CellPress
Partner Journal

Molecular Plant
Research Report

Population genomic analysis reveals domestication of cultivated rye from weedy rye

Yanqing Sun[1,2,3,10], Enhui Shen[1,2,3,10], Yiyu Hu[1], Dongya Wu[1], Yu Feng[4], Sangting Lao[1], Chenfeng Dong[1], Tianyu Du[1], Wei Hua[5], Chu-Yu Ye[1], Jinhuan Zhu[5], Qian-Hao Zhu[6], Daguang Cai[7], Lidia Skuza[8], Jie Qiu[9] and Longjiang Fan[1,2,3,*]

[1]Institute of Crop Science & Institute of Bioinformatics, Zhejiang University, Hangzhou 310058, China

[2]Zhejiang University Zhongyuan Institute, Zhengzhou 450000, China

[3]Shandong (Linyi) Institute of Modern Agriculture of Zhejiang University, Linyi 310014, China

[4]Institute of Ecology, Zhejiang University, Hangzhou 310058, China

[5]Institute of Crops, Zhejiang Academy of Agricultural Sciences, Hangzhou 322105, China

[6]CSIRO Agriculture and Food, GPO Box 1700, Canberra, ACT 2601, Australia

[7]Department of Molecular Phytopathology and Biotechnology, Christian-Albrechts-University of Kiel, 24118 Kiel, Germany

[8]Institute of Biology, University of Szczecin, 71-415 Szczecin, Poland

[9]Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, Shanghai 200235, China

[10]These authors contributed equally to this article.

***Correspondence: Longjiang Fan (fanlj@zju.edu.cn)**

https://doi.org/10.1016/j.molp.2021.12.015

ABSTRACT

Rye (*Secale cereale*) is an important crop with multiple uses and a valuable genetic resource for wheat breeding. However, due to its complex genome and outcrossing nature, the origin of cultivated rye remains elusive. The geneticist N.I. Vavilov proposed that cultivated rye had been domesticated from weedy rye, rather than directly from wild species like other crops. Unraveling the domestication history of rye will extend our understanding of crop evolution and upend our inherent understanding of agricultural weeds. To this end, in this study we generated the 8.5 Tb of whole-genome resequencing data from 116 worldwide accessions of wild, weedy, and cultivated rye, and demonstrated that cultivated rye was domesticated directly from weedy relatives with a similar but enhanced genomic selection by humans. We found that a repertoire of genes that experienced artificial selection is associated with important agronomic traits, including shattering, grain yield, and disease resistance. Furthermore, we identified a composite introgression in cultivated rye from the wild perennial *Secale strictum* and detected a 2-Mb introgressed fragment containing a candidate ammonium transporter gene with potential effect on the grain yield and plant growth of rye. Taken together, our findings unravel the domestication history of cultivated rye, suggest that interspecific introgression serves as one of the likely causes of obscure species taxonomy of the genus *Secale*, and provide an important resource for future rye and wheat breeding.

Key words: *Secale cereale*, domestication, weedy rye, introgression, Vavilovian hypothesis

Sun Y., Shen E., Hu Y., Wu D., Feng Y., Lao S., Dong C., Du T., Hua W., Ye C.-Y., Zhu J., Zhu Q.-H., Cai D., Skuza L., Qiu J., and Fan L. (2022). Population genomic analysis reveals domestication of cultivated rye from weedy rye. Mol. Plant. **15**, 552–561.

INTRODUCTION

As a member of the Triticeae tribe, rye (*Secale cereale*, $2n = 2x = 14$, genome RR) is an important crop for bread making and feed forage worldwide, particularly in central and northeastern Europe. Rye is outstanding for its vigorous growth, high tolerance of abiotic and biotic stress, and exceptional capability to grow on poor soil and in cold climates. The phylogenetically close relationship to wheat (*Triticum aestivum*) facilitated rye as a widely used genetic resource for introducing genes useful for wheat improvement, in the form of chromosome segment introgression (Moskal et al., 2021). Such a well-known example is the creation of the wheat–rye 1RS/1BL chromosome translocation, which has been widely used worldwide because the 1RS

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

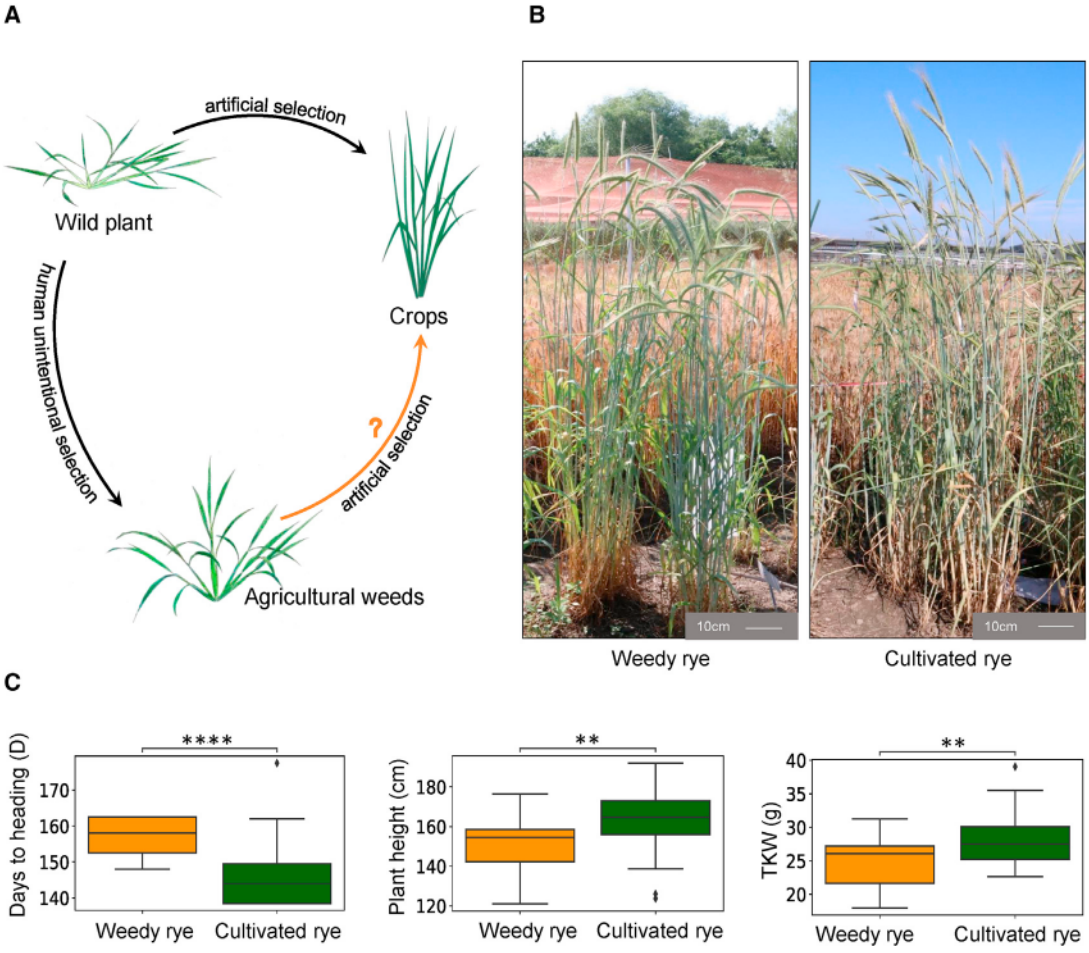

Figure 1. The origin and phenotyping of weedy and cultivated rye.

(A) A diagram showing two routes of crop domestication: directly from wild species or from agricultural weeds derived from wild species, i.e., the Vavilovian hypothesis.

(B) Representative mature plants of weedy (PI 618672) and cultivated (CISE 86) rye grown in the experimental field in Hangzhou, China.

(C) Comparison of agronomic traits between cultivated and weedy rye. The traits are heading date, plant height at maturity, and 1000-kernel weight (TKW). In the boxplots, the horizontal lines show the median values, and whiskers show the 25% and 75% quartile values. Significance test was performed using *t*-test. ****$P < 0.0001$; **$P < 0.01$.

harbors disease-resistance genes: *Pm8* for powdery mildew and *Yr9* for stripe rust diseases in wheat (Crespo-Herrera et al., 2017).

Over 100 years ago, the geneticist Nikolai Ivanovich Vavilov hypothesized that cultivated rye had been derived from agricultural weeds, rather than directly from wild progenitors (Vavilov, 1917). He termed rye as a secondary crop, since it had experienced a transitory phase as a weed that might have acquired certain "domestication" traits under unintentional selection by humans (Ye et al., 2019) and a "real" domestication process after humans discovery of its superior performance under extreme conditions (Vavilov, 1917; McElroy, 2014; Schreiber et al., 2019) (Figure 1A). Another example of such crops is oat (*Avena sativa*). In their journey to becoming a crop, they have an additional evolutionary node (i.e., weeds) compared with regular crops that originated from wild species (Ye and Fan, 2021). However, the hypothesis that weeds could be domesticated to become crops has not yet been fully proven even though it seems to fit with the observations.

The hypothesis (i.e., rye originally domesticated from the weeds growing in wheat and barley fields) has been under investigation for a long time, based on not only plant morphology and on-the-spot investigation (Sakamoto, 1982), but also archaeology (Behre, 1992; Grikpėdis and Matuzevičiūtė, 2016), philology (Sencer and Hawkes, 1980; Behre, 1992), cytology (Khush and Stebbins, 1961), and limited molecular markers. Archaeological findings of rye in the Neolithic period and European Bronze Age (~3500 BC) are scarce. Usually only single grains or small proportions mixed with other primary cereals were evident, indicating a status of rye as a weed at the time. The status of rye changing from preadapted weed into a "fully domesticated" crop probably occurred in the period of the early Iron Age (Behre, 1992; Grikpėdis and Matuzevičiūtė, 2016). With extensive advances in molecular biology technologies, molecular markers such as random amplified polymorphic DNA, amplified fragment length polymorphisms, and SNPs have been widely applied in studying the relationship and genetic diversity of rye (Vences et al., 1987; Chikmawati et al., 2005; Parat et al., 2016; Gholizadeh Sarcheshmeh et al., 2018;

Species[a]	Number of accessions	Average sequencing depth based on clean reads (×)	Mapping rate (%)	Genome coverage (%)	Source
S. cereale subsp. *cereale* (landrace)	48	9.4	99.5	81.5	This study
S. cereale subsp. *cereale* (inbred line)	10	12.5	99.5	77.9	Bauer et al. (2017)
S. cereale subsp. *segetale*, etc. (weedy)	30	9.5	99.4	82.9	This study
S. cereale subsp. *vavilovii*	10	10	98.9	80.8	This study
S. strictum (perennial wild)	10	9.5	99.1	72.2	This study
S. sylvestre (annual wild)	8	9.6	98.9	57.5	This study
Total/average	116	9.8	99.3	79.2	

Table 1. Summary of genome resequencing and mapping results of the rye accessions used in this study.
The reference genome Lo7 was used to estimate mapping rate and genome coverage of clean reads.
[a]See Supplemental Table 1 for accession details.

Larsson et al., 2019; Schreiber et al., 2019). However, most studies have been focused on species taxonomy classification and the design of a high-density genotyping array for breeding, and few have been dedicated to the evolution of rye, leading to largely ambiguous ideas about its origin.

The genus *Secale* comprises only three species with different life cycles and mating systems (Schreiber et al., 2019; Daskalova and Spetsov, 2020). *S. cereale* is an annual species with various domesticated statuses, containing putative wild progenitor *S. cereale* subsp. *vavilovii*, domesticated rye *S. cereale* subsp. *cereale*, and other subspecies considered as weedy or feral rye. *Secale sylvestre* and *Secale strictum* are the other two wild species, characterized by annual selfing and perennial outcrossing, respectively. In general, all three species can be crossed with one another. Previous cross-pollination studies have reported chromosomal translocations as possible causes of reduced fertility in hybrids (Stutz, 1957; Khush and Stebbins, 1961). Cultivated rye has a large genome (nearly 8 Gb in size), and high-quality genome sequences for it have been recently generated (Bauer et al., 2017; Li et al., 2021; Rabanus-Wallace et al., 2021). The genomic assemblies provide an unprecedented opportunity to examine the genetic footprints underlying its unique trajectory originating from weeds and enable the inspection of a composite introgression from the wild population. In this study, we report a genus-level whole-genome sequencing study to determine the domestication history of cultivated rye, with a total of 116 accessions of wild, weedy, and cultivated rye, which were collected mainly from the center of a great diversity of rye, southwestern Asia, and the major growing region of rye in Europe. Our results provide genomic evidence for its origin from the domestication of weedy rye and a valuable genomic resource for Triticeae genetic research and breeding in the future.

RESULTS AND DISCUSSION

Sampling, sequencing, and phenotyping of *Secale* species

The genomic data of 116 worldwide rye accessions, including 58 cultivated ryes (*S. cereale* subsp. *cereale*) (48 landraces and 10 inbred lines), 30 weedy ryes (*S. cereale* subsp. *segetale*, etc.), 10 *S. cereale* subsp. *vavilovii*, 10 perennial wild *S. strictum*, and 8 annual wild *S. sylvestre*, were analyzed to investigate the evolutionary origin of rye (Table 1 and Supplemental Table 1). The data of 105 accessions were newly generated in this study, with a total of 8.5 Tb of high-quality clean reads and an average sequencing depth of ~10-fold for each accession (based on the 7.9-Gb estimated genome size of rye). The clean paired-end reads of each accession were individually mapped against the ~6.7-Gb rye reference genome of the inbred line Lo7 (Rabanus-Wallace et al., 2021). While a >80% genome coverage rate was found for *vavilovii*, weedy, and cultivated rye, the genome coverage rate was relatively low for *S. strictum* and *S. sylvestre* (Table 1 and Supplemental Table 1). A total of 154 792 683 high-quality SNPs with a minor allele frequency (MAF) greater than 0.01 and integrity rate greater than 0.8, and 22 828 272 insertions or deletions (indels) were identified (Supplemental Figure 1 and Supplemental Table 2).

Of the 105 accessions resequenced in this study, 90 had seeds available, while the remaining 15 had only DNA samples available. Seeds of the 90 rye accessions were sown in the field (Hangzhou, China) for phenotyping. Compared with the cultivated rye, the weedy rye had a lower thousand-kernel weight, shorter plant height, stronger tillering, and later heading date (Figure 1B and 1C). Generally, these observations were consistent with those previously reported (Shang et al., 2009; Akhalkatsi, 2016). However, we also found exceptions. Three *S. strictum* accessions (perennial wild) exhibited compact leaves and erect stem, different from the phenotypes of other *S. strictum* accessions (Supplemental Figure 2A and 2B). Four cultivated accessions were more weedy-like. These seven accessions were found to be outliers based on phylogenetic analysis (see next section).

Genomic differentiation between weedy and cultivated rye

The maximum-likelihood phylogenetic tree based on 908 599 high-quality synonymous SNPs showed a clear separation of *S. sylvestre* (annual wild), *S. strictum* (perennial wild), and *S. cereale* subspecies (Figure 2A), a result further supported by principal-component analysis (PCA) (Supplemental Figure 3A and 3B) and consistent with previously reported results (Schreiber et al., 2019). The three *S. strictum* accessions that were phenotypically different from other *S. strictum* accessions were grouped with *S. cereale*. The *S. cereale* subspecies clade included *vavilovii*, weedy, and cultivated rye, indicating a close

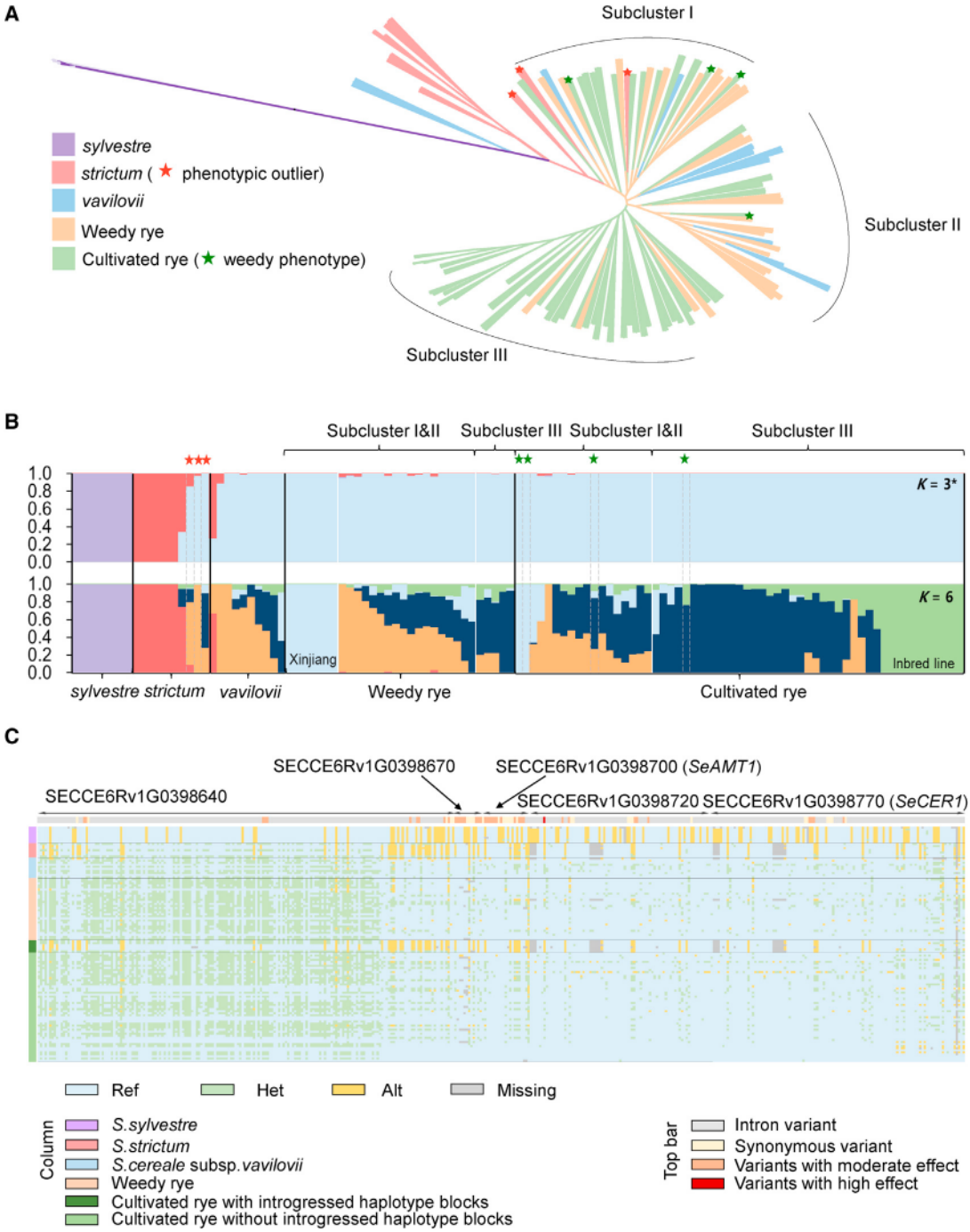

Figure 2. Phylogeny, population structure, and introgression of the 116 *Secale* accessions.

(A) An approximate maximum-likelihood tree of the 116 wild, weedy, and cultivated rye accessions generated using a total of 908 599 synonymous SNPs with 1000 bootstraps. Each branch represents an accession and is color coded to the population to which it belongs. The outlier accessions of *S. strictum* and cultivated rye, according to phenotypic observation, are marked with red and green asterisks, respectively.

(B) Population structure of the 116 accessions with $K = 3$ and $K = 6$. The subcluster information and the outliers presented in **(A)** are shown at the top of the graph. "*" refers to the best grouping number, i.e., with the lowest cross-validation error.

(C) The haplotype block introgressed from *S. strictum* into cultivated rye. Haplotype diversity of the 540 variants identified in five genes among wild (*S. sylvestre*, *S. strictum*, and *S. cereale* subsp. *vavilovii*), weedy, and cultivated rye populations (column). The region of each gene is indicated by a line with an arrowhead at both ends. The genotypes (ref, alt, het, and missing) of each variant are color coded with the genome of the cultivated rye Lo7 as the reference.

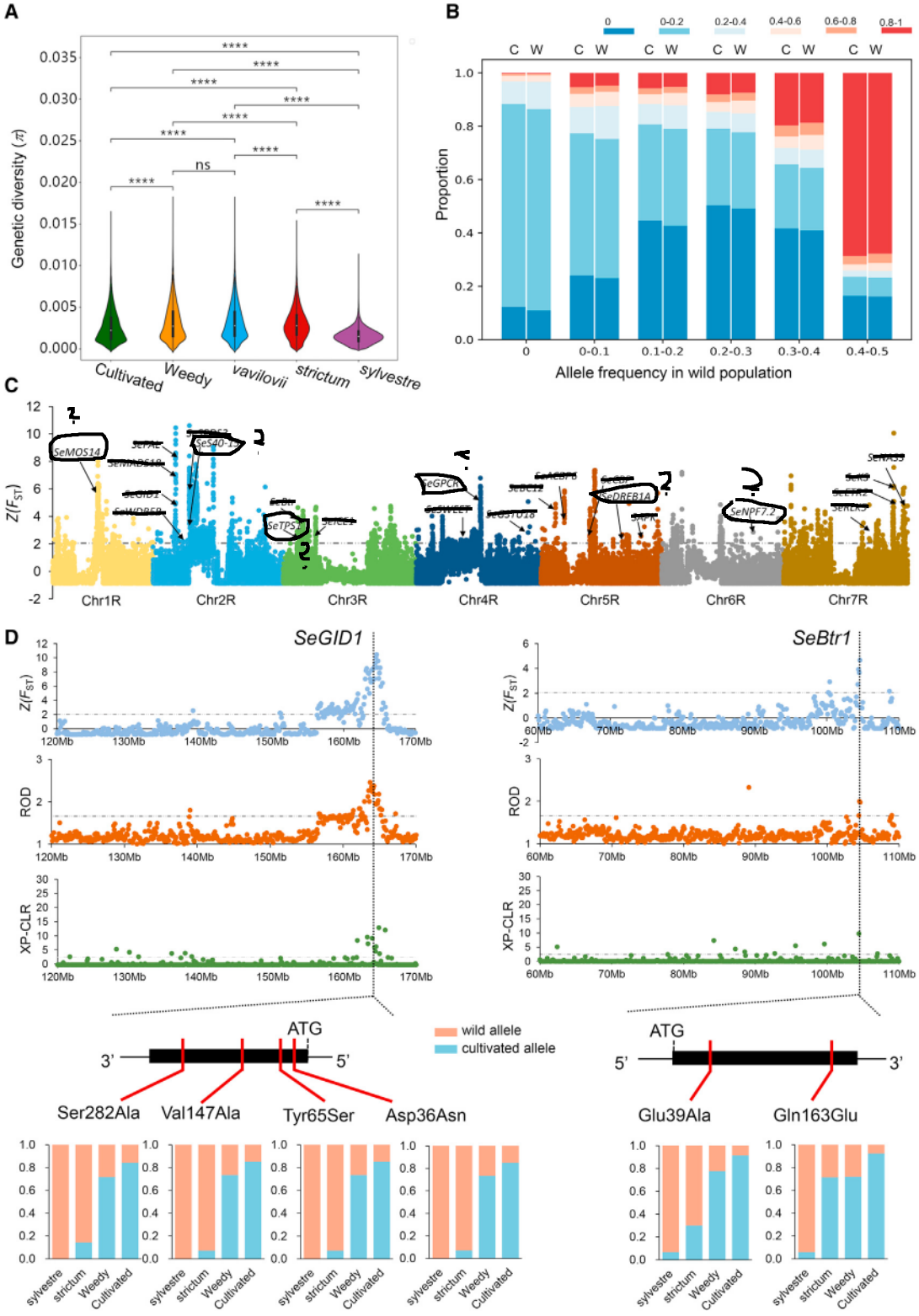

Figure 3. Genomic signatures of human selection on rye during domestication.

(A) Comparison of genetic diversity among wild, weedy, and cultivated rye. ****$P < 0.0001$; ns, no significant difference.

(B) Allele frequency in cultivated and weedy ryes, with the frequency range in wild rye shown on the x axis. "C" and "W" refer to cultivated and weedy rye, respectively.

(C) Distribution of the Z-transformed F_{ST} values between weedy and cultivated rye along seven chromosomes of rye. The genome-wide threshold was defined by the top 5% region. Domestication-related candidate genes are marked by arrows.

(legend continued on next page)

relationship between weedy and cultivated rye. The *S. cereale* subspecies clade could be further separated into three distinct subclusters. While subcluster III included mainly cultivated rye, subclusters I and II contained all three types of rye, including the three *S. strictum* outliers and the four weed-like cultivated rye accessions. Most *vavilovii* accessions (9/10 or 90%) and weedy rye accessions (26/30 or 86.67%) were grouped into subclusters I and II, suggesting that *vavilovii* accessions share a much closer relationship with weedy than with cultivated rye. Subcluster III included 62.5% of the landraces and the 10 inbred lines, implying a different level of domestication of the landraces, with most of them close to the inbred lines. A weak genomic differentiation between the weedy and the cultivated rye populations (F_{ST} = 0.013) was observed, which was further supported by the genetic distance calculated based on pairwise identity-by-state (Supplemental Figure 3C). The phylogenetic relationship was also supported by the population structure analysis using ADMIXTURE, and three ancestral populations (i.e., K = 3) was the best-fit grouping model (with the lowest cross-validation error) (Figure 2B). At K = 6, landraces from subclusters I and II of the phylogenetic tree showed high similarity with the weedy rye of the same subclusters. We noted that the three *S. strictum* outliers clustered together but were separated from other *S. strictum* accessions, casting doubt on them being *de facto S. strictum*. These three *S. strictum* outliers and the four weed-like cultivated rye accessions (possible intermediates of weedy and cultivated rye) were excluded from the following analyses.

Demographic inference and introgression

Six different demographic simulation origin models of cultivated rye were further tested using *fastsimcoal2* (Excoffier et al., 2013) (Supplemental Figure 4). Among the six scenarios, no. 4 showed the lowest delta likelihood and Akaike information criterion (AIC) score (Supplemental Table 3) and fit the genetic data best, indicating that cultivated rye was not derived from wild antecedents directly but shared a common ancestor with weedy rye. Based on model 4, strong and bidirectional gene flow among the cultivated, weedy, and wild rye group was also detected. The earliest divergence between the wild rye and *S. cereale* occurred approximately 6000 years ago, well in agreement with several archaeological records found in central Europe with a ^{14}C age of 4440 BC (Supplemental Table 4). No evident genetic bottleneck was observed in the cultivated rye relative to the weedy rye. The continuous gene flow within the *Secale* genus might have had a profound effect on the domestication of rye.

Further, we used ABBA–BABA statistics (Green et al., 2010; Martin et al., 2015) to detect introgression signals. We observed that the weedy rye from Xinjiang showed distinct genetic components, which are likely to be the result of relatively early differentiation due to geographical isolation, and therefore chose it as the reference population (P1). The f_d statistic, which signifies gene flow when $0 < f_d < 1$ (Martin et al., 2015), was used to calculate the fraction of introgression in cultivated rye from western Asia. As a result, a total of 421.8 Mb of introgressed segments were identified (Supplemental Figure 7). Interestingly, we detected a 2-Mb introgressed haplotype block on chromosome 6R (423–425 Mb), carried by six cultivated ryes (five accessions from an inbred line, AR132), which was similar to that found in *S. strictum* (Figure 2C). In the introgressed block, there are five genes with 540 SNPs and indels, including *SeAMT1* (SECCE6Rv1G0398700), an ortholog of *OsAMT1:1* encoding an ammonium transporter affecting rice grain yield and plant growth (Hoque et al., 2006; Ranathunge et al., 2014), and *SeCER1* (SECCE6Rv1G0398770), potentially regulating environmentally sensitive male sterility (Figure 3) (Zhang et al., 2008b; Ni et al., 2021). Previous breeding efforts allowed self-fertile inbred lines of rye, which were derived from two genetically distinct populations in which the pollen parental lines carry restorer genes allowing reconstitution of male fertility in the final hybrid (Wilde and Miedaner, 2021). *SeCER1* might be one of the candidate key genes related to fertility regulation in rye breeding.

Positive selection in weedy and cultivated rye

We used genetic diversity and cross-population composite likelihood ratio (XP-CLR) to test whether certain regions of the rye genome contain selection signals from the domestication process (i.e., from weedy to domesticated). Assessments of genome-wide nucleotide diversity indicated that cultivated rye harbors the lowest genetic diversity ($\pi = 2.220 \times 10^{-3}$), followed by weedy rye ($\pi = 2.728 \times 10^{-3}$) and *vavilovii* ($\pi = 2.771 \times 10^{-3}$), and then by *S. strictum* ($\pi = 2.772 \times 10^{-3}$) (Figure 3A), suggesting that selection played an important role in diversity reduction during the weedy phase and further during the domestication stage. However, compared with its close relatives wheat (Cheng et al., 2019; Guo et al., 2020; Zhou et al., 2020) and barley (Zeng et al., 2018; Milner et al., 2019), rye did not show apparent reduction in diversity, likely due to its outcrossing nature (mass introgression was suggested by model 4 shown in Supplemental Figure 4). In addition, the allele frequencies calculated based on SNPs seemed to be similar between cultivated and weedy rye, but differed significantly in wild rye (including *S. strictum* and *S. sylvestre*), as demonstrated by the similar distribution patterns of MAF of cultivated and weedy rye in each range of MAF in wild rye (Figure 3B). For the alleles (a total of 15 619 723 SNPs or 10.1% of total SNPs) with a medium frequency ($0.4 < \text{MAF} \leq 0.5$) in the wild rye population, the majority had a frequency less than 0.2 or greater than 0.8 in weedy rye (90.8% of the 15 619 723 SNPs) and cultivated rye (91.9%) (Figure 3B), suggesting that they were almost fixed in the two populations, which might have experienced similar artificial selection by humans, although likely slightly stronger in the cultivated population.

Selective sweeps were determined for weedy and cultivated rye across the whole genome based on Z-transformed F_{ST}, reduction of diversity (ROD = $\pi_{weedy}/\pi_{cultivated}$), and XP-CLR scores. As a

(D) Two examples of candidate domestication-related genes, *SeGID1* and *SeBtr1*. Both genes contain a single exon (represented by black bar). The graphs above the genes show selection signals (Z-transformed F_{ST}, ROD, and XP-CLR values) of the corresponding genes between cultivated and weedy rye. The bar graphs below the genes show the allele frequency of the non-synonymous variants identified in *SeGID1* and *SeBtr1* in four different rye populations.

result, a total of 191 Mb (2.4% of the rye genome) of selective sweeps with 279 genes containing 3450 non-synonymous SNPs and 430 indels was identified with the standard of at least two signals detected (Supplemental Tables 5 and 6). Candidate domestication-related genes, controlling shattering (brittle rachis), grain yield (grain size), disease resistance, etc., were found to be under artificial selection and to have significantly different allele frequency between the wild (*S. sylvestre* and *S. strictum*) and the cultivated/weedy population (Figure 3C). For example, *SeGID1* (SECCE2Rv1G0083180) on chromosome 2R is an ortholog of the rice gibberellin receptor related to plant height. The region containing *SeGID1* was apparently differentiated between weedy and cultivated rye ($Z(F_{ST})$ = 4.412), and its genetic diversity decreased in cultivated rye relative to weedy population (ROD = 2.013) (Figure 3D) (Tanaka et al., 2006; Zhang et al., 2008a). We also noted that the region with *SeBtr1* (SECCE3Rv1G0160350), an ortholog of *Btr1* controlling brittle rachis in barley, was evolved under selection ($Z(F_{ST})$ = 3.881, ROD = 1.665, XP-CLR = 9.752) with the frequency of dominant variants significantly increased from wild rye (*S. sylvestre* and *S. strictum*) to weedy rye and then almost fixed in cultivated rye (Figure 3D), suggesting a continuous effect of artificial selection on the shattering-related genes during rye domestication (Pourkheirandish et al., 2015). In addition, positive selection was evident for *SeETR2*, which potentially affects floral transition and starch accumulation; *SeBC12*, affecting plant height; *SeNPF7.2*, regulating tiller number and grain yield; and *Sebel*, an ortholog of a cytochrome P450, conferring resistance to acetolactate synthase-inhibiting herbicides in rice (for details see Supplemental Figure 5) (Pan et al., 2006; Wuriyanghan et al., 2009; Li et al., 2011; Saika et al., 2014; Fang et al., 2021; Yau et al., 2004).

The role of *S. cereale* subsp. *vavilovii* in rye domestication

Although our results confirmed that cultivated rye was directly domesticated from a weedy population, the roles of wild rye species in the origin of cultivated rye were obscure. According to our analysis, we support the notion that *S. cereale* subsp. *vavilovii* was the immediate wild ancestor, as Vavilov proposed (McElroy, 2014; Daskalova and Spetsov, 2020). The *vavilovii* accessions exhibited little intraspecific substructure and were embedded in the *S. cereale* clade based on principal-component, phylogenetic tree, and genetic structure analyses (Figure 2A and 2B), which has also been confirmed in the study of organelle genomes (Skuza et al., 2019). Phenotypic observation also showed that most *vavilovii* accessions resembled perennial wild rye, with strong tillering at the seedling stage and a crooked stem node after heading, and remained creeping for a longer period of time compared with weedy and cultivated rye, although two accessions (PI 284842 and PI 253957) showed weedy-like characteristics, with larger leaves, relatively erect stems, and no-shattering spikes (Supplemental Figure 2C). In addition, one *vavilovii* accession from Bauer et al. (2017) was found to be far away from other *vavilovii* accessions but with a close genetic affinity to *S. strictum* (Figure 2A). Similar results have been reported in previous studies (Sencer and Hawkes, 1980; Meier et al., 1996). For example, it has been pointed out that, although some *vavilovii* accessions shared a similar karyotype with the *vavilovii* specimen collected by Professor Kuckuck in Northern Iran, some *vavilovii* accessions were most likely weedy rye forms based on their morphological and biochemical characteristics (Vences et al., 1987; Meier et al., 1996). Early studies showed that some *vavilovii* differed from cultivated rye, with at least two structural rearrangements, while the *S. cereale* chromosome karyotype pattern in *vavilovii* has also been observed (Stutz, 1972; Singh, 1977). These observations support *vavilovii* as the likely link between perennial wild rye and weedy/cultivated rye, encompassing two different types: weedy-like and wild-like. One of the possible explanations could be a wrong classification, as suggested by Meier et al. (1996), or a contaminant resulting from natural hybridization. As indicated by demographic simulation and genetic structure, strong gene flows exist between *S. strictum* and *S. cereale* (including *vavilovii*), consistent with previous results suggesting that perennial rye could cross easily to form hybrids (Stutz, 1972). However, our small panel of 10 *vavilovii* accessions may not capture all the genetic diversity of *vavilovii*, and more *bona fide vavilovii* accessions without domesticated admixture are needed to make a firm conclusion on the role of *vavilovii* in the origin and domestication of rye. Further, confirming the possible origin of the outliers by natural hybridization would require generation of an artificial hybrid using the potential parents and comparing their genomic features with those of the outliers.

In conclusion, in addition to providing a valuable genomic resource for rye and wheat research, this study provides a comprehensive and unbiased view of the diversity of different rye populations and genomic evidence for the Vavilovian hypothesis, i.e., the domestication of cultivated rye from weedy relatives. We also identified candidate genes for some important domestication-related traits of rye and provided genetic evidence for introgression from wild *S. strictum* to cultivated rye.

METHODS

Sample collection, phenotyping, and genome resequencing

The 105 accessions of the *Secale* taxa were obtained from several world collections, including the Germplasm Resources Information Network (GRIN; https://npgsweb.ars-grin.gov/gringlobal/search), USA, and Institute of Crop Science, Chinese Academy of Agricultural Sciences (Supplemental Table 1). Seeds (for the 90 accessions with seeds available) were directly sowed in the experimental field of Zhejiang Academy of Agricultural Sciences in 2020–2021 (Hangzhou, China), with six individuals per accession for phenotyping. Phenotypes, including plant height, heading date, grain shattering, and thousand-kernel weight, were recorded. Genomic DNA was extracted from the young leaves using the cetyltrimethylammonium bromide method. Sequencing library construction was constructed using standard protocols of MGI TECH on the MGI2000 platform. All samples were sequenced using the MGI2000 sequencing platform with a paired-end read length of 150 bp and an average clean read sequencing coverage of approximately 10× for each accession. Publicly available genomic data of 11 additional *Secale* accessions (Bauer et al., 2017) were downloaded from NCBI (www.ncbi.nlm.nih.gov). The detailed information of the 116 accessions is given in Supplemental Table 1.

Variant detection and genotyping

The raw paired-end reads were first filtered to get clean data using the NGSQCToolkit v.2.3.3 (Patel and Jain, 2012). The cutoff value for PHRED quality score was set to 20 and the percentage of read length that met the given quality was 70. Clean paired-end reads of each accession were

mapped to the cultivated rye reference genome (Lo7) using BWA-MEM (v.0.7.17-r1188) with the default settings (Li and Durbin, 2009). In light of the limitation of the BWA software regarding chromosome length, we split each chromosome into two parts: the first 500 Mb and the remaining sequence. Reads with an abnormal insert size (greater than 10,000, less than −10,000, or 0) and low mapping quality (<1) were filtered out using BamTools (v.2.5.1) (Barnett et al., 2011). SAMtools v.1.7 (Li et al., 2009) and GATK v.3.7.0 (McKenna et al., 2010) were applied to mark duplicate reads and detect genome-wide variations (SNPs and indels). To have a set of high-quality variants, the SNP calls were filtered according to the following parameters with the customized scripts: biallelic alleles || QD < 2.0 || FS > 60.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0 || SOR > 3.0 || MQ < 40.0. The indels were filtered using "QD < 2.0 || FS > 200.0 || ReadPosRankSum < −20.0". SNPs and indels that did not meet any of the following criteria were further discarded: (1) MAF ≥ 0.01, (2) integrity rate > 0.8, or (3) biallelic sites. The error rate of variant call was as low as 0.16% based on analysis of the proportion of segregating sites using clean reads from the reference accession (Lo7). The effect of the variants was annotated by SnpEff v.5.0 (Cingolani et al., 2012) and summarized by customized scripts. To further avoid potential mistakes on sampling, we estimated missing rate and heterozygous rate for each sample after filtering with VCFtools v.0.1.16 (Danecek et al., 2011). One accession (Lo282) was excluded due to a large missing rate.

Population genetics analysis

A total of 908 599 synonymous SNPs were used for phylogenetic analysis. A phylogenetic tree was constructed using the maximum-likelihood method by IQ-TREE (Minh et al., 2020), with the "-m MFP" option applied to determine the best-fit model and a bootstrap value of 1000. Interactive Tree of Life (Letunic and Bork, 2016) was used to visualize and modify the constructed tree. PCA was performed using the smartPCA script of EIGENSOFT (v.6.1.3) (Price et al., 2006) with the default settings. ADMIXTURE (version 1.3.0) software (Alexander et al., 2009) was used to quantify the genome-wide population structures, with *K* values from 2 to 6 to estimate the standard errors of parameters. To quantify the relatedness between individuals, the pairwise identity-by-state genetic distance matrix of the 116 accessions was calculated using PLINK (v.1.90) (Purcell et al., 2007) with the parameter of distance 1-ibs.

Demographic history inference

The joint site frequency spectrum (SFS) was built using only unlinked synonymous SNPs using easySFS.py (https://github.com/isaacovercast/easySFS). The SNP data of each group were down-projected to an SFS with equivalent sampling sizes across groups to decrease the effects of different levels of missing data between groups. The scenarios for divergence of three groups were set with or without gene flow. We calculated the likelihood function for different demographic scenarios using the software fastsimcoal2 (Excoffier et al., 2013). For each scenario, 100 000 coalescent simulations per likelihood estimation (i.e., *-n* 100 000) and 40 expectation-conditional maximization cycles (-L40) were used as the command line parameters for each run. We ran each model 100 times to obtain the best parameters and estimated likelihoods. The AIC was used to compare different models (Excoffier et al., 2013) and was calculated using the formula AIC = 2*k*–2ln (MaxEstLhood), where *k* is the number of parameters estimated by each model and MaxEstLhood is the ML (max likelihood) function value for each model. The scenario with the lowest AIC value was considered the best. After that, 100 independent DNA polymorphism datasets were simulated as joint conditional SFSs based on the estimated demographic parameters of the best scenario. ML analysis was then applied to each joint SFS over 40 expectation-conditional maximization cycles to obtain confidence intervals for the final estimates.

Detection of selection signals

VCFtools v.0.1.16 (Danecek et al., 2011) was used to calculate the genetic statistics π and F_{ST} across the whole genome with a 100-kb sliding window using the SNPs with an integrity ratio of >0.8. Windows with fewer than 20 SNPs were eliminated. We also performed XP-CLR analysis (Chen et al., 2010) and calculated ROD with a customized script to detect selective sweeps. The XP-CLR score between weedy and cultivated populations was calculated using parameters of –minsnps 20 –size 100000. To detect genes under selection, we ranked the selection sweeps with an XP-CLR score, *Z*-transformed F_{ST}, and ROD value in descending order and considered the top 5% regions with at least two signals detected as selective sweeps. Only genes in the selective sweep regions with potential impacts (MODIFIER, MODERATE, and HIGH in the SnpEff [Cingolani et al., 2012] annotation results) were considered as candidate genes under selection.

Introgression analysis with ABBA–BABA tests

We estimated the f_d values (Martin et al., 2015) across the genome using the python code available at https://github.com/simonhmartin/genomics_general. The sliding window was set with a window size of 100 kb and a step size of 50 kb. The minimum good sites in each window were set to 200 through -m flag. For windows of $D < 0$ or of $D > 0$, but $f_d > 1$, the f_d statistic value becomes meaningless or noisy; therefore, we converted the f_d value to zero. We estimated the f_d statistic value using weedy rye from Xinjiang, China, as P1 and the cultivated rye as P2. *S. strictum* was used as P3 for the analysis. The top 5% regions were considered the candidate introgression region.

Data availability

The genomic resequencing data included in this study have been deposited into the National Genomics Data Center, China (https://bigd.big.ac.cn/), under accession no. PRJCA006012.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This work was supported by the National Natural Science Foundation (grant 9143511), Department of Science and Technology of Zhejiang Province (grant 2020C02002), the Zhejiang Natural Science Foundation (grant LZ17C130001), the Jiangsu Collaborative Innovation Center for Modern Crop Production, and the 111 Project (grant B17039) to L.F.

AUTHOR CONTRIBUTIONS

L.F. conceived the study. Y.S., E.S., Y.H., D.W., Y.F., and C.-Y.Y. analyzed the data. Y.S., Y.H., S.L., C.D., T.D., J.Z., W.H., and C.-Y.Y. performed the phenotyping. Q.-H.Z., D.C., L.S., and J.Q. advised on the data analysis. Q.-H.Z. edited the manuscript. Y.S. and L.F. wrote the manuscript. All authors read and contributed to the manuscript.

ACKNOWLEDGMENTS

We thank Xinming Yang (Institute of Crop Science, Chinese Academy of Agricultural Sciences) for rye genetic materials, Nils Stein (IPK) and Guangwei Li (Henan Agricultural University) for their data on rye genomes, and Xiue Wang (Nanjing Agricultural University) for critical reading of the manuscript. No conflict of interest is declared.

Received: September 16, 2021
Revised: December 17, 2021
Accepted: December 24, 2021
Published: December 27, 2021

REFERENCES

Akhalkatsi, M. (2016). Landraces and wild species of the *Secale* genus in the Georgia (Caucasus ecoregion). Agric. Res. Technol. Open Access J. **1**:1–7.

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. **19**:1655–1664.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., and Marth, G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics **27**:1691–1692.

Bauer, E., Schmutzer, T., Barilar, I., Mascher, M., Gundlach, H., Martis, M.M., Twardziok, S.O., Hackauf, B., Gordillo, A., Wilde, P., et al. (2017). Towards a whole-genome sequence for rye (*Secale cereale* L.). Plant J. **89**:853–869.

Behre, K.E. (1992). The history of rye cultivation in Europe. Veg. Hist. Archaeobot. **1**:141–156.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Res. **20**:393–402.

Cheng, H., Liu, J., Wen, J., Nie, X., Xu, L., Chen, N., Li, Z., Wang, Q., Zheng, Z., Li, M., et al. (2019). Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. Genome Biol. **20**:1–16.

Chikmawati, T., Skovmand, B., and Gustafson, J.P. (2005). Phylogenetic relationships among *Secale* species revealed by amplified fragment length polymorphisms. Genome **48**:792–801.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) **6**:80–92.

Crespo-Herrera, L.A., Garkava-Gustavsson, L., and Åhman, I. (2017). A systematic review of rye (*Secale cereale* L.) as a source of resistance to pathogens and pests in wheat (*Triticum aestivum* L.). Hereditas **154**:14. https://doi.org/10.1186/s41065-017-0033-5.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156–2158.

Daskalova, N., and Spetsov, P. (2020). Taxonomic relationships and genetic variability of wild *Secale* L. species as a source for valued traits in rye, wheat and triticale breeding. Cytol. Genet. **54**:71–81.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. PLoS Genet. **9**:e1003905.

Fang, Z., Wu, B., and Ji, Y. (2021). The amino acid transporter *OsAAP4* contributes to rice tillering and grain yield by regulating neutral amino acid allocation through two splicing variants. Rice **14**:2.

Gholizadeh Sarcheshmeh, P., Mozafari, J., Saeidi Mehrvarz, S., and Shahmoradi, S. (2018). Genetic and ecogeographical diversity of rye (*Secale* L.) species growing in Iran, based on morphological traits and RAPD markers. Genet. Resour. Crop Evol. **65**:1953–1962.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., et al. (2010). A draft sequence of the Neandertal genome. Science **328**:710–722.

Grikpėdis, M., and Matuzevičiūtė, G.M. (2016). The beginnings of rye (*Secale cereale*) cultivation in the East Baltics. Veg. Hist. Archaeobot. **25**:601–610.

Guo, W., Xin, M., Wang, Z., Yao, Y., Hu, Z., Song, W., Yu, K., Chen, Y., Wang, X., Guan, P., et al. (2020). Origin and adaptation to high altitude of Tibetan semi-wild wheat. Nat. Commun. **11**:1–12.

Hoque, M.S., Masle, J., Udvardi, M.K., Ryan, P.R., and Upadhyaya, N.M. (2006). Over-expression of the rice *OsAMT1-1* gene increases ammonium uptake and content, but impairs growth and development of plants under high ammonium nutrition. Funct. Plant Biol. **33**:153–163.

Khush, G.S., and Stebbins, G.L. (1961). Cytogenetic and evolutionary studies in *Secale*. I. some new data on the ancestry of *S. cereale*. Am. J. Bot. **48**:723.

Larsson, P., Oliveira, H.R., Lundström, M., Hagenblad, J., Lagerås, P., and Leino, M.W. (2019). Population genetic structure in Fennoscandian landrace rye (*Secale cereale* L.) spanning 350 years. Genet. Resour. Crop Evol. **0**:1059–1071.

Letunic, I., and Bork, P. (2016). 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.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics **25**:1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; Subgroup, 1000 Genome Project Data Processing (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079.

Li, J., Jiang, J., Qian, Q., Xu, Y., Zhang, C., Xiao, J., Du, C., Luo, W., Zou, G., Chen, M., et al. (2011). Mutation of rice *BC12/GDD1*, which encodes a kinesin-like protein that binds to a GA biosynthesis gene promoter, leads to dwarfism with impaired cell elongation. Plant Cell **23**:628–640.

Li, G., Wang, L., Yang, J., He, H., Jin, H., Li, X., Ren, T., Ren, Z., Li, F., Han, X., et al. (2021). A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. Nat. Genet. **53**:574–584.

Martin, S.H., Davey, J.W., and Jiggins, C.D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol. Biol. Evol. **32**:244–257.

McElroy, J.S. (2014). Vavilovian mimicry: Nikolai Vavilov and his little-known impact on weed science. Weed Sci. **62**:207–216.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**:1297–1303.

Meier, S., Kunzmann, R., and Zeller, F.J. (1996). Genetic variation in germplasm accessions of *Secale vavilovii* Grossh. Genet. Resour. Crop Evol. **43**:91–96.

Milner, S.G., Jost, M., Taketa, S., Mazón, E.R., Himmelbach, A., Oppermann, M., Weise, S., Knüpffer, H., Basterrechea, M., König, P., et al. (2019). Genebank genomics highlights the diversity of a global barley collection. Nat. Genet. **51**:319–326.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37**:1530–1534.

Moskal, K., Kowalik, S., Podyma, W., Łapiński, B., and Boczkowska, M. (2021). The pros and cons of rye chromatin introgression into wheat genome. Agronomy **11**:456.

Ni, E., Deng, L., Chen, H., Lin, J., Ruan, J., Liu, Z., Zhuang, C., and Zhou, H. (2021). *OsCER1* regulates humidity-sensitive genic male sterility through very-long-chain (VLC) alkane metabolism of tryphine in rice. Funct. Plant Biol. **48**:461–468.

Pan, G., Zhang, X., Liu, K., Zhang, J., Wu, X., Zhu, J., and Tu, J. (2006). Map-based cloning of a novel rice cytochrome P450 gene *CYP81A6* that confers resistance to two different classes of herbicides. Plant Mol. Biol. **61**:933–943.

Parat, F., Schwertfirm, G., Rudolph, U., Miedaner, T., Korzun, V., Bauer, E., Schön, C.C., and Tellier, A. (2016). Geography and end use drive the diversification of worldwide winter rye populations. Mol. Ecol. **25**:500–514.

Patel, R.K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One **7**:e30619.

Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M., Azhaguvel, P., Sakuma, S., Dhanagond, S., Sharma, R., et al. (2015). Evolution of the grain dispersal system in barley. Cell **162**:527–539.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38**:904–909.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**:559–575.

Rabanus-Wallace, M.T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., Baez, M., Houben, A., Mayer, K.F.X., Guo, L., et al. (2021). Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. Nat. Genet. **53**:564–573.

Ranathunge, K., El-Kereamy, A., Gidda, S., Bi, Y.-M., and Rothstein, S.J. (2014). *AMT1;1* transgenic rice plants with enhanced NH4(+) permeability show superior growth and higher yield under optimal and suboptimal NH4(+) conditions. J. Exp. Bot. **65**:965–979.

Saika, H., Horita, J., Taguchi-Shiobara, F., Nonaka, S., Nishizawa-Yokoi, A., Iwakami, S., Hori, K., Matsumoto, T., Tanaka, T., Itoh, T., et al. (2014). A novel rice cytochrome P450 gene, *CYP72A31*, confers tolerance to acetolactate synthase-inhibiting herbicides in rice and Arabidopsis. Plant Physiol. **166**:1232–1240.

Sakamoto, S. (1982). The Middle East as a cradle for crops and weeds. In Biology and Ecology of Weeds (Dordrecht: Springer), pp. 97–109.

Schreiber, M., Himmelbach, A., Börner, A., and Mascher, M. (2019). Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. Evol. Appl. **12**:66–77.

Sencer, H.A., and Hawkes, J.G. (1980). On the origin of cultivated rye. Biol. J. Linn. Soc. **13**:299–313.

Shang, H., Chen, G., Hou, Y., Li, W., and Wei, Y. (2009). Analysis of main agronomic characters in *Secale*. J. Sichuan Agric. Univ. **27**:409–414.

Singh, R.J. (1977). Cross compatibility, meiotic pairing and fertility in 5 *Secal*e species and their interspecific hybrids. Cereal Res. Commun. **5**:67–75.

Skuza, L., Szućko, I., Filip, E., and Strzała, T. (2019). Genetic diversity and relationship between cultivated, weedy and wild rye species as revealed by chloroplast and mitochondrial DNA non-coding regions analysis. PLoS One **14**:1–21.

Stutz, H.C. (1957). A cytogenetic analysis of the hybrid *Secale cereale* L. x *Secale montanum* Guss. and its progeny. Genetics **42**:199–221.

Stutz, H.C. (1972). On the origin of cultivated rye. Am. J. Bot. **59**:59–70.

Tanaka, N., Matsuoka, M., Kitano, H., Asano, T., Kaku, H., and Komatsu, S. (2006). gid1, a gibberellin-insensitive dwarf mutant, shows altered regulation of probenazole-inducible protein (PBZ1) in response to cold stress and pathogen attack. Plant Cell Environ. **29**:619–631.

Vavilov, N.I. (1917). O proiskhozhdenii kulturnoi rzhi [On the origin of the cultivated rye]. Bull. Bur. Appl. Bot. **10**:561–590.

Vences, F.J., Vaquero, F., and Pérez de la Vega, M. (1987). Phylogenetic relationships in *Secale* (Poaceae): an isozymatic study. Plant Syst. Evol. **157**:33–47.

Wilde, P., and Miedaner, T. (2021). Hybrid rye breeding. In The Rye Genome, M.T. Rabanus-Wallace and N. Stein, eds. (Cham: Springer International Publishing), pp. 13–41.

Wuriyanghan, H., Zhang, B., Cao, W.-H., Ma, B., Lei, G., Liu, Y.-F., Wei, W., Wu, H.-J., Chen, L.-J., Chen, H.-W., et al. (2009). The ethylene receptor *ETR2* delays floral transition and affects starch accumulation in rice. Plant Cell **21**:1473–1494.

Yau, C.P., Wang, L., Yu, M., Zee, S.Y., and Yip, W.K. (2004). Differential expression of three genes encoding an ethylene receptor in rice during development, and in response to indole-3-acetic acid and silver ions. J. Exp. Bot. **55**:547–556.

Ye, C., and Fan, L. (2021). Orphan crops and their wild relatives in the genomic era. Mol. Plant **14**:27–39.

Ye, C.Y., Tang, W., Wu, D., Jia, L., Qiu, J., Chen, M., Mao, L., Lin, F., Xu, H., Yu, X., et al. (2019). Genomic evidence of human selection on Vavilovian mimicry. Nat. Ecol. Evol. **3**:1474–1482.

Zeng, X., Guo, Y., Xu, Q., Mascher, M., Guo, G., Li, S., Mao, L., Liu, Q., Xia, Z., Zhou, J., et al. (2018). Origin and evolution of qingke barley in Tibet. Nat. Commun. **9**:1–11.

Zhang, Y., Zhu, Y., Peng, Y., Yan, D., Li, Q., Wang, J., Wang, L., and He, Z. (2008a). Gibberellin homeostasis and plant height control by EUI and a role for gibberellin in root gravity responses in rice. Cell Res. **18**:412–421.

Zhang, D.-S., Liang, W.-Q., Yuan, Z., Li, N., Shi, J., Wang, J., Liu, Y.-M., Yu, W.-J., and Zhang, D.-B. (2008b). Tapetum degeneration retardation is critical for aliphatic metabolism and gene regulation during rice pollen development. Mol. Plant **1**:599–610.

Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., Xu, D., Chen, H., Wang, Y., Wang, Y. ge, et al. (2020). *Triticum* population sequencing provides insights into wheat adaptation. Nat. Genet. **52**:1412–1422.

Molecular Plant, Volume *15*

Supplemental information

Population genomic analysis reveals domestication of cultivated rye from weedy rye

Yanqing Sun, Enhui Shen, Yiyu Hu, Dongya Wu, Yu Feng, Sangting Lao, Chenfeng Dong, Tianyu Du, Wei Hua, Chu-Yu Ye, Jinhuan Zhu, Qian-Hao Zhu, Daguang Cai, Lidia Skuza, Jie Qiu, and Longjiang Fan

Supplemental information for “Population genomic analysis reveals domestication of cultivated rye from weedy rye” by Sun et al.

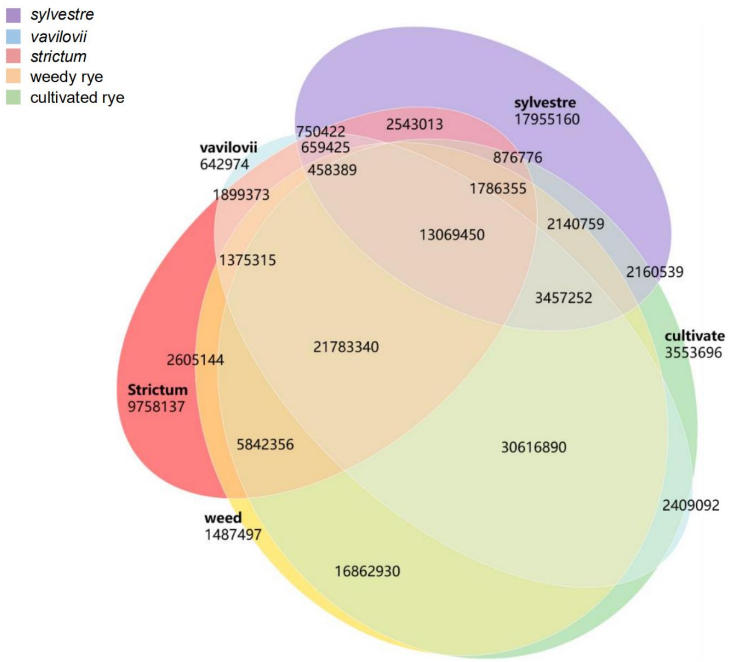

Figure S1. Venn diagram showing the population-specific and overlapping SNPs among different populations. A total of 2,409,092 and 1,487,497 SNPs were specific to cultivated and weedy ryes, respectively, and 95,559,323 SNPs (89.61% of the SNPs detected in the weedy population) were detected in both populations.

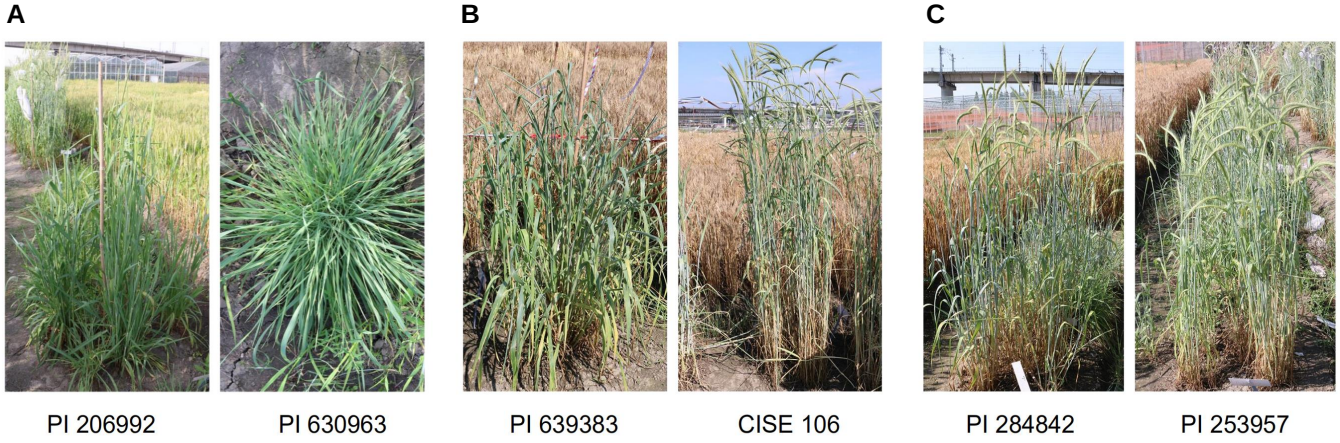

Figure S2. Phenotypes of the outliers and weedy-like *S. cereale* subsp. *vavilovii*. (A) Comparison of the phenotype (compact leaves and erect stem) of an accession classified as *S. strictum* (left) with the phenotype of a typical *S. strictum* accession (right). **(B)** Weedy-like cultivated accession (left) and the typical phenotype of a cultivated rye (right). **(C)** Phenotypes (larger leaves and relatively erect stems) of two accessions (left: PI 284842; right: PI 253957) of the weed-like *S. cereale* subsp. *vavilovii*.

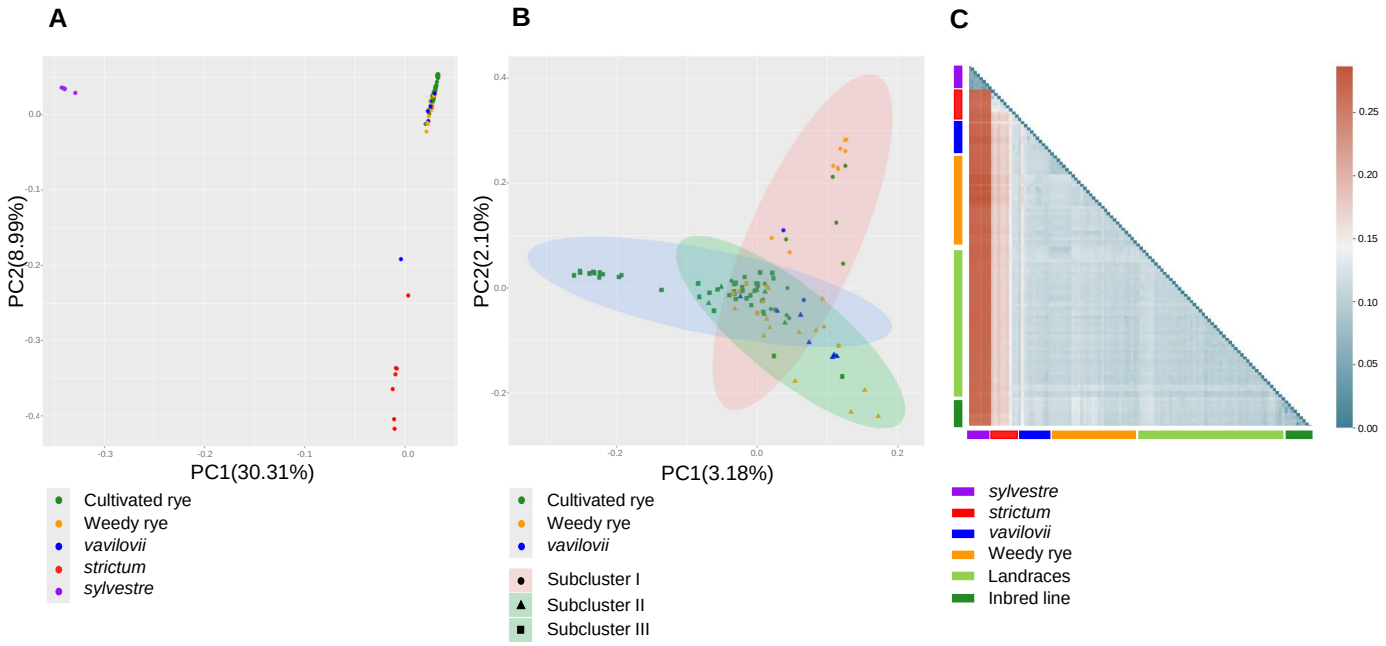

Figure S3. PCA and pairwise genetic distance of the 116 *Secale* accessions. (A) PCA plots of the 116 *Secale* accessions. The different rye populations are color-coded. **(B)** PCA plots of the 89 accessions in the subclusters I, II, and III that were clustered in phylogenetic tree. **(C)** Pairwise genetic distance of the 116 *Secale* accessions.

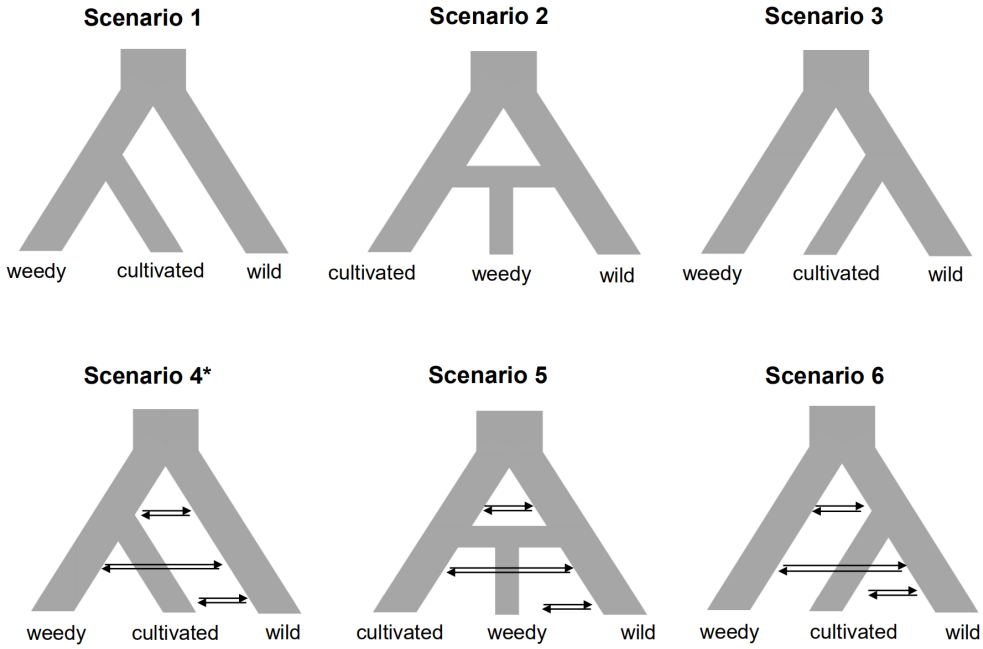

Figure S4. Six different demographic simulation models. Scenarios 1 and 4, cultivated rye originating from weedy rye without or with mass introgression; scenarios 2 and 5, weedy rye originating from hybrids of cultivated and wild rye without or with introgression; scenarios 3 and 6, cultivated rye originating from wild rye without or with introgression. "*" (Scenario 4) refers to the best model.

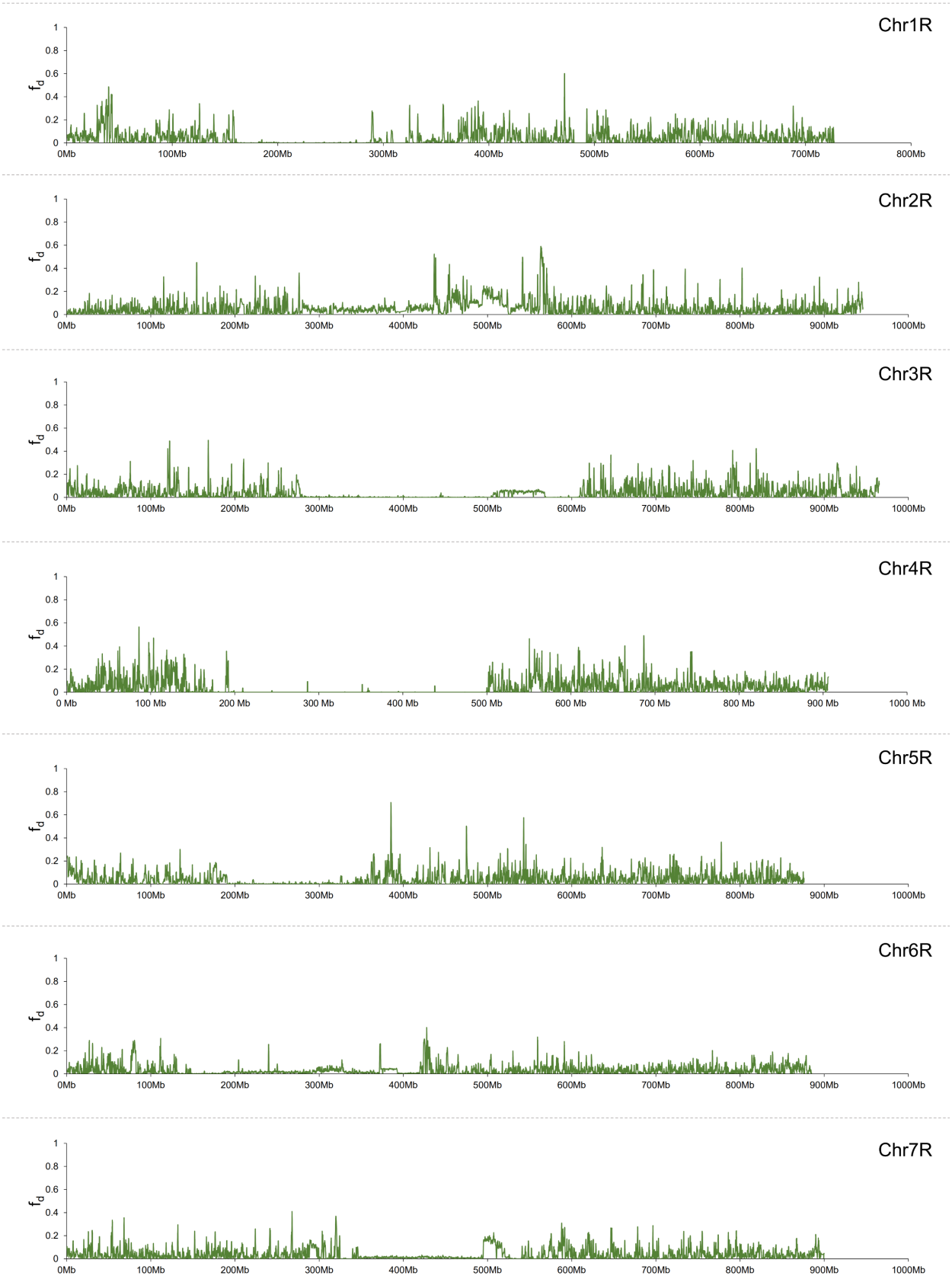

Figure S5. Introgression from *S.strictum* to cultivated rye across seven chromosomes. Gene flow from *S.strictum* to cultivated rye using weedy rye from Xinjiang as P1 as indicated by f_d.

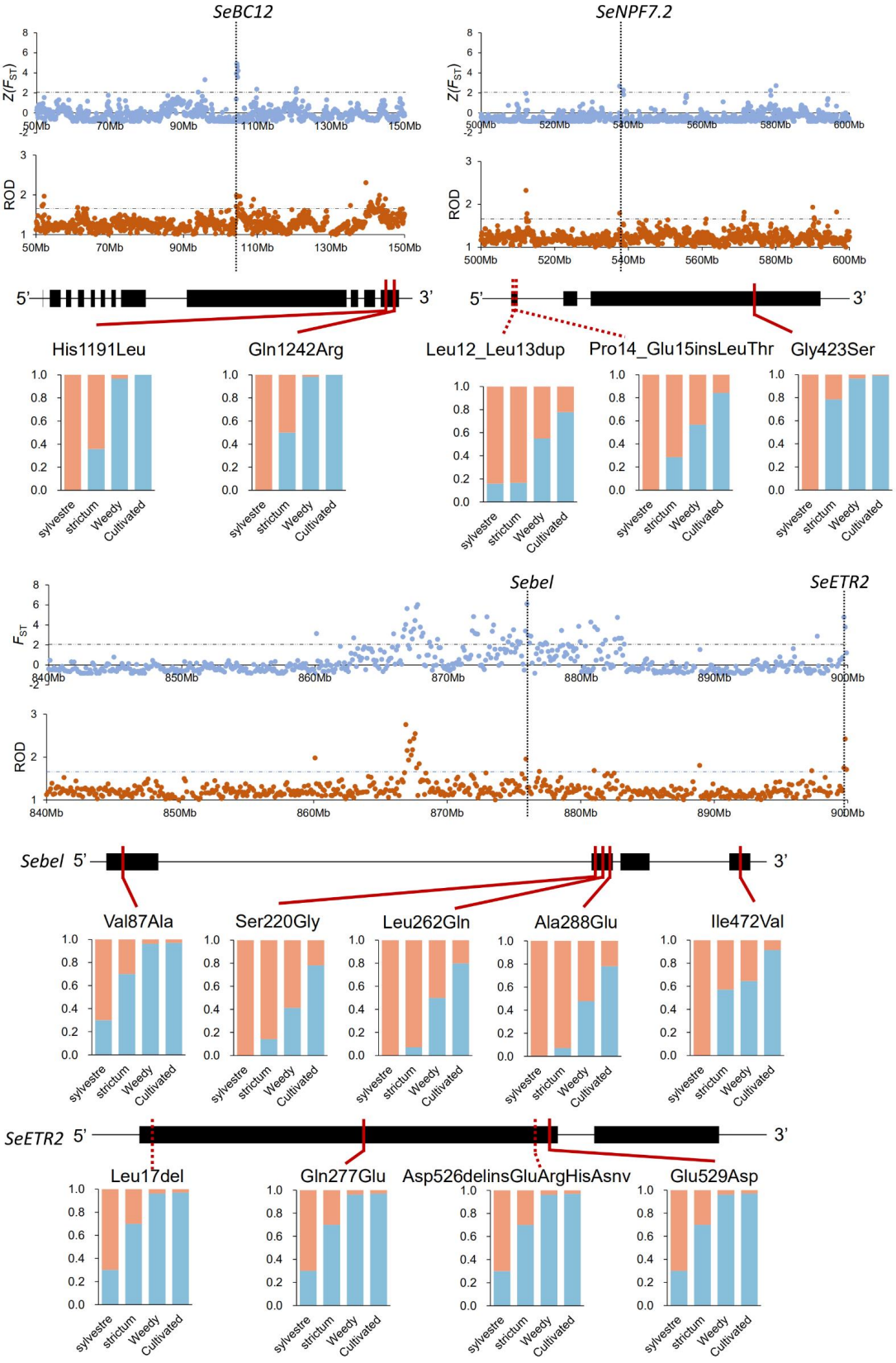

Figure S6. Four examples of candidate domestication genes. Gene structures are illustrated in black bars (exons) and black lines (introns). The graphs above the gene structure represent selection signals (Z-transformed F_{ST} and ROD) on the four candidate domestication genes *SeETR2*, *SeBC12*, *SeNPF7.2*, and *Sebel* between cultivated and weedy rye. The graphs below the gene structure represent the allele frequency of the non-synonymous variants on the top of each graph in the four rye populations.

Table S2 Quality control of SNPs and indels for each chromosome.

Chromosome	Raw SNPs	Filtration[1]		maf1mm2		Raw indel	Filtration[2]	
		Valid SNPs	Valid percentage	Valid SNPs	Valid percentage		Valid indels	Valid percentage
chr1R	59,485,348	34,102,213	57.33%	19,641,473	33.02%	3,284,632	2,878,014	87.62%
chr2R	73,340,679	42,624,409	58.12%	24,407,455	33.28%	3,938,037	3,447,458	87.54%
chr3R	75,446,742	43,541,074	57.71%	24,983,354	33.11%	3,888,519	3,399,971	87.44%
chr4R	70,701,729	40,194,949	56.85%	20,886,947	29.54%	3,885,142	3,401,608	87.55%
chr5R	67,513,909	37,209,993	55.11%	20,606,566	30.52%	3,724,018	3,267,886	87.75%
chr6R	67,007,036	38,793,875	57.90%	22,085,040	32.96%	3,639,470	3,180,250	87.38%
chr7R	69,279,747	39,589,250	57.14%	22,181,848	32.02%	3,705,525	3,253,085	87.79%
Total	482,775,190	276,055,763	57.18%	154,792,683	32.06%	26,065,343	22,828,272	87.58%

Filteration[1]: 1.quality >30; DP>10;QD>=2;MQ>=40;FS<=60;MQRankSum>=-12.5;ReadPosRankSum>=-8;SOR<=3

Filteration[2]: QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0

Table S3 Likelihood comparison of demographic models for divergence history among the three group.

Model name	# parameter	Delta likelihood	AIC	ΔAIC
Model 1	9	22940.261	2219092.99	40587.87
Model 2	9	22689.402	2217939.75	39434.63
Model 3	9	35516.862	2277010.38	98505.26
Model 4	17	14123.241	2178505.12	0
Model 5	17	15198.768	2183460.1	4954.98
Model 6	17	16544.055	2189653.38	11148.26

Table S4 Inferred parameters under the best fitted model.

Parameter	Point estimation	Median estimation	95% Confidence Interval	
			Lower bound	Upper bound
NCUL	1,130	1,443	1,116	3,239
NWILD	1,188	1,139	1,080	2,457
NWEED	2,159	3,186	1,517	6,007
NCUWE	10,424	4,830	151	10,653
NANC	8,920	17,943	8,771	28,373
T1	536	58	33	1,349
T2	7,306	5,999	1,569	7,470
$M_{WILD->CUL}$	7.69E-08	9.90E-06	1.45E-07	8.34E-05
$M_{CUL->WILD}$	8.34E-04	1.59E-03	1.32E-06	3.02E-03
$M_{WEED->WILD\ (0-T1)}$	4.23E-04	1.79E-05	1.59E-07	3.00E-03
$M_{WILD->WEED\ (0-T1)}$	9.92E-05	1.31E-04	2.60E-05	2.31E-04
$M_{CUL->WEED}$	1.35E-02	5.86E-03	5.23E-07	2.01E-02
$M_{WEED->CUL}$	8.14E-03	7.88E-05	1.06E-07	9.39E-03
$M_{WILD->CUWE\ (T1-T2)}$	4.01E-07	1.29E-06	4.35E-08	1.75E-03
$M_{CUWE->WILD\ (T1-T2)}$	3.15E-06	3.71E-04	3.56E-07	5.62E-04