CellPress
PARTNER JOURNAL

Molecular Plant
Research Article

Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence

Dezhi Wu[1,8], Zhe Liang[2,8], Tao Yan[1], Ying Xu[1], Lijie Xuan[1], Juan Tang[3], Gang Zhou[3], Ulrike Lohwasser[4], Shuijin Hua[5], Haoyi Wang[1], Xiaoyang Chen[7], Qian Wang[1], Le Zhu[1], Antony Maodzeka[1], Nazim Hussain[1], Zhilan Li[1], Xuming Li[3], Imran Haider Shamsi[1], Ghulam Jilani[6], Linde Wu[3], Hongkun Zheng[3], Guoping Zhang[1], Boulos Chalhoub[1], Lisha Shen[2,*], Hao Yu[2,*] and Lixi Jiang[1,*]

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

[2]Temasek Life Sciences Laboratory and Department of Biological Science, National University of Singapore, Singapore 117543, Singapore

[3]Biomarker Technologies Corporation, Beijing 101300, China

[4]Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research, 06466 Stadt Seeland, Germany

[5]Institute of Crop and Nuclear Agricultural Sciences, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

[6]Office of Research, Innovation & Commercialization, PMAS-Arid Agricultural University Rawalpindi, 46300 Rawalpindi, Pakistan

[7]Institute of Crop Science, Jinhua Academy of Agricultural Sciences, Jinhua 321017, China

[8]These authors contributed equally to this article.

***Correspondence: Lisha Shen (lisha@tll.org.sg), Hao Yu (dbsyuhao@nus.edu.sg), Lixi Jiang (jianglx@zju.edu.cn)**

https://doi.org/10.1016/j.molp.2018.11.007

ABSTRACT

Rapeseed (*Brassica napus*), an important oilseed crop, has adapted to diverse climate zones and latitudes by forming three main ecotype groups, namely winter, semi-winter, and spring types. However, genetic variations underlying the divergence of these ecotypes are largely unknown. Here, we report the global pattern of genetic polymorphisms in rapeseed determined by resequencing a worldwide collection of 991 germplasm accessions. A total of 5.56 and 5.53 million single-nucleotide polymorphisms (SNPs) as well as 1.86 and 1.92 million InDels were identified by mapping reads to the reference genomes of "Darmor-*bzh*" and "Tapidor," respectively. We generated a map of allelic drift paths that shows splits and mixtures of the main populations, and revealed an asymmetric evolution of the two subgenomes of *B. napus* by calculating the genetic diversity and linkage disequilibrium parameters. Selective-sweep analysis revealed genetic changes in genes orthologous to those regulating various aspects of plant development and response to stresses. A genome-wide association study identified SNPs in the promoter regions of *FLOWERING LOCUS T* and *FLOWERING LOCUS C* orthologs that corresponded to the different rapeseed ecotype groups. Our study provides important insights into the genomic footprints of rapeseed evolution and flowering-time divergence among three ecotype groups, and will facilitate screening of molecular markers for accelerating rapeseed breeding.

Key words: *Brassica napus*, genome resequencing, selective sweep, ecotype divergence, GWAS, flowering-time trait

Wu D., Liang Z., Yan T., Xu Y., Xuan L., Tang J., Zhou G., Lohwasser U., Hua S., Wang H., Chen X., Wang Q., Zhu L., Maodzeka A., Hussain N., Li Z., Li X., Shamsi I.H., Jilani G., Wu L., Zheng H., Zhang G., Chalhoub B., Shen L., Yu H., and Jiang L. (2019). Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence. Mol. Plant. **12**, 30–43.

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

INTRODUCTION

Rapeseed (*Brassica napus*) is an important source of edible oil and protein-rich livestock feed. *B. napus* (AACC) ancestrally originated from an interspecific hybridization between two diploid progenitors, *Brassica rapa* (AA) (*n* = 10) and *Brassica oleracea* (CC) (*n* = 9), less than 7500 years ago (Nagaharu, 1935; Chalhoub et al., 2014). Despite its shorter evolutionary history compared with that of its parental species, rapeseed has adapted to diverse climate zones and latitudes and formed three main ecotype groups, namely the winter, semi-winter, and spring types. Europe is the origin center of the ancient rapeseed cultivars. The identity of the species "Rapa" in the Middle Ages is not clear, and might refer to *B. rapa*, *B. oleracea*, or *B. napus*. Rapeseed (*B. napus*) could have been cultivated in Mediterranean areas as early as the Classical times (between the 8th century BC and the 6th century AD) by ancient Greeks and Romans. Medieval scholars recorded the preferential cultivation of *B. napus* in dry sandy soils in contrast to the cultivation of *B. rapa* in low and wet lands (Fussell, 1955). *B. napus* was introduced to China at the end of the 1940s, and successfully replaced the native turnip type (*B. rapa*) due to its vigorous vegetative growth and much higher yield. After intensive selection in China, rapeseed has become a semi-winter type with a biennial habit, but moderate vernalization requirement, that is acclimated to the mild winter and the multiple cropping system in the Yangtze River region (Liu, 2000). The prevalent type of rapeseed in Australia is the semi-winter type, which is sown before winter and has a mild vernalization requirement (Chen et al., 2008). The rapeseed with low erucic acid content in oil and low glucosinolate levels in seed meal known as "Canola" was first bred in Canada and then in Europe in the 1970s, and spread to other parts of the world as the gene source for "double-low" quality in breeding programs (Dupont et al., 1989).

To survive in a given environment, flowering at the right time is essential, and flowering time plays a critical role in determining the life-cycle period, yield, and seed quality in rapeseed. In the model plant *Arabidopsis*, multiple flowering pathways respond to environmental and developmental signals and converge at the transcriptional regulation of *FLOWERING LOCUS T* (*FT*) (Amasino, 2010; Srikanth and Schmid, 2011; Andrés and Coupland, 2012), which encodes a protein that acts as a mobile florigen signal moving from leaves to the shoot apical meristem to trigger flowering (Corbesier et al., 2007; Jaeger and Wigge, 2007). The expression of *FT* is activated by CONSTANS (CO) in the photoperiod pathway (Suarez-Lopez et al., 2001; An, 2004), and is repressed by FLOWERING LOCUS C (FLC) (Searle et al., 2006), whose mRNA expression is repressed by both the vernalization and autonomous pathways (Sheldon, 1999; Michaels and Amasino, 2001).

The domestication and breeding of crops gives rise to selective sweeps, the reduction or elimination of variation among the nucleotides neighboring a mutation in DNA (Maynard Smith and Haigh, 1974). Comparative population genomics allows the discovery of selective sweeps through the measurement of linkage disequilibrium (LD) to determine whether a given haplotype is over-represented in a population. The occurrence of strong LD indicates a selective sweep, in which the loci under selection can be identified (Slatkin, 2008). Performing genome-wide association studies (GWAS) based on single-nucleotide polymorphisms (SNPs) is an effective strategy for identifying candidate genes and/or quantitative loci that control important agronomic and quality traits in various field crops. The recent development of genome-sequencing technology has enabled the effective identification of SNPs and other genetic variations in a large genetic population (Zhou et al., 2015; Wang et al., 2017; Du et al., 2018). Although genetic variations among winter, semi-winter, and spring ecotypes of rapeseed have been reported (Schiessl et al., 2017a; Wei et al., 2017), the key allelic variations underlying the divergence of the different *B. napus* ecotypes are not fully understood.

In this study, we resequenced a worldwide collection of 991 *B. napus* germplasm accessions, including 658 winter types, 145 semi-winter types, and 188 spring types, from 39 countries. We identified the global pattern of genetic polymorphism in *B. napus* and revealed the paths of allelic drift showing the splits and mixtures of populations among the major origins. We also studied the selective sweeps produced during natural and artificial selection, and uncovered the genetic basis underlying the divergence of the main ecotypes. GWAS of the flowering-time trait identified SNPs in the promoter regions of *FT* and *FLC* orthologs, which specifically correspond to the three rapeseed ecotype groups. This study provides important insights into the genomic footprints of rapeseed evolution and flowering-time divergence among the three ecotype groups, and will facilitate screening of molecular markers for accelerating rapeseed breeding.

RESULTS

Resequencing 991 Accessions of *B. napus*

We generated a total of 7.82 Tb of clean reads from whole-genome resequencing of 991 *B. napus* accessions originating from 39 countries across the world (Supplemental Table 1) with an average of ~6.6-fold-coverage (Supplemental Table 2). After SNP and InDel calling, 5.56 million SNPs and 1.86 million InDels were detected by comparing the 991 rapeseed genomes with the reference genome ("Darmor-*bzh*") (Table 1). We examined the distribution of polymorphisms in *B. napus* genomic regions and found that 202 323, 62 834, 939 553, 899 982, and 2 778 189 SNPs were located within exons, introns, and upstream regions (within 5 kb upstream of transcription start sites), downstream regions (within 5 kb downstream of transcription stop sites), and intergenic regions, respectively (Supplemental Figure 1 and Supplemental Table 3). In coding regions, we annotated 79 018 nonsynonymous, 2992 splicing, 235 start-loss, 1413 stop-gain, and 476 stop-loss SNPs (Supplemental Figure 1 and Supplemental Table 3), which led to amino acid changes, splicing variants, longer transcripts, or premature stop codons. Moreover, 20 852, 38 625, 381 648, and 358 535 InDels were located within exons, introns, upstream regions, and downstream regions, respectively (Supplemental Figure 1 and Supplemental Table 4).

Among the 19 chromosomes of *B. napus*, the numbers of SNPs ranged from 155 121 (Chr.A10) to 400 436 (Chr.C03) (Supplemental Table 5). Comparison of the SNP distribution between the A and C subgenomes showed that the frequency of SNPs in the A subgenome (7.9 SNPs/kb) was higher than that in the C subgenome (5.9 SNPs/kb) (Supplemental Figure 2 and

Parameters			Groups		
			Spring	Semi-winter	Winter
	No. of accessions	991	188	145	658
Variation[a]	SNPs	5 559254	5 150 108	5 229 868	5 372 677
	Private SNPs	–	31 830	68 248	141 454
	InDels	1 858 671	1 740 570	1 776 121	1 824 101
	Private InDels	–	10 897	27 826	53 328
	π (10^{-4})	–	9.92	10.16	8.42
LD[b]	r^2 = (half_maxR^2)	0.30	0.31	0.32	0.30
	LD decay (kb; r^2 = half_maxR^2)	12.10	17.40	24.55	11.27
	LD decay (kb; r^2 = 0.1)	465.5	566.3	NA	384.5
	LD blocks	553 398	451 228	279 780	442 589
	<1 kb	544 441	435 378	230 860	440 520
	1–10 kb	8505	15 219	43 460	1969
	>10 kb	452	631	5460	100
			S versus W	SW versus W	S versus SW
Population[c] divergence	F_{ST}, top 1% threshold		0.48	0.52	0.45
	Regions		61	56	64
	Genes		1404	1228	1370

Table 1. Summary of Genomic Polymorphisms and Variants in Different Rapeseed Ecotype Groups.

NA, not available; S, spring; SW, semi-winter; W, winter.

[a]Private SNPs and InDels are variations specific to each group. π, nucleotide diversity within each group.

[b]LD blocks were defined using PLINK software (www.cog-genomics.org/plink2; v1.9).

[c]F_{ST} values were calculated for each 100-kb sliding window with a step size of 10 kb for each SNP among the populations. The regions and nonsynonymous SNPs with the top 1% of F_{ST} values were considered.

Supplemental Table 5). After filtering, we identified 2 753 575 high-confidence SNPs (missing data < 50%, minor allele frequency [MAF] > 5%) that were used for subsequent analyses.

Population Structure and Genomic Variation among the 991 Accessions

The 991 rapeseed (*B. napus*) germplasm accessions include 658 winter, 145 semi-winter, and 188 spring types (Supplemental Table 1), which correspond roughly to their geographical distributions in Europe, Asia, and the Eurasian continent/North America, respectively (Figure 1A). We performed phylogenetic and population structure analyses, and principal component analysis (PCA), and found that the 991 rapeseed accessions were clustered into three groups (K = 3), namely groups 1, 2, and 3 (Figure 1B–1D; Supplemental Figure 3). Based on population structure and PCA, groups 1, 2, and 3 contain 604, 174, and 213 accessions, respectively (Supplemental Figure 4A and Supplemental Table 6), and roughly correspond to the winter, semi-winter, and spring ecotype groups (Figure 1C). Principal component 1 (PC1), representing 11.19% of the total variation, distinguished the winter-type accessions from the semi-winter and spring types, while PC2, representing 6.90% of the total variation, distinguished the semi-winter types from the spring types (Figure 1D). The geographic distribution of rapeseed revealed by population structure and PCA (Figure 1B) was comparable with the original information regarding the country of origin and ecotype on the accession passports (Figure 1A), suggesting that the winter, semi-winter, and spring ecotypes are genetically different from each other. Nevertheless, there are a number of accessions (88; 8.9%) (Supplemental Table 6) with different classifications based on the two grouping systems. We analyzed these 88 accessions by performing further PCA. These accessions were not well separated into ecotypes by PC1 and PC2 (Supplemental Figure 4B), indicating the occurrence of genomic modifications along with their adaptation to different environments.

Our analysis uncovered 5 372 677, 5 229 868, and 5 150 108 SNPs, as well as 1 824 101, 1 776 121, and 1 740 570 InDels in the winter, semi-winter, and spring types, respectively, as compared with the reference genome "Darmor-*bzh*" (Table 1). The nucleotide diversity (π) of winter types (8.42×10^{-4}) was lower than that of semi-winter ($\pi = 10.16 \times 10^{-4}$) and spring ($\pi = 9.92 \times 10^{-4}$) types (Table 1 and Supplemental Figure 5). The existence of low-frequency SNPs in winter types is mainly because mapping was performed against "Darmor-*bzh*," which is also a winter-type accession. We further analyzed LD (indicated by r^2) throughout the genome. The physical distances between SNPs (half of the maximum value) were 11.27 kb (r^2 = 0.30), 24.55 kb (r^2 = 0.32), and 17.40 kb (r^2 = 0.31) for the winter, semi-winter, and spring ecotypes, respectively (Figure 1E and Table 1). The magnitude of LD decay varied drastically among different chromosomes and different subgenomes (Supplemental Figure 6). 442 589, 279 780, and 451 228 LD blocks were detected in winter-, semi-winter-, and spring-type genomes, respectively. Furthermore, a total of 100, 5460, and 631 large LD structures (>10 kb) were

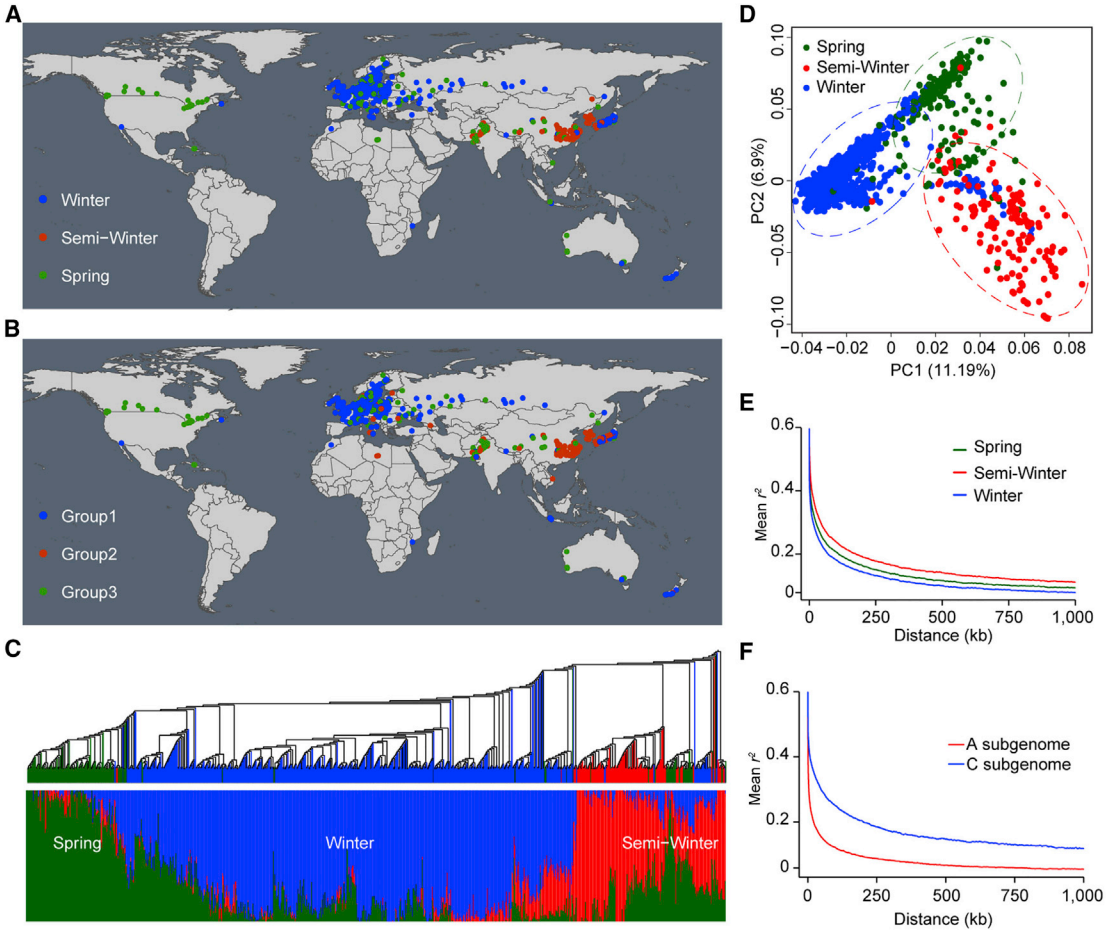

Figure 1. Distribution, Population Structure, PCA, and LD Decay of the 991 Rapeseed Germplasm Accessions.

(A) The geographic distribution of the 991 rapeseed accessions based on country of origin and ecotype information.

(B) The geographic distribution of the 991 rapeseed accessions based on population structure and PCA.

(C) Population structure analysis using 293 498 SNPs (missing data < 50%, MAF > 5%, r^2 < 0.2). The genetic groups, namely groups 1, 2, and 3, roughly correspond to the winter, semi-winter, and spring ecotype groups, but these classifications do not overlap for 88 (8.9%) accessions.

(D) PCA plot of the first two components (PC1 and PC2) of the 991 accessions. PC1, representing 11.19% of the total variation, separates the winter-type accessions from the semi-winter- and spring-type accessions, whereas PC2, representing 6.90% of the total variation, distinguishes the semi-winter types from the spring types.

(E) Genome-wide average LD decay in the winter, semi-winter, and spring ecotypes. The green, red, and blue curves display the rate of LD decay over distance (kb) in the spring, semi-winter, and winter ecotypes, respectively.

(F) Genome-wide average LD decay for the two subgenomes of *B. napus*. The blue and red curves show the rate of LD decay over distance (kb) in the C and A subgenomes, respectively.

detected in the winter-, semi-winter-, and spring-type genomes (Table 1), respectively, indicating a higher frequency of genetic recombination, which breaks the genomic LD structure, in winter-type rapeseed than in the other types (Supplemental Figure 7). Moreover, the LD decay rate was faster in the A subgenome than in the C subgenome (Figure 1F and Supplemental Figure 6), suggesting a higher frequency of genetic recombination in the A subgenome.

Paths of Allelic Drift among the Major Sites of Origin

In order to understand population splits and mixtures among the major origin sites, we calculated drift parameters in light of the allelic variations across the 991 genomes, and drew trees showing the relationships within the winter-type (Figure 2A) and spring-type (Figure 2B) germplasm accessions. After further calculating the number of significant gene-flow paths on the world map, we found four major paths each within the winter-type and spring-type rapeseed accessions (Figure 2C). The direction of gene flow for the winter-type rapeseed accessions was from France to Russia (weight = 0.5), France to Sweden via Germany (weight = 0.3), Poland to Japan (weight = 0.3), and Japan to China (weight = 0.3), while spring-type rapeseed spread from France to Russia (weight = 0.5), France to Sweden via Germany (weigh = 0.4), Sweden to Canada (weight = 0.2), and Canada to China (weight = 0.5). Notably, there were no significant inter-ecotype drift events.

Selective Signals Produced by Natural and Artificial Selection in *B. napus*

To determine the gene modifications occurring along the history of natural and artificial selection, we analyzed selective sweeps

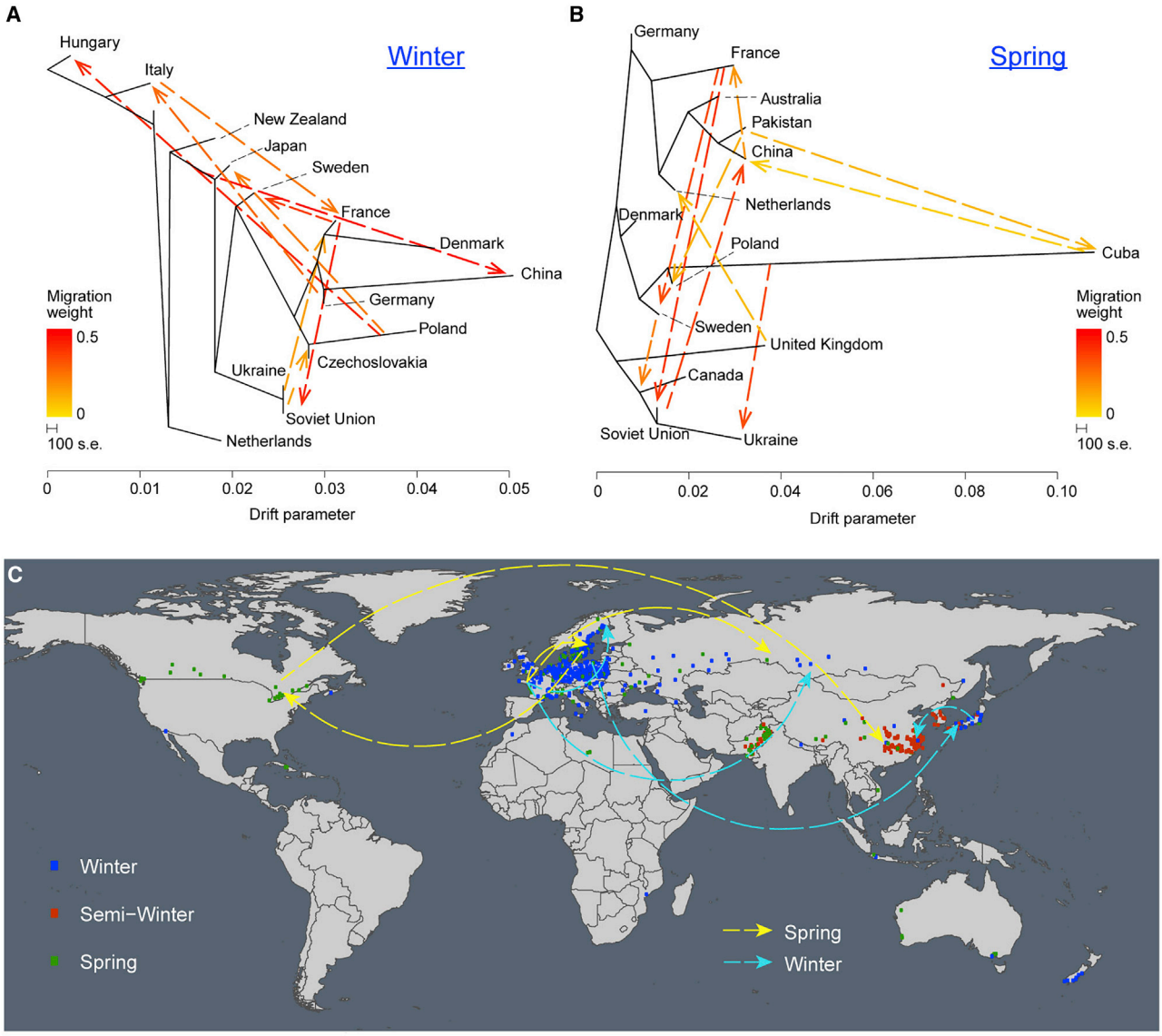

Figure 2. Allelic-Drift Paths Showing Population Splits and Mixtures among the Major Sites of Origin of the Accessions.

(A and B) Inferred rapeseed tree showing the major sites of origin of the winter-type **(A)** and spring-type **(B)** germplasm accessions and the directions of gene-flow in each population. Arrows indicate the direction of gene flow, while the line colors represent the migration weight based on sample number. Horizontal branch length is proportional to the amount of genetic drift that has occurred on the branch. Scale bars represent a 100-fold average SE for the entries in the sample covariance matrix.

(C) Gene-flow paths visualized on the world map. The yellow and lake-blue arrows show the direction of gene flow in the spring- and winter-type populations, respectively. The blue, red, and green dots show the distribution of the winter, semi-winter, and spring accessions on the world map.

among the three ecotype groups (Figures 3A and 4). We scanned all genomic regions in 100-kb sliding windows (a step of 10 kb), and defined the regions containing nonsynonymous SNPs with the top 1% of population fixation statistics (F_{ST}) values as significant different windows. There were 61, 56, and 64 selective sweeps unveiled between spring and winter types, semi-winter and winter types, and spring and semi-winter types, respectively (Table 1 and Figure 4). The genes located in these selective-sweep regions could potentially play important roles. In particular, the strongest signal of a selective sweep was found on Chr.A10 between the winter and spring types (Figure 3A). There were 56 annotated genes in this region, including an *FLC* ortholog (*BnaA10g22080D*) and an *AIL6* ortholog (*BnaA10g21750D*) encoding an AP2-like ethylene-responsive transcription factor (ERF). Another strong selective signal was found in a large region on Chr.A02 containing 377 annotated genes, including orthologs of *FT* (*BnaA02g12130D*), *FY* (*BnaA02g01670D*), *CONSTANS*-like (*BnaA02g13920D*), *ERF118* (*BnaA02g13950D*), and *ERF118*-like (*BnaA02g13960D*), and the histidine transporter-like 2 ethylene transporter (*BnaA02g13290D*) (Figure 3A; Supplemental Tables 7 and 8). Some regions on Chr.C01, C02, C08, and C09 that encode several MADS-box genes displayed weak selective-sweep signals (Figure 3A and Supplemental Table 8). In total we found 1404 genes in the selective-sweep regions between the winter and spring types, which included genes involved in flowering-time control, ethylene biosynthesis and

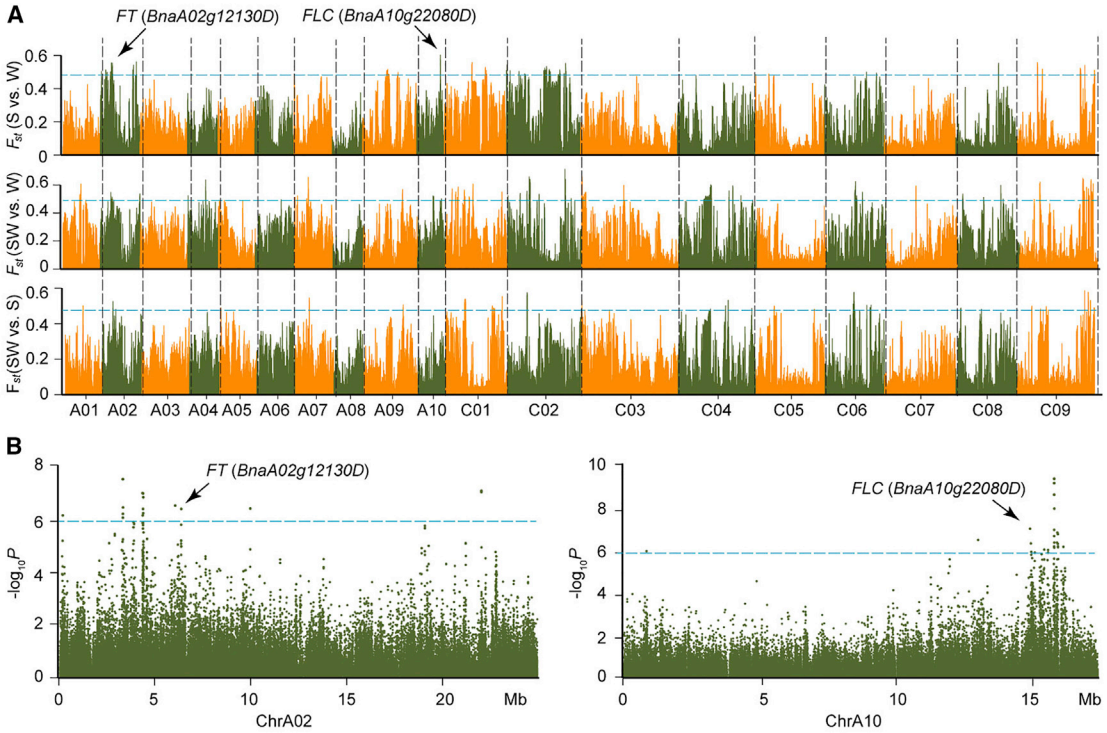

Figure 3. Selective-Sweep Signals between the Three *B. napus* Ecotype Groups.

(A) Genome-wide distribution of selective-sweep signals identified through comparisons between spring and semi-winter ecotypes (upper panel), semi-winter and winter ecotypes (middle panel), and spring and winter ecotypes (lower panel). The blue dashed lines represent the thresholds (top 1% of F_{ST} values).

(B) Manhattan plots of GWAS results for the flowering-time trait on Chr.A02 and Chr.A10. The identified SNPs were in located the regions from 4 to 7 Mb on Chr.A02 (left) and the regions from 14 to 17 Mb on Chr.A10 (right). The blue dashed lines indicate the significance threshold ($-\log_{10}P = 6.0$).

signaling, and plant response to stresses (Supplemental Figure 8 and Supplemental Table 8). Moreover, a total of 1228 and 1370 genes were identified within the sweep regions underlying the population divergence between the semi-winter and winter types, and spring and semi-winter types, respectively (Supplemental Figure 8). Similarly, we identified many genes relevant to either flowering-time regulation, such as *CONSTANS*-like and MADS-box genes, or phytohormone signaling, such as ERFs, as well as auxin- and gibberellin-induced genes (Supplemental Tables 9 and 10). Gene ontology analysis revealed that these genes are involved in multiple biological process, cellular component, and molecular function (Supplemental Figures 9–11).

Identification of Genes Associated with Flowering Time in *B. napus*

Flowering time is one of the most important traits associated with crop domestication and breeding, and also largely distinguishes the three ecotype groups of *B. napus*. As expected, we observed variation in flowering time among different accessions, ranging from 125 to 191 days (from sowing to flowering), with the latest flowering mainly exhibited by the winter-type accessions (Supplemental Table 1). Since the strongest selection signals were detected in regions on Chr.A02 and Chr.A10 containing a variety of flowering time genes (Figure 3A), we performed GWAS for flowering time with a total of 2 753 575 SNPs across the 991 accessions to narrow down these candidate genes (Figure 3B and Supplemental Figure 12). The Manhattan plot showed that many SNPs in the 4–7 Mb region of Chr.A02 were significantly associated with flowering time ($-\log_{10}P > 6$) (Figures 3B and 5). Interestingly, an ortholog of *Arabidopsis FT* (*BnaA02g12130D*) was identified within the 6.37–6.38 Mb region (Supplemental Figure 12 and Supplemental Table 11), where SNPs were strongly associated with flowering time (Figure 5). Further LD analysis of this 100-kb (6.37–6.38 Mb) region revealed that the SNPs associated with flowering time were located at ~3 kb (6 372 995–6 375 936 kb) upstream of the start codon of the *FT* ortholog (Figure 5A and 5B; Supplemental Tables 12 and 13).

Similarly, *BnaA10g22080D*, a putative ortholog of the *Arabidopsis FLC* gene, was identified within the region between 14.99 and 15.01 Mb on Chr.A10 (Figure 3B and Supplemental Table 11). Both GWAS and LD analyses of the ~100-kb sequences surrounding *BnaA10g22080D* (~50 kb upstream and ~50 kb downstream of the *FLC* ortholog) revealed that the SNPs in the 5′ regulatory region (14 995 616–14 998 616 kb; 3 kb upstream of the start codon of the *FLC* ortholog) were significantly linked with flowering-time variation among all accessions (Figure 6A and 6B; Supplemental Tables 14 and 15).

Analyses Using the SNP Set Generated by Mapping Reads to the "Tapidor" Genome

To verify the reliability of the above results (i.e., the analyses of population structure, selective signals, and GWAS of flowering time), we called SNPs by mapping reads to another reference genome, "Tapidor" (Bayer et al., 2017). We compared the

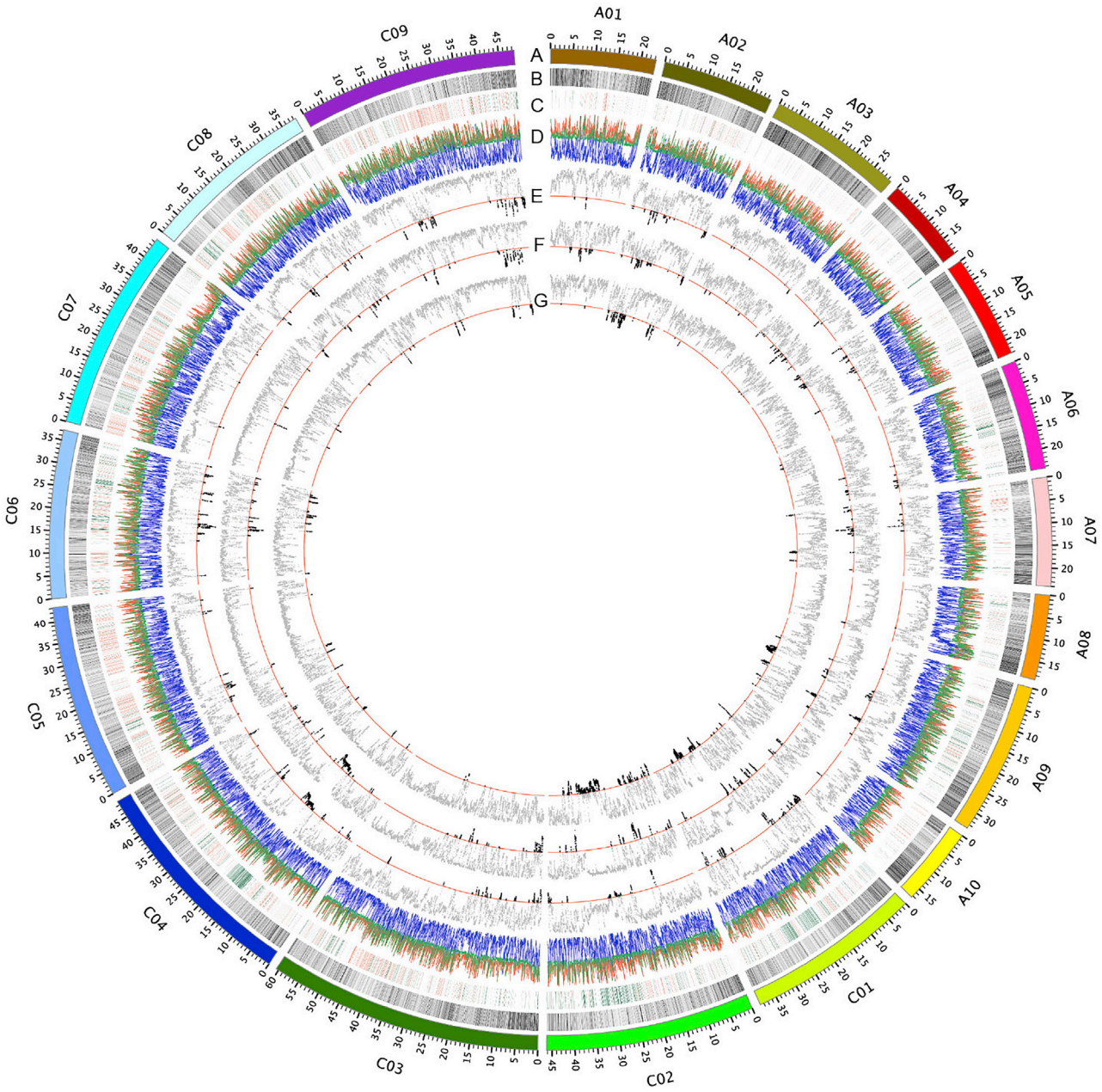

Figure 4. Circos Plot Showing Genetic Diversity and Selective Signals among Three Ecotype Groups.

(A) Sizes of the 19 chromosomes of *B. napus*.

(B) Heatmap view of SNP density for each chromosome.

(C) Distribution of weak LD regions (green) and strong LD regions (red) in 10-kb windows across the 19 chromosomes.

(D) Comparison of genetic diversity (π) among winter, semi-winter, and spring ecotypes.

(E) Scatter-spot distribution of F_{ST} in spring versus winter ecotypes.

(F) Scatter-spot distribution of F_{ST} in semi-winter versus winter ecotypes.

(G) Scatter-spot distribution of F_{ST} in semi-winter versus spring ecotypes. The top 1% of F_{ST} values above the red threshold lines are indicated by black spots inside the circle.

percent coverage of reads on the reference genomes obtained using the two mapping approaches (Supplemental Table 2). The average coverage depths were 6.6- and 6.8-fold when mapping against "Darmor-*bzh*" and "Tapidor," respectively. A total of 5.53 million SNPs and 1.92 million InDels were obtained by mapping the reads to the "Tapidor" genome (Supplemental Table 16). Based on population structure and PCA, groups 1, 2, and 3 contained 604, 174, and 213 accessions, respectively, which is identical to the result obtained using SNPs identified from mapping to the "Darmor-*bzh*" reference genome (abbreviated as SNP-Darmor) (Supplemental Figure 13A and Supplemental Table 6). We also compared the LD decay rates between "SNP-Darmor" and "SNP-Tapidor." The LD decay rates calculated using both SNP sets were faster in winter types

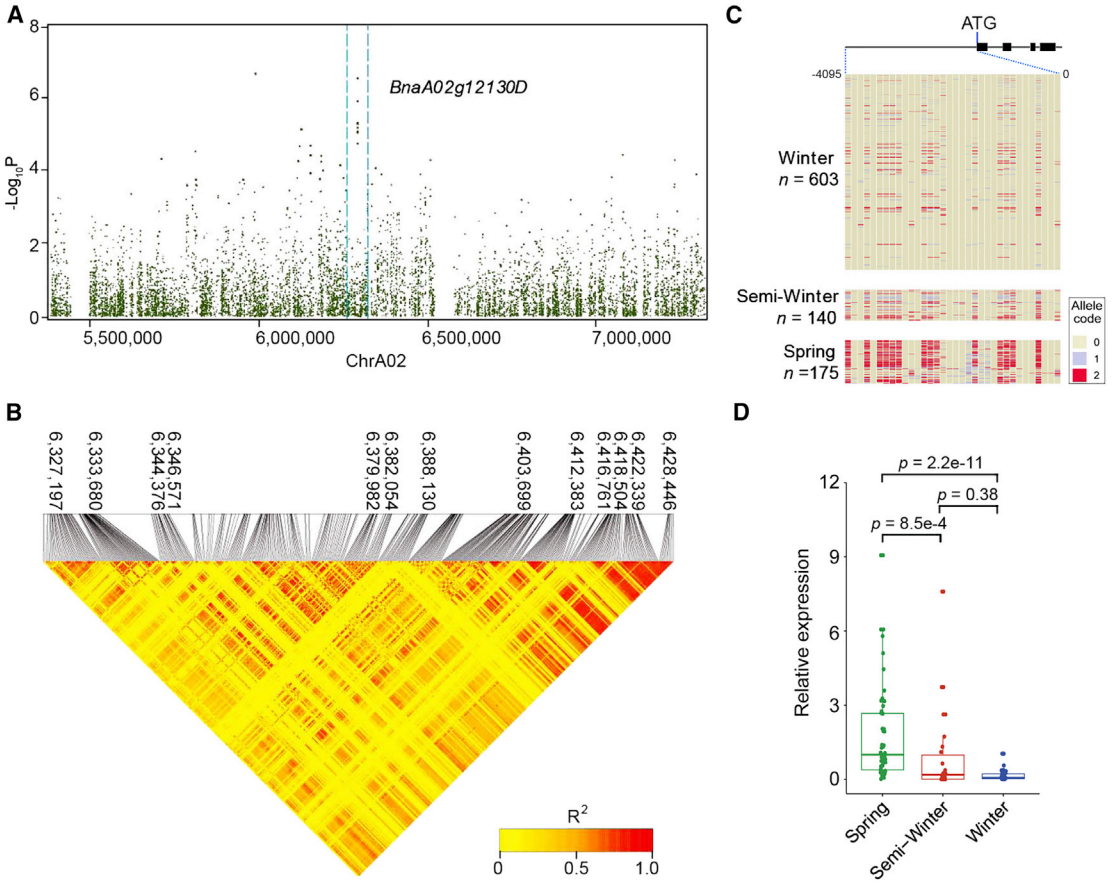

Figure 5. The SNPs Responsible for Flowering-Time Variation in the Region from ~6.37 to ~6.38 Mb on Chr.A02.

(A) Local Manhattan plot showing the ~100-kb region surrounding the *FT* ortholog (*BnaA02g12130D*). The spots above the threshold ($-\log_{10}P = 6.0$) in the candidate region (between the two blue vertical dotted lines) are SNPs significantly associated with flowering-time variation. Statistical analysis was performed with the two-tailed chi-square test.

(B) LD heat map of the ~100-kb *FT* region. The color key indicates r^2 values.

(C) Comparison of conserved SNPs specific to the three ecotype groups in the 0–3 kb region upstream of the start codon of the *FT* ortholog (*BnaA02g12130D*). The curcuminoid, blue, and red colors indicate SNPs homozygous for the reference genotype (0), heterozygous SNPs (1), and SNPs homozygous for the nonreference allele (2), respectively.

(D) Expression levels of the *FT* ortholog (*BnaA02g12130D*) in different ecotypes revealed by qRT–PCR analysis. *BnaACTIN7* was included as an internal reference, and the mid-value of *BnaA02g12130D* expression in the spring cultivars was set to 1. *p* values indicate the significance of pairwise comparisons.

than in semi-winter and spring types, and faster in the A subgenome than in the C subgenome (Supplemental Figure 13B and 13C). Selective-sweep analysis and GWAS for flowering time confirmed that the *FT* ortholog on A02 (*BnaA02g12130D*) and the *FLC* ortholog on A10 (*BnaA10g22080D*) were significantly associated with flowering-time variation (Supplemental Figure 13D and 13E). Taken together, analyses of both SNP sets using population genetic approaches, such as PCA, LD rate, selective signals, and GWAS, led to the same conclusions.

Identification of SNPs Corresponding to Ecotype Divergence

To find out whether the SNPs in the putative promoter regions of *FT* and *FLC* orthologs correspond to the divergence of the three rapeseed ecotype groups, we compared the SNPs in a 6136-bp region from 6 372 895 to 6 379 031 bp on Chr.A02 that contains the coding and upstream sequences of the putative *FT* ortholog (*BnaA02g12130D*), and the SNPs in a 7335-bp region from 14 995 720 to 15 003 055 bp on Chr.A10 that contains the coding and upstream regions of the putative *FLC* ortholog (*BnaA10g22080D*), across the genomes of the three rapeseed ecotype groups. Notably, nucleotide sequences in the coding regions of both the *FT* and *FLC* orthologs were quite conserved in all three ecotype groups (Supplemental Tables 12–15). However, 10 out of 32 SNPs in the 5′ upstream region of *BnaA02g12130D* were different among the three groups. In particular, 95% ± 2% of the winter-type accessions had the same nucleotide as the reference ("Darmor-*bzh*") at the positions 2999, 1267, 918, 627, and 490 bp upstream of the *BnaA02g12130D* start codon (Supplemental Table 12), whereas in 85% ± 6% of the spring-type accessions, the nucleotides at these sites were altered in different ways, such as conversion from guanine (G) to cytosine (C) and from adenine (A) to G. Only 22% ± 3% of the semi-winter-type accessions shared the same changes as the spring-type accessions at these sites except at the site 490-bp

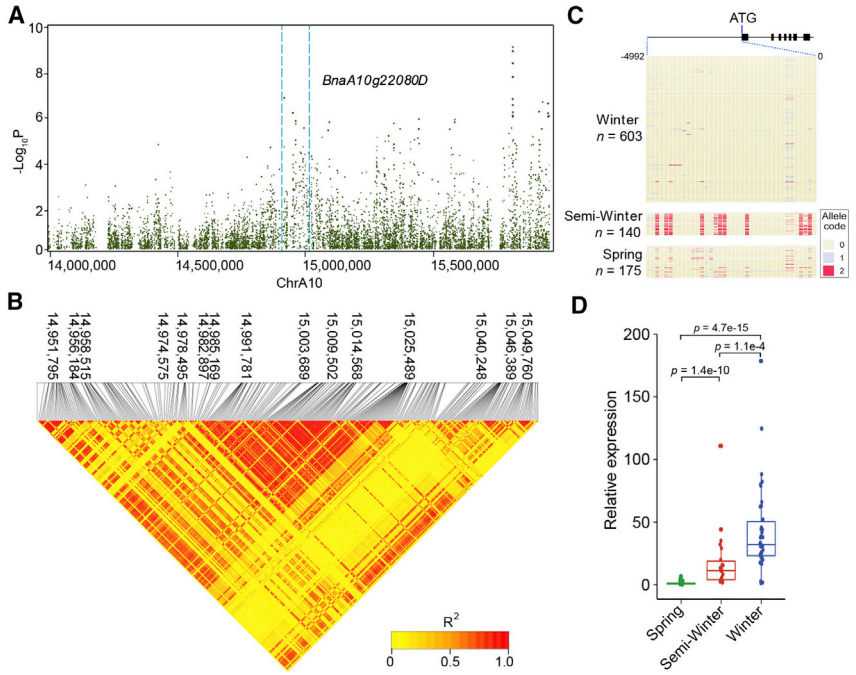

Figure 6. The SNPs Responsible for Flowering-Time Variation in the Region from ~14.99 to ~15.00 Mb on Chromosome A10.

(A) Local Manhattan plot showing the ~100-kb region surrounding the *FLC* ortholog (*BnaA10g22080D*). The spots above the threshold ($-\log_{10}P > 6.0$) in the candidate region (between the two blue vertical dotted lines) are significant SNPs associated with flowering-time variation. Statistical analysis was performed with the two-tailed chi-square test.

(B) LD heat map of the ~100-kb *FLC* region. The color key indicates r^2 values.

(C) Comparison of conserved SNPs specific to the three ecotype groups in the 0–5 kb region upstream of the start codon of the *FLC* ortholog (*BnaA10g22080D*). The curcuminoid, blue, and red colors indicate SNPs homozygous for the reference allele (0), heterozygous SNPs (1), and SNPs homozygous for the nonreference allele (2), respectively.

(D) Expression levels of the *FLC* ortholog (*BnaA10g22080D*) in different ecotypes revealed by qRT–PCR analysis. *BnaACTIN7* was included as an internal reference, and the mid-value of *BnaA02g12130D* expression in the spring cultivars was set to 1. *p* values indicate the significance of the pairwise comparisons.

upstream, where 40% of the semi-winter-type accessions displayed the nucleotide change from T to C observed in most of the spring-type accessions (Figure 5C). Similarly, the frequencies of various SNPs in the putative promoter region of the *FLC* ortholog (*BnaA10g22080D*) were very different among the three ecotypes (Figure 6C and Supplemental Table 14).

To investigate whether the ecotype-specific SNPs were associated with *FT* or *FLC* transcription in *B. napus*, we compared the expression levels of various *FT* and *FLC* orthologs in the leaves of 52 spring, 39 winter, and 21 semi-winter types out of the 991 accessions by performing qRT-PCR using primers specific to *FT* (*BnaA02g12130D*) and *FLC* (*BnaA02g12130D*) (Supplemental Figure 14 and Supplemental Table 17) and using published transcriptome data (Harper et al., 2012; Havlickova et al., 2018) reported for 124 out of the 991 accessions (Supplemental Tables 18 and 19). Our results and the published transcriptome data both showed that the expression levels of *BnaA02g12130D* were significantly higher in the leaves of spring-type accessions than in those of the other two ecotype groups (Figure 5D and Supplemental Figure 15A), whereas the expression of *BnaA10g22080D* was the lowest in the leaves of spring-type accessions (Figure 6D and Supplemental Figure 15B). In contrast to the significant differences in the expression of both *FT* and *FLC* orthologs between spring- and winter-type accessions, the differences in expression between semi-winter-type accessions and the other two types of accessions were relatively small (Figures 5D and 6D). These results indicate that the ecotype-specific SNPs are associated with the expression levels of *FT* and *FLC* orthologs, and could contribute to the differences in flowering time observed among the three ecotype groups.

Furthermore, we analyzed the patterns of SNPs in the putative genes regulating ethylene biosynthesis and signaling. There were 12–15 putative ethylene genes each in the selective-sweep regions between the spring and winter, spring and semi-winter, and semi-winter and winter types. However, only the putative *ERF119* ortholog on Chr.A02 and the putative *AP2-like ERF* (*AIL6*) ortholog on Chr.A10 displayed SNPs corresponding to ecotype divergence (Supplemental Figure 16A and 16C). Moreover, these ecotype-specific SNPs were associated with the transcription levels of *ERF119* and *AP2-like ERF* (*AIL6*) in the published transcriptome data (Supplemental Figure 16B and 16D) (Harper et al., 2012; Havlickova et al., 2018), which could contribute to the differences in vegetative growth period displayed by the three ecotype groups.

DISCUSSION

In this study, we resequenced and characterized the genetic diversities of 991 natural accessions of rapeseed, most of which are from major rapeseed production regions (i.e., European countries, China, Canada, and Australia). Population-genetics analyses of the two sets of SNPs generated by mapping reads to two different reference genomes, the "Darmor-*bzh*" and "Tapidor" genomes, led to consistent conclusions.

Based on population structure and PCA, the 991 resequenced rapeseed accessions were clustered into three groups (Figure 1B and 1D), which roughly correspond to the three ecotype groups, namely the winter, semi-winter, and spring rapeseed types (Figure 1C). *B. napus* is derived from natural hybridization between *B. rapa* (AA) and *B. oleracea* (CC) (Chalhoub et al., 2014). In general, there were higher densities of SNPs and InDels in the A subgenome than in the C subgenome of *B. napus*, suggesting a higher degree of genetic diversity in the A subgenome. Consistent with this, the faster LD decay rate in the A subgenome than in the C subgenome (Figure 1F) indicates a higher frequency of genetic recombination in the A subgenome.

The higher degree of genetic diversity in the A subgenome than in the C genome could be due to more frequent outcrossing between *B. napus* and *B. rapa* than between *B. napus* and *B. oleracea*. The A genome of a modern *B. napus* plant could be more diverse because it is more likely to be subjected to the introduction of new diversity from *B. rapa*, whereas C genome diversity is restricted more or less to the few donors of the original allopolyploidization events that created the species. Moreover, there are different levels of genetic diversity in natural populations of *B. rapa* and *B. oleracea*. Calculation of the extent of LD in broadly representative populations of *B. rapa* and *B. oleracea* has revealed that the distance required for LD to decay to half of its maximum value is much smaller in the *B. rapa* population (2.9 kb) than in the *B. oleracea* population (36.8 kb) (Cheng et al., 2016). Overall, rapeseed breeding suffers from low genetic diversity, especially for the Canola type (Dupont et al., 1989). Artificially resynthesized rapeseed has been a solution for introducing genetic variation (Becker et al., 1995; Cleemput and Becker, 2012; Girke et al., 2012). Given our understanding of the asymmetric evolution of rapeseed subgenomes, selection of appropriate cabbage parents with a relatively wild C genome is of particular interest for increasing genetic variation in rapeseed.

By calculating the drift parameters on the basis of allelic variation across the 991 genomes, we have generated a world map of the major gene-flow paths (Figure 2C). At least three major gene-flow paths are in line with the known breeding history of rapeseed. Firstly, the flow of winter-type rapeseed genes from France to Russia (weight = 0.5) and from France to Sweden via Germany (weight = 0.3) is supported by early documentation that rapeseed was first cultivated in Mediterranean areas in Classical times, and then likely introduced to other parts of Europe by the Romans (Fussell, 1955). Secondly, the "Japan–China" winter rapeseed flow path is consistent with a report that the first popular winter-type rapeseed variety in China was introduced from Japan in the 1930s, and was renamed Shengliyoucai in the 1950s (Wang et al., 2017). Thirdly, the "Canada–China" spring rapeseed flow path is supported by the fact that Chinese breeders popularly used Canadian cultivars such as "Oro," "Midas," "Tower," and "Wester" as "double-low quality" donors to breed their own Canola-type rapeseed in breeding programs sponsored by International Development Research Center in Canada in the early 1980s. However, the drift-path calculation relies on correct information about the origin of each accession in the passport. We cannot exclude the possibility that the country of origin is actually the country of the donor, but not the true origin. However, such a problem may not considerably change the general conclusions of the present study.

There are three rapeseed ecotype groups that have been formed by natural and artificial selection in various eco-environments. Adaptation and breeding processes gave rise to numerous selective sweeps because some alleles were under strong selection, which caused a large reduction in nearby genetic variation (Maynard Smith and Haigh, 1974; Rubin et al., 2010). Selective-sweep analyses have unveiled 61, 56, and 64 selective sweeps between spring and winter types, semi-winter and winter types, and spring and semi-winter types, respectively (Figure 3 and Table 1). The spring and winter rapeseed accessions have an annual and biennial growth habit, respectively, and their life cycles differ by a period of approximately 90–150 days. The adaptation from winter to spring environments (and vice versa) resulted in allelic changes at 1404 loci located in 61 selective-sweep regions, detected using SNP-Darmor (Supplemental Table 8). To adapt to a shorter (or longer) life cycle requires not only a change in flowering time but also changes in other relevant traits, such as the duration of the vegetative and reproductive periods. Natural or artificial selection on flowering-time genes, such as orthologs of *FY*, *FT*, and *FLC*, could result from flowering-time adaptation. On the other hand, the selection of genes involved in ethylene biosynthesis or signaling pathways, such as *1-aminocyclopropane-1-carboxylate synthase 2* (*ACS2*) (*BnaC05g00530D*) and a variety of *ERFs* (*BnaA02g01570D*, *BnaA02g01920D*, *BnaA02g13950D*, *BnaA02g13960D*, and *BnaA02g27680D*) (Supplemental Table 8) could result from adaptive changes in growth period. ACS2 catalyzes the biosynthesis of ACC, which is the precursor of ethylene (Schellingen et al., 2014), and *ERFs* are transcription factor genes that encode ethylene-responsive element binding proteins responsible in part for mediating the response to ethylene signaling (Ohme-Takagi and Shinshi, 1995). Ethylene plays important roles in leaf and petal senescence, as well as in fruit ripening (Pratt et al., 1948; Ju et al., 2015). Selective-sweep analyses revealed numerous genetic footprints left during natural and artificial selections, i.e., modification of genes that regulate flowering time, hormone synthesis and signaling, and stress responses, leading to the differences between semi-winter and winter types, and between semi-winter and spring types (Supplemental Figures 9–11 and Supplemental Tables 9 and 10).

GWAS identified 22 SNPs ($-\log_{10}P > 6.0$) linked with 37 genes that are strongly associated with flowering time and two genes associated with ethylene signaling and response. These include genes orthologous to many flowering and ethylene signaling pathway genes in *Arabidopsis* (Supplemental Figure 12 and Supplemental Table 11). Of these genes, *FT* and *FLC* are the most important adaptive genes; both are found in single copy in *Arabidopsis* but duplication events have given rise to multiple copies in the *B. napus* lineage (Guo et al., 2014; Schiessl et al., 2017b). Not all paralogous genes in a polyploid genome such as that of *B. napus* are equally important in terms of function (Hua et al., 2009). GWAS conducted using millions of SNPs has been a powerful method for identifying functional copies among various paralogs in polyploid crops (Fang et al., 2017; Guo et al., 2018). In accordance with the selective-sweep analyses, GWAS analysis revealed that SNPs in the regions containing the *FT* ortholog (*BnaA02g12130D*) on Chr.A02 and the *FLC* ortholog (*BnaA10g22080D*) on Chr.A10 are significantly associated with flowering time. Interestingly, the coding regions of these two genes are well conserved among the three ecotype groups, whereas some SNPs in the putative promoter regions are specific to a single group (Figures 5C and 6C). Since the expression levels of the *FT* and *FLC* orthologs are significantly higher and lower, respectively, in spring-type accessions than in the other two types of accessions before the floral transition (Figures 5D and 6D), the ecotype-specific SNPs in the *FT* and *FLC* orthologs could contribute to the differential expression of these genes in different ecotypes, and thus the control of flowering time in response to the environmental conditions to which the three ecotype groups are geographically adapted.

The sequencing coverage depth in this study did not permit us to perform reliable structural-variation analyses, such as analysis of gene presence and absence and copy-number variation among the three ecotype groups. A recent study based on the sequencing of 280 *B. napus* inbred lines revealed that post-polyploidization morphotype diversification is associated with gene copy-number variation, and that 12 haplotypes, including *Bna.FLC.A10*, *Bna.VIN3.A02*, and the *Bna.FT* promoter on C02_random, are indicative of the diversification of winter and spring types (Schiessl et al., 2017a). However, since the *FT* ortholog on Chr.C02 is expressed at very low levels in all ecotypes tested (Supplemental Figure 14A) (Wang et al., 2012), it is unlikely that this *FT* ortholog directly contributes to the genetic basis of flowering-time divergence among the three rapeseed ecotype groups.

Taken together, this study reports the global pattern of genetic polymorphism in rapeseed revealed by resequencing a worldwide collection of 991 germplasm accessions, and sheds new light on the genomic footprints generated during the process of natural and artificial selection and flowering-time divergence among three major ecotype groups. Our study provides by far the largest genetic resource for screening molecular markers for the genetic improvement of rapeseed.

METHODS

Sampling

A total of 991 *B. napus* accessions from 39 countries were collected for genome resequencing. Of these, 821 were maintained by the Leibniz Institute of Plant Genetics and Crop Plant Research (https://gbis.ipk-gatersleben.de/gbis2i/faces/index.jsf) in Gatersleben, Germany, and 170 were preserved by the Provincial Key Laboratory of Crop Gene Resources of Zhejiang University. The collection consists of 658 winter, 145 semi-winter, and 188 spring types based on the information in the passport of each accession as well as several years' field observations (Supplemental Table 1). Biased weight was given to genetic materials from the regions important for rapeseed production, such as Europe, Canada, China, and Australia. Detailed information about the country of origin and ecotype for each accession is shown in Supplemental Table 1. A world map showing origin and ecotype information was made using the R package ggplot2 (Wickham, 2016). Seeds of all accessions were sown in mixed nutrient soil in a greenhouse at Zhejiang University, China. Young leaves from 21-day-old seedlings of each accession were sampled for genomic DNA extraction.

DNA Extraction and Sequencing

Genomic DNA from young leaves was extracted using a cetyltrimethylammonium bromide (CTAB)-based protocol (Uzunova et al., 1995). The concentration and quality of the total genomic DNA were determined using a NanoDrop2000 Spectrophotometer (Thermo Fisher Scientific). DNA libraries for Illumina sequencing were constructed for each accession according to the manufacturer's specifications (Illumina, CA, USA). After DNA library construction, sequencing was performed on an Illumina HiSeq XTen platform by a commercial service (Biomarker Technologies, Beijing, China), with 150-bp read lengths. Raw reads were filtered based on the following criteria: pair-end reads with >10% "N" bases; average base-quality less than 20 (Phred-like score); quality score of 3′ bases $\leq$40. Finally, 7.9 Tb (Q30 = 92.5%) of high-quality sequences were obtained for subsequent analyses (Supplemental Table 2).

SNP and InDel Calling

All clean reads for each accession were mapped to two *B. napus* reference genomes, the "Darmor-*bzh*" genome (*B. napus* v4.1 genome, http://www.genoscope.cns.fr/brassicanapus/data/) and the "Tapidor" genome (https://www.ncbi.nlm.nih.gov/bioproject/342383) using the MEM algorithm of Burrows–Wheeler Aligner (Li and Durbin, 2009) (BWA v0.7.5a-r405). The average mapping rates were 97.7% and 95.0%, and the average coverage rates were 6.6-fold and 6.8-fold for the "Darmor-*bzh*" genome and "Tapidor" genome, respectively (Supplemental Table 2). The mapping results were sorted, and duplicate reads were marked using SAMTOOLS (Li et al., 2009) (v1.1) and PICARD (http://broadinstitute.github.io/picard/;v1.94). Local realignment around InDel regions was performed by InDel-Realigner in GATK (McKenna et al., 2009) (v3.7). SNPs and InDels within the 991 accessions were called using the HaplotypeCaller module in GATK and were filtered with the following parameters: QD < 2.0 || MQ < 40.0 || FS > 60.0 || QUAL <30.0 || MQrankSum < -12.5 || ReadPosRankSum < -8.0 -clusterSize 2 -clusterWindowSize 5. A total of 5 559 254 and 5 526 961 SNPs, and 1 858 671 and 1 919 465 InDels were obtained across the 991 genomes by mapping reads to the "Darmor-*bzh*" and "Tapidor" reference genomes, respectively (Supplemental Tables 3 and 4). The SNPs identified by GATK were further filtered: only SNPs with a minor allele frequency greater than 5% and less than 50% missing data were considered as high-quality SNPs. Finally, 2 753 575 ("Darmor-*bzh*" reference genome) and 2 703 562 ("Tapidor" reference genome) high-quality SNPs were obtained. SNP annotation was performed on the basis of the *B. napus* v4.1 genome using snpEff software (Cingolani et al., 2012), and SNPs were categorized into intergenic regions, upstream or downstream regions, and exons or introns. SNPs in coding exons were further classified as synonymous SNPs or nonsynonymous SNPs. InDels in exons were grouped according to whether they led to a frameshift.

Phylogenetic and Population Structure Analysis

To analyze the phylogenetic relationships of the 991 accessions, we constructed an unrooted phylogenetic tree using the neighbor-joining method with the Kimura 2-parameter model in MEGA5.2 software (Tamura et al., 2011), with 1000 bootstrap replicates. To minimize the contribution from regions of extensive strong LD, we scanned the whole genome with a sliding window of 100 kb (in steps of 10 SNPs), and removed any SNPs correlated with other SNPs within the window with a correlation coefficient (r^2) > 0.2 using the PLINK software (www.cog-genomics.org/plink2; v1.9). In total, 293 498 ("Darmor-*bzh*" reference genome) and 386 947 ("Tapidor" reference genome) high-confidence SNPs were used to infer the population structure within 991 *B. napus* accessions using ADMIXTURE (Alexander et al., 2009) (v1.22), in haploid mode with five different random seeds and *K* values (the putative number of populations) ranging from 2 to 10. We assessed the number of sub-populations using the ΔK method (Evanno et al., 2005). The Q matrix from five replicates for each *K* value was merged using CLUMPP software (https://web.stanford.edu/group/rosenberglab/clumpp.html) and stacked assignment bar plots were generated using the R package Pophelper (http://royfrancis.github.io/pophelper).

Principal Component Analysis and Linkage Disequilibrium Analysis

PCA was performed using the smartPCA program from the EIGENSOFT package (https://github.com/DReichLab/EIG; v.6.0.1). The first two PCs were used to separate samples from different groups. PC1 separated the winter type from the other two types, while PC2 mainly separated the semi-winter type from the winter and spring types. For LD analysis, complete and partial LD was calculated between each pair of SNPs using PLINK software (www.cog-genomics.org/plink2; v1.9). The squared correlation coefficient (r^2) values and the significance of any LD detected between polymorphic sites ($P < 0.05$) were analyzed for all chromosomes with a 1000-kb window. LD blocks were defined with the following parameters: –ld-window-r2 0 –ld-window 99 999 –ld-window-kb 1000.

Recombination Rate Analysis

To ensure imputation quality, we filtered out SNPs with a low MAF (<0.05) and low sequence coverage (missing data > 20%). After genotype imputation using Beagle software (Browning and Browning, 2008), we obtained 2 135 013 SNPs for recombination rate analysis. We calculated the recombination rate for all chromosomes with a 100-kb window (in steps of 10 kb) using the R package FastEPRR (Gao et al., 2016).

Allelic-Drift Analysis

To explore the pattern of splits and mixtures in spring- and winter-type rapeseed populations, we used TREEMIX (Pickrell and Pritchard, 2012) (v.1.13) to construct a population graph depicting the relationships between the two rapeseed populations. This method fits a bifurcating tree based on the population allele frequency covariance matrix and adds migration/admixture edges to the tree to improve the fit. Based on the results, we mapped the gene-flow information and migration events onto the world map using the R package ggplot2 (Wickham, 2016).

Selective Sweep Analyses

The F_{ST} and cross-population composite likelihood ratio test (XP-CLR) scores were calculated using PopGenome (https://popgenome.weebly.com/) and the XP-CLR package (Chen et al., 2010), respectively. The F_{ST} values were estimated for 100-kb sliding windows with a 10-kb step size along each chromosome. The average F_{ST} was the value for the whole genome across different groups. Sliding windows with F_{ST} values greater than 99% of genome-wide F_{ST} values were selected as significant windows. The overlapping significant windows were then merged into fragments, which were considered highly diverged regions across groups. The nonsynonymous SNPs within the diverged regions with the top 1% of F_{ST} values were selected, and the corresponding genes were considered selective sweep candidates.

Whole-genome screening of selection was also performed using XP-CLR, a method based on modeling the likelihood of multilocus allele frequency differentiation between two populations. Genetic distances between SNPs were interpolated according to their physical distances in an ultra-high-density genetic map (Yang et al., 2017). The program XP-CLR was run for each chromosome with the following parameters: -w1 0.005 200 10 000 -p0 0.95. After obtaining the original XP-CLR results, each chromosome was divided into sliding windows (100-kb window with a 10-kb step size), and the average XP-CLR score was calculated. Sliding windows with an average XP-CLR score greater than the 95th percentile of the genome-wide average XP-CLR score were selected as significant windows. Overlapping significanct windows were then merged into fragments, which were considered highly diverged regions across groups. An integrated figure of SNP density, LD structure, recombination rate, and selective-sweep signals on the 19 chromosomes was generated using Circos software (http://circos.ca/).

Flowering-Time Measurement

The 991 rapeseed germplasm accessions were grown in the experimental field of Changxing Agricultural Experiment Station of Zhejiang University (31°02′N and 119°93′E) during the rapeseed production season from October 2017 to May 2018. The rapeseed materials were sown in nursery beds in a plastic tunnel before being transplanted into the experimental field. Twelve plants for each accession were transplanted in a block (160 × 40 cm) with three replicates. In spring 2018, the flowering time for each accession was recorded when 50% of the plants had visible open flowers on the main inflorescence.

GWAS

GWAS of flowering time was based on the 2 753 575 ("Darmor-*bzh*" reference genome) and 2 703 562 ("Tapidor" reference genome) high-quality SNPs (MAF > 0.05, missing rate < 0.5) from 991 accessions. The Q value, which represents population structure, was calculated using EIGENSOFT software. The *K* value, which represents the genetic relationship between samples, was calculated using PLINK software. The general linear model and mixed linear model (MLM) in TASSEL (http://www.maizegenetics.net/tassel) were used to test the associations. Both Q and *K* values were used in the MLM. The Emmax test was performed using EMMAX software (http://csg.sph.umich.edu//kang/emmax/download/index.html). The *p* value was calculated for each SNP and $-\log_{10}P > 6$ was defined as the suggestive threshold and genome-wide control threshold.

Promoter Sequence and Expression Analyses of *FLC* and *FT* Orthologs

To analyze the distribution of SNPs in the putative promoter regions of the *FLC* and *FT* orthologs in different rapeseed ecotypes, we compared the SNPs in a region ~5 kb upstream of the *FLC* and *FT* orthologs according to the "Darmor-*bzh*" reference genome in the 991 accessions, excluding 88 accessions with different classifications based on the two grouping systems. "0", "1" and "2" indicate SNPs homozygous for the reference allele (0), heterozygous SNPs (1), and SNPs homozygous for the nonreference allele (2), respectively. A total of 124 accessions including 65 winter, eight semi-winter, and 51 spring types used for an associative transcriptomics study (Harper et al., 2012; Havlickova et al., 2018) were selected to calculate the expression levels of *BnaA10g22080D* and *BnaA02g12130D*. The data were downloaded from NCBI (https://www.ncbi.nlm.nih.gov) and mapped to the *B. napus* reference genome with STAR (Dobin et al., 2013) (v.2.5.4a). We calculated the expression levels of each gene in transcript per million using RSEM (Li and Dewey, 2011) (v.1.3.0).

Expression Analysis

A total of 112 rapeseed accessions (39 winter, 21 semi-winter, and 52 spring) out of the 991 accessions were grown in a growth chamber under long-day conditions (14 h light/10 h dark) (Supplemental Table 19). One month after sowing, the second true leaves from four seedlings were sampled for RNA extraction using an RNA extraction kit (cat. no. R6827-1, Omega, USA). RNA was reverse transcribed with the HiScript II Q RT SuperMix for qPCR (+gDNA wiper) Kit (cat. no. R223-01, Vazyme, China) and qRT–PCR experiments were performed with ChamQ Universal SYBR qPCR Master Mix (cat, no. Q711-02, Vazyme) according to the manufacturer's specifications. The primers used for the *FT* ortholog (*BnaA02g12130D*) were 5′-CGA AGA AGT TAG ATG AGC CTC TTC T-3′ and 5′-ACC ATA TAA ATA AAA ACA CTC CCC C-3′, and those for the *FLC* ortholog (*BnaA10g22080D*) were 5′-ATA TGG ATG TCT CAC CAG GAC AAA T-3′ and 5′-CGA GCC GGA GAG AGA GTA TAG ATT A-3′, which were designed to specifically amplify the putative *FT* and *FLC* orthologs on Chr.A02 and Chr.A10, respectively (Supplemental Figure 16). The expression of *BnACTIN7* (5′-GGA GCT GAG AGA TTC CGT TG-3′ and 5′-GAA CCA CCA CTG AGG ACG AT-3′) was used as an internal control. The difference between the cycle threshold (Ct) of target genes and the Ct of the control gene ($\Delta Ct = Ct_{target\ gene} - Ct_{control}$) was used to calculate the normalized expression of target genes.

ACCESSION NUMBERS

All of the raw reads of the rapeseed accessions generated in this study have been deposited in the public database of National Center of Biotechnology Information under SRP155312 (https://www.ncbi.nlm.nih.gov/sra/SRP155312).

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

The work was supported by the National Key Basic Research Project (no. 2015CB150205), Natural Science Foundation of China, China (no. 31671597, 31370313, 31670283), Sino-German Science Center for Research Promotion, China (GZ 1099), Jiangsu Collaborative Innovation Center for Modern Crop Production, China, and the Singapore National Research Foundation Investigatorship Program, Singapore (NRF-NRFI2016-02).

AUTHOR CONTRIBUTIONS

L.J., H.Y., L.S., and G.Z. conceived the project; D.W., Z.L., T.Y., Y.X., L.X., U.L., S.H., Z.L., H.W., Q.W., L.Z., A.M., I.H.S., G.J., N.H., and X.C. performed the experiments; D.W., Z.L., T.Y., J.T., G.Z., X.L., L.W., B.C., U.L., B.C., I.H.S., G.J., and H.Z. conducted bioinformatics analyses and were involved in many discussions; D.W., Z.L., T.Y., and X.Y. analyzed data; L.J., H.Y., L.S., and D.W. wrote the paper.

ACKNOWLEDGMENTS

No conflict of interest declared.

Received: July 29, 2018
Revised: November 17, 2018
Accepted: November 18, 2018
Published: November 21, 2018

REFERENCES

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

Amasino, R. (2010). Seasonal and developmental timing of flowering. Plant J. **61**:1001–1013.

An, H. (2004). CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*. Development **131**:3615–3626.

Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. Nat. Rev. Genet. **13**:627–639.

Bayer, P.E., Hurgobin, B., Golicz, A.A., Chan, C.K., Yuan, Y., Lee, H., Renton, M., Meng, J., Li, R., Long, Y., et al. (2017). Assembly and comparison of two closely related *Brassica napus* genomes. Plant Biotechnol. J. **15**:1602–1610.

Becker, H.C., Engqvist, G.M., and Karlsson, B. (1995). Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. Theor. Appl. Genet. **91**:62–67.

Browning, B.L., and Browning, S.R. (2008). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. **84**:210–223.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science **345**:950–953.

Chen, S., Nelson, M.N., Ghamkhar, K., Fu, T., and Cowling, W.A. (2008). Divergent patterns of allelic diversity from similar origins: the case of oilseed rape (*Brassica napus* L.) in China and Australia. Genome **51**:1–10.

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

Cheng, F., Sun, R., Hou, X., Zheng, H., Zhang, F., Zhang, Y., Liu, B., Liang, J., Zhuang, M., Liu, Y., et al. (2016). Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. Nat. Genet. **48**:1218–1224.

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. Fly **6**:80–92.

Cleemput, S., and Becker, H.C. (2012). Genetic variation in leaf and stem glucosinolates in resynthesized lines of winter rapeseed (*Brassica napus* L.). Genet. Resour. Crop Evol. **59**:539–546.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., et al. (2007). FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. Science **316**:1030–1033.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**:15–21.

Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., Li, N., Zhang, X., Sun, J., Liu, M., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat. Genet. **50**:1–7.

Dupont, J., White, P.J., Johnston, K.M., Heggtveit, H.A., Mcdonald, B.E., Grundy, S.M., and Bonanome, A. (1989). Food safety and health-effects of canola oil. J. Am. Coll. Nutr. **8**:360–375.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. **14**:2611–2620.

Fang, L., Wang, Q., Hu, Y., Jia, Y., Chen, J., Liu, B., Zhang, Z., Guan, X., Chen, S., Zhou, B., et al. (2017). Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. Nat. Genet. **49**:1089–1098.

Fussell, G.E. (1955). History of cole. Nature **9**:48–51.

Gao, F., Ming, C., Hu, W., and Li, H. (2016). New software for the fast estimation of population recombination rates (FastEPRR) in the genomic era. G3 (Bethesda) **6**:1563–1571.

Girke, A., Schierholt, A., and Becker, H.C. (2012). Extending the rapeseed gene pool with resynthesized *Brassica napus* II: Heterosis. Theor. Appl. Genet. **124**:1017–1026.

Guo, Y., Hans, H., Christian, J., and Molina, C. (2014). Mutations in single FT- and TFL1-paralogs of rapeseed (*Brassica napus* L.) and their impact on flowering time and yield components. Front. Plant Sci. **5**:1–12.

Guo, Z., Liu, G., Röder, M.S., Reif, J.C., Ganal, M.W., and Schnurbusch, T. (2018). Genome-wide association analyses of plant growth traits during the stem elongation phase in wheat. Plant Biotechnol. J. **16**:2042–2052.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P., and Bancroft, I. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nat. Biotechnol. **30**:798–802.

Havlickova, L., He, Z., Wang, L., Langer, S., Harper, A.L., Kaur, H., Broadley, M.R., Gegas, V., and Bancroft, I. (2018). Validation of an updated associative transcriptomics platform for the polyploid crop species *Brassica napus* by dissection of the genetic architecture of erucic acid and tocopherol isoform variation in seeds. Plant J. **93**:181–192.

Hua, S., Shamsi, I.H., Guo, Y., Pak, H., Chen, M., Shi, C., Meng, H., and Jiang, L. (2009). Sequence, expression divergence, and complementation of homologous ALCATRAZ loci in *Brassica napus*. Planta **230**:493–503.

Jaeger, K.E., and Wigge, P.A. (2007). FT protein acts as a long-range signal in *Arabidopsis*. Curr. Biol. **17**:1050–1054.

Ju, C., Van de Poel, B., Cooper, E.D., Thierer, J.H., Gibbons, T.R., Delwiche, C.F., and Chang, C. (2015). Conservation of ethylene as a plant hormone over 450 million years of evolution. Nat. Plants **1**:14004.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**:323.

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. (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079.

Liu, H.L. (2000). Genetics and Breeding in Rapeseed (Beijing, China: Chinese Agricultural Universities Press).

Maynard Smith, J., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. **23**:22–35.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2009). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**:1297–1303.

Michaels, S.D., and Amasino, R.M. (2001). Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization. Plant Cell **13**:935–941.

Nagaharu, U. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn. J. Bot. **7**:389–452.

Ohme-Takagi, M., and Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell **7**:173–182.

Pratt, H.K., Young, R.E., and Biale, J.B. (1948). The identification of ethylene as a volatile product of ripening avocados. Plant Physiol. **23**:526–531.

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

Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R.S., Sherwood, E., Webster, M.T., Jiang, L., Ingman, M., Sharpe, T., Ka, S., et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. Nature **464**:587–591.

Schellingen, K., Van Der Straeten, D., Vandenbussche, F., Prinsen, E., Remans, T., Vangronsveld, J., and Cuypers, A. (2014). Cadmium-induced ethylene production and responses in *Arabidopsis thaliana* rely on ACS2 and ACS6 gene expression. BMC Plant Biol. **14**:214.

Schiessl, S., Huettel, B., Kuehn, D., Reinhardt, R., and Snowdon, R. (2017a). Post-polyploidisation morphotype diversification associates with gene copy number variation. Sci. Rep. **7**:41845.

Schiessl, S.V., Huettel, B., Kuehn, D., Reinhardt, R., and Snowdon, R.J. (2017b). Flowering time gene variation in *Brassica* species shows evolutionary principles. Front. Plant Sci. **8**:1–13.

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Kröber, S., Amasino, R.A., and Coupland, G. (2006). The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. Genes Dev. **20**:898–912.

Sheldon, C.C. (1999). The FLF MADS box gene: a repressor of flowering in arabidopsis regulated by vernalization and methylation. Plant Cell **11**:445–458.

Slatkin, M. (2008). Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. **9**:477–485.

Srikanth, A., and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. Cell. Mol. Life Sci. **68**:2013–2037.

Suarez-Lopez, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. Nature **410**:1116–1120.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.

Uzunova, M., Ecke, W., Weissleder, K., and Röbbelen, G. (1995). Mapping the genome of rapeseed (*Brassica napus* L.). I. Construction of an RFLP linkage map and localization of QTLs for seed glucosinolate content. Theor. Appl. Genet. **90**:194–204.

Wang, J., Hopkins, C.J., Hou, J., Zou, X., Wang, C., Long, Y., Kurup, S., King, G.J., and Meng, J. (2012). Promoter variation and transcript divergence in Brassicaceae lineages of *FLOWERING LOCUS T*. PLoS One **7**:e47127.

Wang, X., Long, Y., Wang, N., Zou, J., Ding, G., Broadley, M.R., White, P.J., Yuan, P., Zhang, Q., Luo, Z., et al. (2017). Breeding histories and selection criteria for oilseed rape in Europe and China identified by genome wide pedigree dissection. Sci. Rep. **7**:1916.

Wei, D., Cui, Y., He, Y., Xiong, Q., Qian, L., Tong, C., Ding, Y., Li, J., and Qian, W. (2017). A genome-wide survey with different rapeseed ecotypes uncovers footprints of domestication and breeding. J. Exp. Bot. **68**:4791–4801.

Wickham, H. (2016). ggplot2—Elegant Graphics for Data Analysis, 2nd edn (New York: Springer-Verlag).

Yang, Y., Shen, Y., Li, S., Ge, X., and Li, Z. (2017). High density linkage map construction and QTL detection for three silique-related traits in orychophragmus violaceus derived *Brassica napus* population. Front. Plant Sci. **8**:1512.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. **33**:408–414.