tajima, F. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
60. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
61. Pfeifer, B., Wittelsburger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: an efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).
62. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
63. Shin, J. H., Blay, S., McNeney, B. & Graham, J. LDheatmap: an R function for graphical display of pairwise linkage disequilibrium between single nucleotide polymorphisms. *J. Stat. Softw.* **16**, Code Snippet 3 (2006).
64. Huang, X. et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
65. Qiu, J. et al. Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nat. Commun.* **8**, 15323 (2017).
66. Wang, W. et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43–49 (2018).
67. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
68. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).

## Acknowledgements

We thank H. Yamaguchi, S. Ge and G. Chen for their useful comments. This work was supported by the National Natural Science Foundation (grant number 9143511), the Zhejiang Natural Science Foundation (grant number LZ17C130001), the Jiangsu Collaborative Innovation Center for Modern Crop Production, the 111 Project (grant number B17039) and the China Agriculture Research System (grant number CARS-01-02A).

## Author contributions

L.F. conceived the study. D.W., C.-Y.Y., L.J., J.Q., M.C. and F.L. analysed the data. W.T., Y.L. and X.Y. performed the phenotyping. M.P.T., K.M.O., Y.W. and H.X. advised on the data analysis. M.P.T. edited the manuscript. C.-Y.Y., L.F. and D.W. wrote the manuscript. All authors read and contributed to the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41559-019-0976-1>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to L.F.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Wget (version 1.11.4) was used to download raw reads of Echinochloa oryzicola, Oryza sativa and O. rufipogon from NCBI SRA. Morphological data was collected manually.

Data analysis

For variation calling and annotation, we used NGS QC Toolkit(v2.3.3), Bowtie2 (v2.2.1), SAMtools (v0.1.19), GATK (v2.3) and SnpEff (v3.6). For phylogeny analysis, FastTree (v2.1), SNPRelate (v1.0.1), iTOL (itol.embl.de) were used. For population historical demography analysis, SMC++ (v1.13.1) and PSMC (v0.6.5-r67) were used. For selection analysis, we used VCFtools (v0.1.15), PopGenome (v2.2.4) and BLASTP (v2.6.0). For gene expression analysis, we used TopHat (v2.1.1) and Cufflinks (v2.2.1). For haplotype analysis, we used Beagle (v5.0), Plink (v1.9) and LDheatmap (v0.99-5). Custom codes are available at [https://github.com/bioinplant/Vavilovian\\_mimicry](https://github.com/bioinplant/Vavilovian_mimicry).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the raw reads generated in this study have been deposited in NCBI under the accession number PRJNA526761

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	To uncover a striking example of Vavilovian mimicry in barnyardgrass ( <i>Echinochloa crus-galli</i> ), We have carried out comparisons of mimetic and nonmimetic <i>E. crus-galli</i> populations in the Yangtze River basin based on phenotypic analysis and genome resequencing, which provided first clear genomic evidence of human selection in Vavilovian mimicry.
Research sample	328 accessions of <i>Echinochloa crus-galli</i> from the Yangtze River basin were phenotyped at the seedling stage and genotyped by whole-genome resequencing.
Sampling strategy	We focused the morphology of <i>E. crus-galli</i> at the seedling stage, thus the traits were recorded manually at the tillering stage three weeks after the three-leaf-stage seedlings were transplanted into the paddy fields. Whole-genome resequencing was carried by Illumina HiSeq 4000 platform in a routine protocol.
Data collection	Phenotypic information was collected by Lu Group. Raw reads of WGS and RNA-seq was generated by Illumina platform in Shengting Biotech, Taizhou, China. Public data was downloaded from NCBI.
Timing and spatial scale	Samples were planted and phenotyped at the seedling stage in 2017 and 2018.
Data exclusions	No data were excluded from the analyses.
Reproducibility	This is not applicable as there is no experiments involved.
Randomization	This is not applicable as there is no experiments involved.
Blinding	All designs were transparent to every participants.
Did the study involve field work?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

## Field work, collection and transport

Field conditions	Experimental paddy fields at the China National Rice Research Institute (CNRRI), with average annual 18 centigrade and 1400 mm annual rainfall.
Location	Fuyang, Hangzhou, China (30.0818° N, 119.9307° E)
Access and import/export	The Fan group worked in Zhejiang University, Hangzhou, China and Lu group worked in CNRRI, Fuyang, China.
Disturbance	We used fields which were originally used for rice planting.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging