Lawrence Berkeley National Laboratory

Title

Young inversion with multiple linked QTLs under selection in a hybrid zone

Permalink

Journal

ISSN

Authors

Lee, Cheng-Ruei
Wang, Baosheng
Mojica, Julius P
et al.

Publication Date

DOI

OPEN

Young inversion with multiple linked QTLs under selection in a hybrid zone

Cheng-Ruei Lee[1,2]*†, Baosheng Wang[1,3]†, Julius P. Mojica[1], Terezie Mandáková[4], Kasavajhala V. S. K. Prasad[5], Jose Luis Goicoechea[6], Nadeesha Perera[1], Uffe Hellsten[7], Hope N. Hundley[7], Jenifer Johnson[7], Jane Grimwood[8], Kerrie Barry[7], Stephen Fairclough[7], Jerry W. Jenkins[7], Yeisoo Yu[9], Dave Kudrna[10], Jianwei Zhang[10], Jayson Talag[10], Wolfgang Golser[10], Kathryn Ghattas[1], M. Eric Schranz[11], Rod Wing[8], Martin A. Lysak[6], Jeremy Schmutz[7], Daniel S. Rokhsar[7] and Thomas Mitchell-Olds[1]*

Fixed chromosomal inversions can reduce gene flow and promote speciation in two ways: by suppressing recombination and by carrying locally favoured alleles at multiple loci. However, it is unknown whether favoured mutations slowly accumulate on older inversions or if young inversions spread because they capture pre-existing adaptive quantitative trait loci (QTLs). By genetic mapping, chromosome painting and genome sequencing, we have identified a major inversion controlling ecologically important traits in *Boechera stricta*. The inversion arose since the last glaciation and subsequently reached local high frequency in a hybrid speciation zone. Furthermore, the inversion shows signs of positive directional selection. To test whether the inversion could have captured existing, linked QTLs, we crossed standard, collinear haplotypes from the hybrid zone and found multiple linked phenology QTLs within the inversion region. These findings provide the first direct evidence that linked, locally adapted QTLs may be captured by young inversions during incipient speciation.

Chromosome inversions play an important role in local adaptation and speciation[1,2], and selectively important inversions have been identified in many species[3,4]. Selection due to different environmental factors or stages in the life cycle[1] may favour inversions carrying locally adapted alleles at several loci. In addition, established inversions are predicted to accumulate selectively important genetic differences, which may contribute to reproductive isolation during speciation[1].

Although few studies have identified the actual loci that influence selection on inversions[2,4,5], rearrangements may be favoured due to gene alterations near breakpoints[6], chromatin changes[7] or combinations of advantageous, co-adapted alleles[8]. Inversions suppress recombination, so locally advantageous alleles may segregate together, causing higher fitness than recombinant haplotypes[9]. Most evolutionary studies have focused on widespread, older inversions, so we have little knowledge of the evolutionary processes that guide their initial increase in frequency. For example, do inversions drift to higher frequency, and then acquire new advantageous mutations after they are common? Or are multiple linked, advantageous alleles captured in a new inversion, allowing them to spread together? Analysis of younger inversions may elucidate the evolutionary forces controlling the initial spread of chromosome inversions, which therefore influence their role in adaptation and speciation[4,8].

Related species often differ for chromosome inversions that carry locally favoured alleles at multiple loci[10,11]. A key distinction among models for the evolution of inversions is whether early frequency increase is due to genetic drift or natural selection. Genetic drift might predominate initially, with subsequent accumulation of advantageous variants[12]. Alternatively, the Kirkpatrick–Barton model[9] argues that linked, locally adapted alleles exist first, and subsequently are captured within a new, selectively favoured inversion[13]. In this 'inversion-late' evolutionary sequence[1,5], linked quantitative trait loci (QTLs), similar to the ancestral haplotype that gave rise to the inversion, may still exist in non-inverted genotypes[9]. Here, we test these predictions of the Kirkpatrick–Barton model. First, we introduce ecologically diverged subspecies of *Boechera stricta*. Next, we examine a young inversion to infer the selective forces controlling its early increase in frequency. Finally, we cross collinear, standard genotypes from the hybrid zone to ask whether old, linked QTLs can be found within the inversion region.

Results and discussion

Young inversion in a hybrid zone. Previously, we showed that *B. stricta* (a close relative of *Arabidopsis*[14]) has two ecologically differentiated subspecies ('*East*' and '*West*')[15,16], which form a contact zone across >200 km in the northern Rocky Mountains. Throughout our study area, local water availability is the best predictor of habitat

[1]Department of Biology, Duke University, Box 90338, Durham, North Carolina 27708, USA. [2]Institute of Ecology and Evolutionary Biology and Institute of Plant Biology, National Taiwan University, Taipei 10617, Taiwan ROC. [3]Department of Plant Ecology and Genetics, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden. [4]Plant Cytogenomics Group, Central European Institute of Technology, Masaryk University, Kamenice 5, Brno CZ-62500, Czech Republic. [5]Department of Biology, Colorado State University, Fort Collins, Colorado 80523, USA. [6]Arizona Genomics Institute and BIO5 Institute, School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA. [7]Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. [8]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA. [9]Phyzen Genomics Institute, Phyzen Inc., Seoul 151–836, South Korea. [10]Arizona Genomics Institute and BIO5 Institute, School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA. [11]Biosystematics Group, Wageningen University and Research Center, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. †These authors contributed equally to this work. *e-mail: chengrueilee@ntu.edu.tw; tmo1@duke.edu
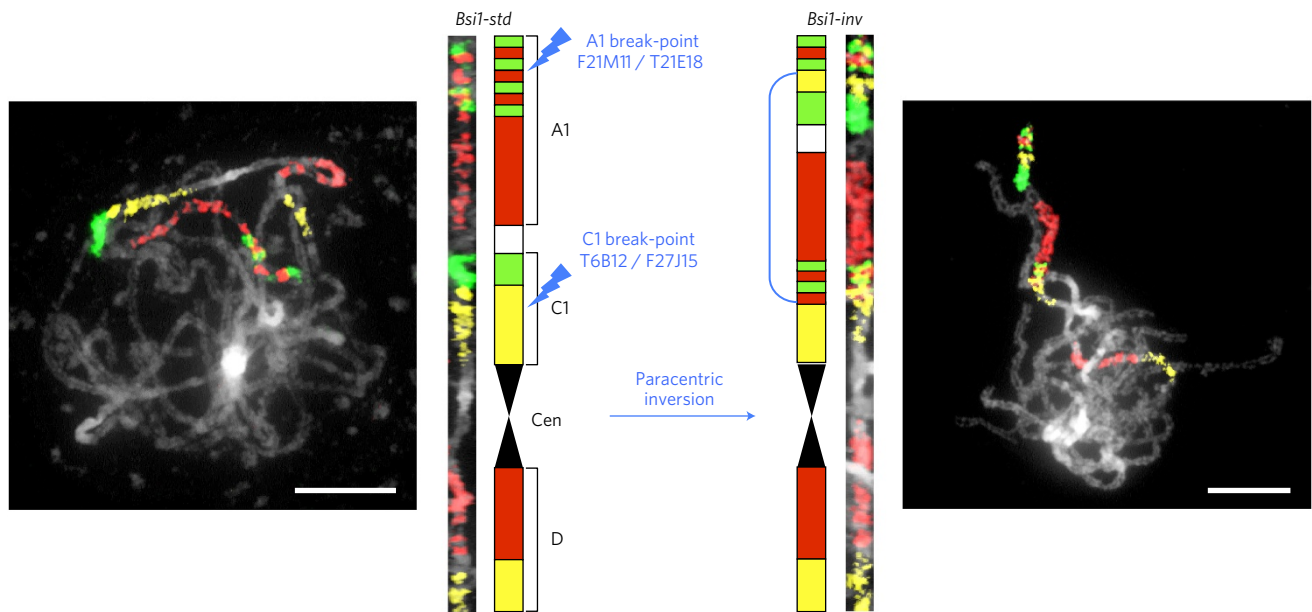
1

Figure 1 | Comparative chromosome painting of chromosome 1 in *Bsi1-std* and *Bsi1-inv* haplotypes. Images showing *in situ* chromosomal localization of painting probes on *B. stricta* pachytene chromosomes. Left, *Bsi1-std*. Right, *Bsi1-inv*. Straightened images and diagrammatic representations of the two haplotypes are shown beside the corresponding image. The curved blue line encompasses the inversion region. Differential labelling of BAC contigs (F21M11, T21E18, T6B12 and F27J15) identified the breakpoints and extent of the paracentric inversion in the *Bsi1-inv* genotype. Scale bars, 10 μm. Cen, centromere. See Supplementary Figure 3 for a detailed cytogenetic analysis.

divergence between these subspecies[15], with *West* genotypes growing in sites with more constant and abundant water supply. In and near the hybrid zone, the *East* and *West* subspecies show significant ecological differentiation across local environmental gradients, and the geographic distribution of the inversion falls within the typical range of hybrid zone habitats. The *West* subspecies has a significantly faster growth rate, larger leaf area, less succulent leaves, delayed reproductive time and longer flowering duration[16], and *West* × *East* crosses found QTLs for flowering traits, leaf number, defensive chemistry, herbivore resistance, cold tolerance, overwinter survival, fecundity and lifetime fitness[17–20]. In addition, Q_{ST}–F_{ST} analysis showed that phenology and some morphological traits have experienced divergent selection between subspecies[16].

In our initial cross[17], genetic mapping in *West* × *East* recombinant inbred lines (RILs) identified a region of suppressed recombination (Supplementary Fig. 1) on linkage group 1 (LG1). Here, we combined short-read sequencing, end-sequencing of bacterial artificial chromosomes (BACs), linkage mapping and physical mapping by Whole Genome Profiling (WGP) to assemble a high-quality genomic sequence (Supplementary Fig. 2) that enabled evolutionary analysis of the inversion. Chromosome painting (Fig. 1 and Supplementary Fig. 3) cytogenetically verified the presence of an 8.4 Mb paracentric inversion in the *West* genotype (*Bsi1*; Fig. 2a and Supplementary Fig. 4), spanning ~31 cM (~4% of the genome, the 'inversion region'; Supplementary Table 1). We developed primers to score the inversion using PCR (Fig. 2b), showing that the common, non-inverted allele (*Bsi1-std*) has the ancestral orientation found in closely related *Capsella* and *Arabidopsis lyrata*[21], and is found in >200 *East* and *West* populations across the species range (Fig. 3a and Dryad Data Archive—see Data availability section). The derived, inverted allele (*Bsi1-inv*) was identified in a cluster of *West* and hybrid populations in the contact zone (Fig. 3b), where it has risen to high frequency. Genotypes carrying the inversion span a narrow geographic range (~14 km, the 'inversion zone'), where pollen and geological analyses[22,23] suggest that conditions suitable for *B. stricta* existed during the Last Glacial Maximum, ~21 thousand years ago (ka). Single-nucleotide polymorphisms (SNPs) in the inversion region show that

the most similar *Bsi1-inv* and *Bsi1-std* haplotypes (Supplementary Fig. 5) are located only 1.7 km apart (Fig. 3b, left). Using measured mutation rates[24] and assuming two years per generation, the inversion dates to ~2.1–8.8 ka (Methods). Because these widely distributed subspecies are ecologically diverged[16], while the *Bsi1* inversion originated very recently, this polymorphism is compatible with the inversion-late Kirkpatrick–Barton model.

Previous analysis of a *West-inv* × *East-std* mapping population grown in the inversion zone found many QTLs controlling components of fitness[25] and strong selection on flowering time[17], influenced by recent climate warming[26]. Native *West Bsi1-inv* alleles had higher fitness than *East Bsi1-std* alleles[25]. The inversion (spanning ~4% of the genome) explains 7.5% of heritable variation for lifetime fecundity, and inversion homozygotes differ by 0.56 genetic standard deviations for this trait ($t = 3.64$, d.f. = 165, $P = 0.0004$). Here, we focused on the effects of the inversion by examining flowering time in two F7 near-isogenic line (NIL) families from this cross. We found no segregation distortion or deviation from Mendelian ratios (Supplementary Table 2) in either family, and the inversion had significant effects on flowering time in these NIL families in the greenhouse (F-ratio, $F_{2,282} = 21.81$; $P < 0.001$), where the *Bsi1-std* haplotype flowered ~2.0 days faster than the *Bsi1-inv* haplotype (Supplementary Fig. 6). Thus, the inversion controls 40% of the average 5.0 day difference in flowering time between subspecies (Supplementary Table 2d). While this pedigree cannot resolve QTLs within the inversion, this cross shows that the inversion has significant effects on an ecologically important trait that experiences natural selection[17]. Gene(s) within the *Bsi1* inversion control flowering in the field, and thus may contribute to reproductive isolation during speciation[27].

To test for phenotypic effects of the inversion or transcription changes near the breakpoints, we compared closely related inverted and standard haplotypes from the inversion zone, using a sympatric *West-inv* × *West-std* F2 cross (Fig. 3b, left). We found high seed germination (>99%), and no *Bsi1* segregation distortion or deviation from Mendelian ratios (Supplementary Table 3). Multivariate analysis of variance (MANOVA) showed a significant effect of *Bsi1* on a suite of phenological and morphological traits
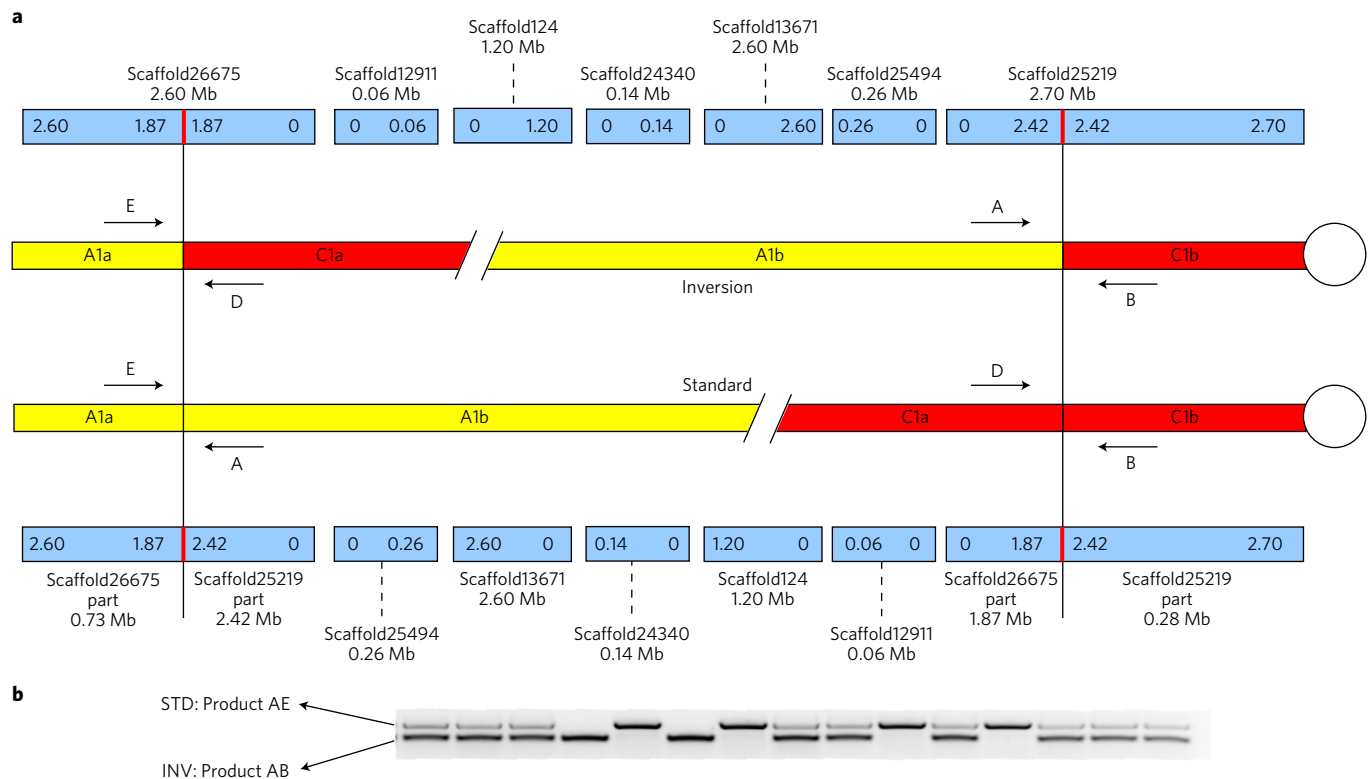
Figure 2 | Inversion region and PCR genotyping. a, Chromosome blocks (yellow and red) are shown for the derived (*Bsi-inv*, upper) and ancestral (*Bsi-std*, lower) haplotypes. Centromeres are indicated by white circles. Scaffolds (blue, above and below) are labelled with their name and size. Breakpoints are shown as vertical red lines in the scaffolds. Arrows indicate PCR primers. There are 1,591 annotated genes in the inversion region. **b**, Gel image showing PCR products, allowing codominant identification of inversion (INV) genotypes. STD, standard.

($P = 0.005$ by permutation test, $R^2 \approx 4\%$; Supplementary Table 4), so we analysed individual traits by analysis of variance (ANOVA). In this cross, flowering time was delayed by 0.9 days ($P = 0.030$ by permutation test) in *Bsi1-inv* homozygotes (accounting for 18% of the difference between subspecies), which may increase reproductive isolation from the early-flowering *East* genotypes[16]. Finally, we tested for functional changes attributable to the inversion breakpoints; these breakpoints do not disrupt existing open reading frames nor do they create new ones (Supplementary Fig. 7). Using replicated F3 homozygotes from this cross, we found no significant expression differences between inversion and standard haplotypes for any of the seven genes flanking the inversion breakpoints (Supplementary Table 5), although power to detect subtle, quantitative differences is limited. Thus, this comparison of *West-inv* versus *West-std Bsi1* haplotypes found genetic differences for ecologically important traits, but little evidence for functional changes at the inversion breakpoints.

Evidence for natural selection. To test whether selection had favoured this inversion, we analysed patterns of population genetic variation (Fig. 4 and Supplementary Fig. 8). *Bsi1-inv* genotypes show lower LG1 polymorphism, lower Tajima's *D* and more linkage disequilibrium (LD) than *Bsi1-std* individuals. These patterns are compatible with neutral drift in this partially inbreeding species, or with a selective sweep on the inversion[28]. In contrast, young, derived mutations at high frequency suggest positive directional selection[29], so we asked whether the frequency of *Bsi1-inv* is typical of the frequencies of all private derived SNP alleles in the same population (the 54 similar genotypes in the left portion of Supplementary Fig. 9c and the left portion of Fig. 3b). The inversion is at relatively high frequency (0.63) in this group, but it is not fixed. We found 2,416 SNPs that, like the inversion, are confined to this population.

Among these, only 63 SNPs (2.6%) have derived allele frequencies greater than the frequency of the *Bsi1-inv* inversion. Hence, the frequency of the *Bsi1-inv* inversion allele is higher than 97.4% of comparable derived alleles in this population—the inversion is a high frequency outlier, supporting the hypothesis of positive directional selection.

QTLs within the inversion. During speciation, co-adapted gene complexes within inversions might reduce gene flow if they preserve favourable combinations of alleles at multiple loci, reducing the frequency of disadvantageous allelic combinations. Although inversions can be engineered for proof of function in some organisms[30], this is infeasible in *Boechera*. However, the Kirkpatrick–Barton model predicts that relatives of the ancestral haplotype that gave rise to the inversion might still exist as non-inverted genotypes nearby. To test for such linked QTLs, we crossed collinear *West-std × East-std* genotypes from the contact zone and tested for QTLs within the inversion region, using freely recombining F3:F4 families. We found several multivariate QTLs altering ecologically important phenology and development traits (Fig. 5a). To clarify differences among these QTLs, we examined differences in their pleiotropic effects (Fig. 5b), using discriminant function analysis (DFA) to find the axis of greatest divergence between the multivariate trait means at each QTL peak. Each DFA axis quantifies the direction of pleiotropy that controls the effects of a locus, and QTLs influencing these composite traits were mapped across the inversion region. We found several distinct QTL peaks, which show divergent patterns of pleiotropy among these linked loci. Finally, we estimated the time of divergence between these parental genotypes (Fig. 5c). These results show that the QTL haplotypes diverged at least 50–100 ka, and are therefore much older than the inversion, which arose less than 10 ka.
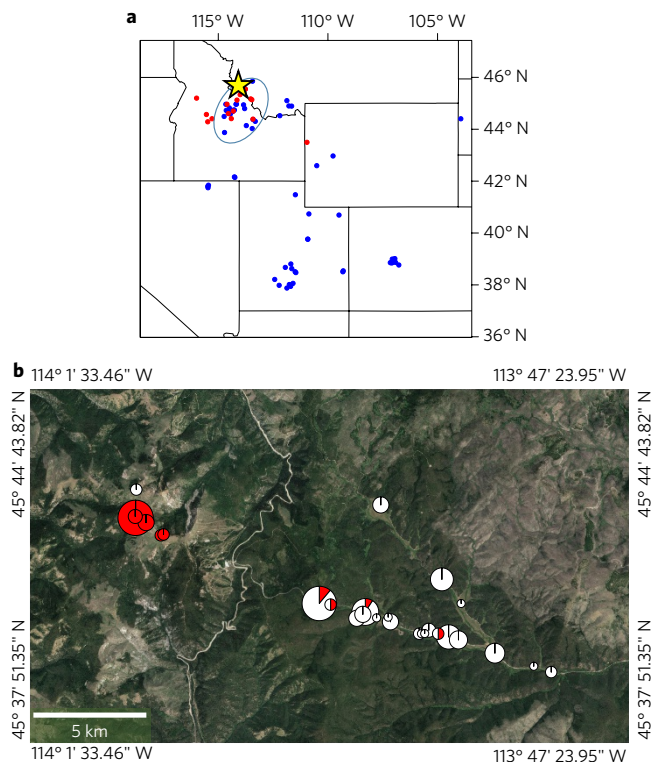
Figure 3 | Geographic locations of subspecies collections and inversion zone genotypes. a, Species-wide samples, with each dot indicating one genotype from a population. Subspecies assignments show *East* (blue) and *West* (red). The ellipse indicates the contact zone where subspecies overlap, and the star shows the inversion zone. *N* = 83 genotypes. The collinear cross comes from within the hybrid zone (ellipse), but outside the inversion zone (star). **b**, Close-up of populations near the inversion zone, with pie diagrams indicating the inversion frequency (red) in each population, sized in proportion to sample size. Forested, unsuitable habitat probably limits gene flow. *N* = 122 genotypes. Individual populations ranged in size from 1 to 19. Map image in **b**: © 2017 Google – Map Data.

In summary, population genetic evidence for a selective sweep corresponds with our ecological findings: the inversion affects multiple ecologically important traits, including flowering differences that are expected to increase reproductive isolation between subspecies. The inversion occurs in a hybrid zone, and this genomic region contains multiple ecologically important QTLs, as predicted by the Kirkpatrick–Barton model. However, beyond the potential advantage conferred by recombination suppression[9], closely related *West-inv* and *West-std* haplotypes show divergent phenotypic effects, which might contribute to selection in the hybrid zone. Our analysis of a young inversion is compatible with evolutionary predictions that linked, locally adapted QTLs may be captured by new inversions and contribute to local adaptation or incipient speciation[9].

Methods

Study area. The inversion was originally detected[21] at the Lost Trail Meadow site (~2,500 m elevation) about 4 km northwest from Lost Trail Pass on the Montana–Idaho border. Pollen and geological analyses[22,23] show that climate and vegetation in this area was strongly influenced by Pleistocene climatic changes, although the site apparently was unglaciated during the Last Glacial Maximum about 21 ka. Inferred plant communities[22,23] suggest that conditions suitable for *B. stricta* may have existed at or near our study site during the Last Glacial Maximum. We searched extensively in nearby *B. stricta* habitats; the Dryad Data Archive contains information on 122 genotypes from the inversion zone (Fig. 3) and 83 comparison genotypes from the across-the-species range.

Experimental pedigrees. Details on experimental crosses and genotype locations are in the Dryad Data Archive.

DNA extraction for genomic sequencing. *Seed sterilization.* Seeds of *B. stricta* ecotypes (LTM and SAD12) were initially surface sterilized with ethanol (95%) for 2 min followed by treatment with 15% sodium hypochlorite solution (with a few drops of Tween-20) for 25 min on a rotatory shaker (Spex CertiPrep; 125 rpm) at room temperature. Surface sterilized seeds were thoroughly washed multiple times with sterile water and were suspended in 0.1% agar. Seeds were kept for stratification in a cold room at 4 °C for 4 d under dark conditions.

Raising of aseptic seedling cultures. To raise aseptic seedlings of LTM and SAD12 ecotypes, surface sterilized and stratified seedlings were inoculated in 250 ml flasks containing 50 ml of sterile half-strength MS liquid medium (pH 5.7) with 0.5 g l⁻¹ MES and 2% sucrose. About 70–80 seeds were inoculated per flask. To obtain etiolated seedlings, the flasks were covered with aluminium foil and kept on a bench top rotatory shaker at 110 rpm for 20 d at room temperature.

Extraction of genomic DNA from etiolated seedlings. Freshly grown 20-day-old etiolated seedlings of both ecotypes were used for extraction of nuclear DNA. Isolation of nuclei was performed according to ref. [18] with slight modifications, as it allowed isolation of clean high-molecular-size nuclear DNA with minimal contamination from organelle DNA. Briefly, about ten flasks of etiolated seedling cultures were removed from the growing medium and thoroughly washed with ice-cold water and placed in ice-cold ethyl ether for 3 min. Subsequently the seedlings were washed five times with ice-cold TE buffer (pH 7.0). The plant material was quickly blotted on sterile filter paper and homogenized in MEB buffer using a commercial blender. The homogenate was filtered through four layers of cheesecloth followed by two layers of mira cloth. Triton X-100 was added to the filtrate at 0.5% concentration and centrifuged to collect the pellet. The pellet was suspended in MPDB solution and gently layered on 37.5% percoll made with MPDB. Nuclei were collected as a pellet after centrifugation, and the pellet was washed by resuspending it in 20 ml of MPBD followed by centrifugation. The nuclei pellet was again suspended in a MPDB solution and high-molecular-weight nuclear DNA was extracted using the Qiagen genomic tip 100/G protocol as per the manufacturer's instructions, with some modifications. The pellet consisting of nuclei was resuspended in the lysis buffer and incubated at 40 °C for 15 min. RNase A and T1 were added to the suspension and further incubated at 37 °C for an additional 30 min. Subsequently, Proteinase K was added to the suspension at a final concentration of 150 µg ml⁻¹ and incubated for 3 h at 45 °C with gentle shaking. The suspension was cleared by centrifugation at 8,000 rpm for 15 min. The cleared suspension was added to the pre-equilibrated G-100 column and further purification was performed as per the protocol provided by the manufacturer. The eluate consisting of genomic DNA was further precipitated using isopropanol, washed with 70% ethanol and subsequently suspended in TE (pH 8.0) buffer.

Total RNA isolation. To collect RNA from roots, the LTM-genotype seeds were surface sterilized with 12% bleach and grown under aseptic conditions on Whatman filter paper bridges in glass tissue culture tubes containing half-strength MS media with 1% sucrose. Each tube was inoculated with two to three seeds and allowed to grow. The roots obtained from the multiple seedlings were pooled and RNA was isolated using a plant RNeasy kit from Qiagen. For other tissues, we extracted RNA from soil-grown plants at several stages: 15-day-old seedlings, rosette leaves from juvenile plants, cauline leaves from plants aged 2–3 months. Finally, from 7-month-old plants, we extracted RNA from inflorescences, stems, flowers and siliques.

For isolation of total RNA from various tissues at different developmental stages, the tissues were snap frozen in liquid nitrogen and stored at −80 °C. The frozen tissue was homogenized with a pestle and mortar using liquid nitrogen. About 150 mg of homogenized tissue was used for isolation. Total RNA was isolated using a plant RNeasy kit. Each of the independent total RNA preparations was subjected to DNase treatment to remove traces of contaminating genomic DNA using RNase-free DNase (Promega). For each tissue type, three independent isolations were pooled together and used for further analysis.

Sequencing, assembly and annotation. Eighteen genomic DNA libraries and one transcriptome complementary DNA (cDNA) library were prepared (Dryad Data Archive) for sequencing using Illumina and Roche 454. The genome assembly of *B. stricta* employed Meraculous (May 2013 build)[31], as described on Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Bstricta): contigs were assembled using a size of 51 oligomers and a minimum depth of 8 from 2 × 150 Illumina HiSeq reads from an unamplified 250 bp whole genome shotgun fragment library. Several discrete rounds of scaffolding were performed with libraries containing inserts ranging in size from 250 bp to 40 kb (sequenced with various Illumina technologies). At each round of scaffolding the minimum pairing threshold was chosen by exhaustive optimization of N50 scaffold length. Assembly gaps were closed with 2 × 150 Illumina HiSeq reads. The resulting assembly comprised 171.9 Mb of contigs in 196.5 Mb of scaffold.

The genome assembly was soft-masked to highlight consensus repeat family sequences predicted *de novo* by RepeatModeler v.1.0.7. A total of 37,384 RNA-seq transcript assemblies were constructed from 118.7 million 2 × 150 paired-end
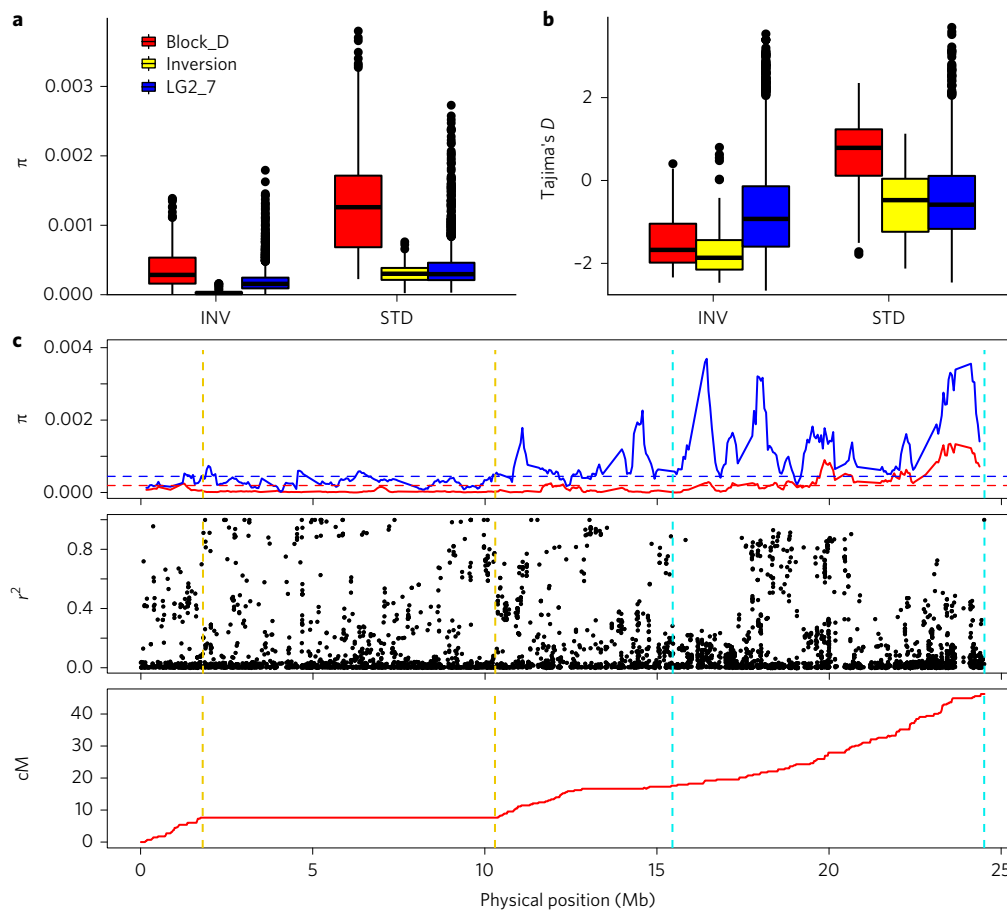
Figure 4 | Population genetic variation for inversion and standard groups. a,b, Nucleotide diversity (π, **a**) and Tajima's *D* (**b**) of the inverted genomic region (yellow), block D of chromosome 1 (red), and chromosomes 2–7 (blue) in the inversion (INV) and standard (STD) genotypes. Block D (south end of LG1) is treated separately because it has unusually high polymorphism in related species. In these box plots, the median is shown by a horizontal line, while the bottom and top of each box represents the first and third quartiles. The whiskers extend to 1.5 times the interquartile range. Outliers are represented by black dots. **c**, Distribution of population genetic statistics along chromosome 1. Nucleotide diversity (π) in INV (red) and STD (blue) genotypes, with genome-wide averages shown as dashed horizontal lines. Linkage disequilibrium (r^2) between the inversion and SNPs is shown for all INV and STD genotypes from the inversion zone, and the relationship between physical and linkage maps. The dashed vertical lines mark the inverted (gold) and block D (light blue) regions. $N = 122$ genotypes, except for the LD (r^2) analysis where four admixed individuals were excluded. LD for 10 comparator SNPs with derived allele frequencies similar to *Bsi1-inv* is shown in Supplementary Figure 8b.

Illumina RNA-seq reads using PERTRAN (Shengqiang Shu, Joint Genome Institute in-house pipeline, http://jgi.doe.gov/wp-content/uploads/2013/11/CSHL-PERTRAN-Shengqiang-Shu-FINAL.pdf). Loci were determined by BLAT transcript assembly alignments, BLASTX alignments of proteins from arabi (*Arabidopsis thaliana*), rice, soybean, grape, maize or *Chlamydomonas reinhardtii* genomes, and/or BLASTX alignments of UniProtKB/Swiss-Prot to the *B. stricta* genome. Gene models were predicated by homology-based predictors, mainly FGENESH+ and GenomeScan.

The best scored predictions for each locus were selected using multiple positive factors, including expressed sequence tag and protein support, and one negative factor: overlap with repeats. The selected gene predictions were improved using PASA r2011_05_20. Improvement includes adding untranslated regions, splicing correction, and adding alternative transcripts. PASA-improved gene model proteins were subject to analysis of their protein homology to the above-mentioned proteomes, in order to obtain a Cscore and protein coverage. Cscore is a protein BLASTP score ratio to mutual best hit BLASTP score, and protein coverage is the highest percentage of protein aligned to the best of the homologues. PASA-improved transcripts were selected on the basis of Cscore, protein coverage, expressed-sequence-tag coverage, and the percentage of their coding regions that overlapped with repeats. A transcript was selected if both its Cscore and protein coverage were ≥0.5, or if it exhibited expressed-sequence-tag coverage, but <20% of its coding region overlapped with repeats. For gene models where >20% of the coding region overlapped with repeats, to be selected, its Cscore and homology coverage had to be at least 0.9 and 70%, respectively. The selected gene models were subject to Pfam (protein family database) analysis and gene models whose protein included >30% transposable-element-associated Pfam domains were removed.

There are 1,591 annotated genes in the inversion region (below), of which 408 are found in the QTL regions within the inversion.

Whole genome profiling. A Hind III BAC library of *B. stricta* (Bs_LBa) was built from fresh leaf tissues from the LTM reference genotype[18]. The library contains 18,432 clones, arrayed into 48–384 well plates, with an average insert size of 150 Kb and an estimated genome depth of 11×. The library or clones of interest from it are publicly available at the Arizona Genomics Institute (AGI) web page (http://www.genome.arizona.edu/orders/).

A WGP physical map of *B. stricta* was built following a previously published protocol[32]. Briefly, BAC library clone plates were arrayed into two- and three-dimensional pools, and BAC DNA was extracted from the pooled plates. BAC DNA was digested with EcoRI/MseI and ligated to their respective adapters (P5-EcoRI-tag barcode and P7-MseI), followed by PCR amplification and sequencing with an Illumina HiSeq 2500.

Resulting sequences were deconvoluted with KeyGene proprietary scripts (licensed to AGI), to generate the tag file (Dryad Data Archive), which were assembled with FPC[33] v.9.4. Four test assemblies were performed with a fixed tolerance value of 0 and cutoff values of 1×10^{-25}, 1×10^{-20}, 1×10^{-15} and 1×10^{-10}, to choose the one producing the best results; based on the number of contigs and number of clones included in those contigs.

Questionable clones were eliminated from the best project (1×10^{-15}) and the contigs were merged at a cutoff value of 1×10^{-9}. Singletons (clones that do not assemble with these parameters), were added at reduced stringency (1×10^{-9}). This new project was subject to manual editing, after linking the tag, BAC-end and genome-scaffold (v 1.0) sequences to the physical map[34]. Using the SyMAP
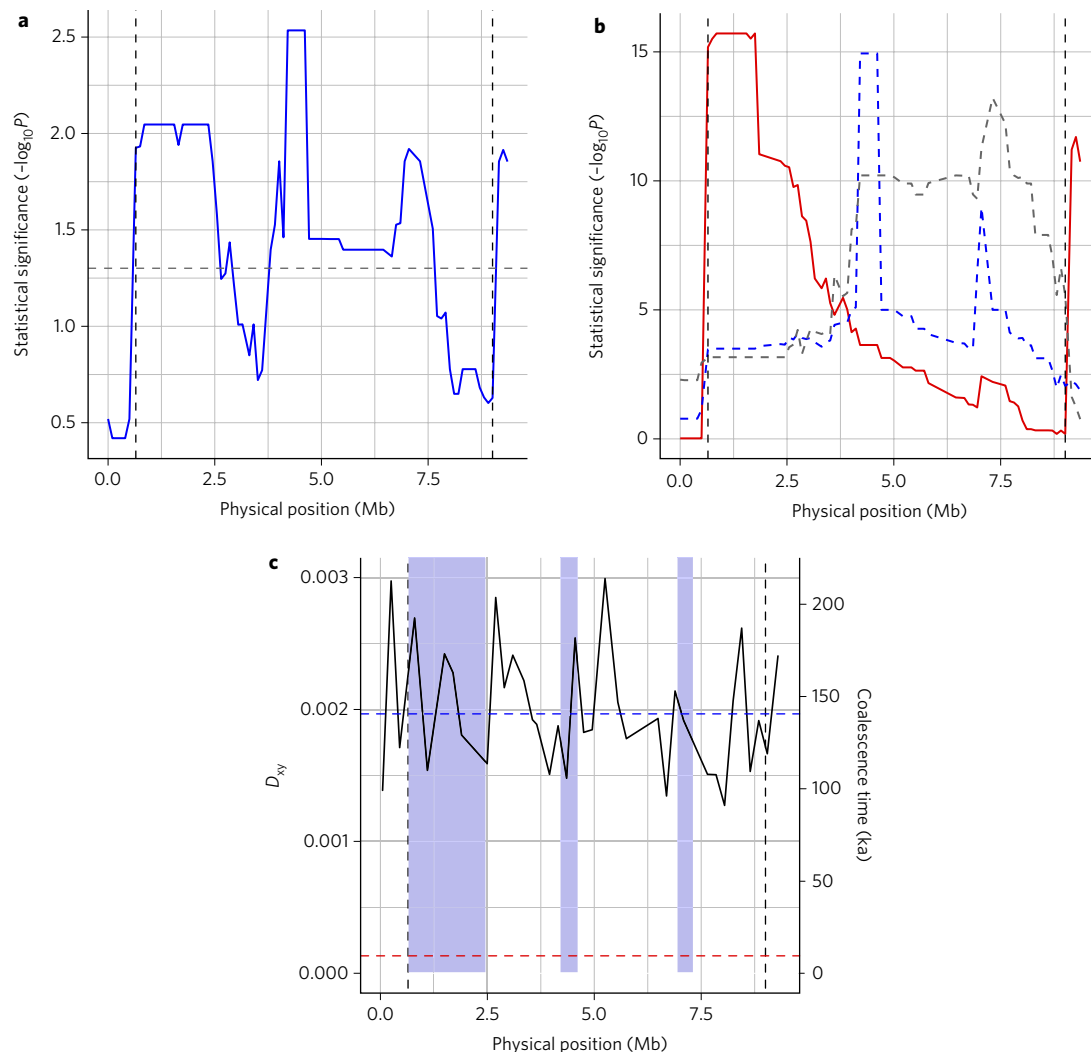
Figure 5 | QTLs within the inversion. A collinear cross shows that several QTLs in the inversion region influence phenology and development traits. Inversion breakpoints are indicated by vertical lines. **a**, Multivariate QTL mapping indicates several QTL peaks (blue line) exceeding the significance threshold (horizontal dashed line); $N = 1,714$ F4 individuals in 153 families. **b**, Testing the hypothesis that three linked QTLs have different pleiotropic effects. Plots for three QTLs (left, centre and right; shown in solid red, dotted blue and dashed grey, respectively) quantify evidence that a locus with different patterns of pleiotropy occurs at each peak. Discriminant function analysis was used to identify new composite trait axes defined at each peak, and evidence for these composite traits was mapped across the inversion region. The left and centre QTLs show little overlap, suggesting different patterns of pleiotropy. **c**, Comparison of molecular divergence and time to coalescence of the *East* and *West* genotypes in the collinear QTL mapping cross. The horizontal red dashed line at 8.8 ka is the upper confidence interval for the age of the inversion. The vertical axes show nucleotide divergence (D_{xy}, left axis), and coalescence time (ka, right axis). The blue dashed line indicates the mean genome-wide values of D_{xy} and the coalescence time between these *East* and *West* alleles. The horizontal axis shows position across the inversion region in Mb, beginning at marker Scaffold26675_2450000. Vertical blue shading indicates the QTL regions $\pm \log P > 1.6$ confidence intervals, with 408 annotated genes in these QTL regions. Divergence (D_{xy}) between these genotypes was calculated in 200-kb non-overlapping windows, and the coalescence time was estimated using $T = D_{xy}/2\mu$, where μ is 7×10^{-9} per site per generation. Mean D_{xy} in the three QTLs (left to right) were 0.00206, 0.00169 and 0.00186, corresponding to coalescent times of 147 ka, 121 ka and 132 ka, respectively. The mean D_{xy} in the whole inversion region was 0.00191, and the average coalescence time was 136 ka.

package[35], the final edited map was aligned to pseudomolecules of *B. stricta* and *Capsella rubella*[36]. Synteny with *A. lyrata* was visualized using MUMmer[37].

The integration of the WGP physical map with the genetic map and sequence scaffolds (via the WGP tags and BAC-end sequencing) showed high concordance among these three genomic resources. WGP enabled regions with low recombination in the linkage map to be ordered and oriented on the basis of their correspondence with the physical map contigs. In addition, we were able to integrate some scaffolds that were not linked in the genetic map, providing more robust pseudomolecules.

Variant discovery. We aligned each genotype read to the *B. stricta* LTM hardmasked reference (B. strict_278_hardmasked) with BWA[38]. We used GATK[39] for base quality recalibration, indel realignment, simultaneous SNP and indel discovery via HaplotypeCaller and joint genotyping using the default hard filtering parameters as prescribed by GATK best practice[39].

Ordering and orienting scaffolds into linkage groups. We created a genetic map based on 159 sixth-generation RILs bred from two parent individuals: LTM (the reference individual) and SAD12. We (re-)sequenced LTM to a genomic depth of ~400× using paired-end Illumina sequencing. These reads were aligned to the reference sequence using bwa mem[40] and samtools mpileup[41] to visualize the alignment. We conservatively verified 105,074,494 positions as homozygous by requiring unambiguous alignment of a single base type with a depth ranging from 160–600. Next, we performed a similar analysis on a set of paired-end Illumina reads from SAD12 (depth ~170×) and compiled a catalogue of 442,637 discriminatory positions throughout the genome. These markers were homozygous for different bases in LTM and SAD12, allowing identification of ancestry within the RILs. These results imply a nucleotide divergence of 0.000425 between these two accessions.

We initially aligned sequences from 199 barcoded RILs, sequenced to modest depth (a few fold), to the reference sequence and genotyped these at all sites

possible within the discriminatory catalogue. During this process, we eliminated a number of contaminated RILs, RILs with too little sequence for reliable genotyping, and RILs that appeared to be heterozygous over most of the genome, suggesting that the source DNA had not originated from a single RIL. The scaffolds in the assembly with sizes ≥20 kb were binned into 20-kb blocks and each block with at least three genotypeable sites was genotyped as either LTM, SAD12 or heterozygous on the basis of the consensus of genotypeable sites. A catalogue of breakpoints was constructed, as a list of scaffolds and bins where one or more RILs changed genotype. From this data, we identified 9 redundant RILs, and after removal of these, we were left with 159 RILs to use in the analyses. In addition, we found two misjoins in the original assembly, which identified themselves as being apparently unlinked neighbouring bins. These were Scaffold25219, between bin 134 and 137 (removed bins 135,136 and renamed bins 137–139 to Scaffold100000 bins 0–2) and Scaffold19424 between bins 344 and 345 (renamed Scaffold19424 bins 345–391 to Scaffold200000 bins 0–46). In the final data set, 2,664 crossovers were located (16.7 per RIL), and 2.4% of all blocks were genotyped as heterozygous (an average of 3.12% is expected for F6 RILs, with a large variance).

Next, we calculated a matrix of recombination fractions, r, between each pair of bins, defined as the fraction of RILs for which the bins differ in genotype. Bins in each pair with $r < 0.15$ were clustered by a single-linkage approach. (Briefly, all pairs of bins were compared, and pairs with $r < 0.15$ were assigned to the same cluster. Two clusters were merged if any pair of members had $r < 0.15$.) This algorithm resulted in seven distinct linkage groups. The bins in each linkage group were used as markers for input into MSTmap[42] (v.1). Finally, individual bins within assembled scaffolds were rearranged into the order in which they occur in the scaffolds, and in the orientation, if known, consistent with the genetic map. This step is necessary as the sizes of the 20-kb bins are often smaller than the map resolution, so that the bins appear in random order on the map. During this step, the cumulative physical length of the map (in bases) was also inserted into the map. In addition, occasional unmapped bins, the position of which could be inferred from the context of adjacent mapped bins, were inserted at the appropriate locations (Dryad Data Archive). Finally, from the matrix of markers and genotypes, we inferred the locations of recombination events near and within the inversion (Supplementary Fig. 1).

Identification of centromere positions in the *B. stricta* reference genome. To identify the centromere position for each *B. stricta* linkage group, we downloaded sequences of 16 *A. thaliana* centromeric BAC clones and looked for their homologues in the *B. stricta* reference genome using BLAST. These 16 clones were mapped on pericentromeric regions of *B. stricta* by chromosome painting[43], thus the positions of their homologues on *B. stricta* linkage groups indicate the margins of centromeres. The lengths of predicted centromere regions were 2–3 Mb for linkage groups 1, 2, 5, 6 and 7, and 6–7 Mb for linkage groups 3 and 4. As expected, all of these predicted centromere regions showed low recombination rates, except for the parts of these regions on linkage groups 3 and 4 (Supplementary Fig. 4 and Supplementary Table 6).

Identification of a paracentric inversion by comparative chromosome painting. *Painting probes.* Previous F2 genetic mapping in the LTM × SAD12 F2 population found an extensive block of recombination suppression on LG1 (Bs1). On the basis of the published karyotype structure of *B. stricta* SAD12[43], we designed BAC-based painting probes to verify the position and orientation of ancestral genomic blocks on the seven chromosomes of *B. stricta* LTM by comparative chromosome painting. The *A. thaliana* BAC contigs used as painting probes are listed in ref. [44]. Chromosome-specific BAC contigs were arranged and differentially labelled following the organization of genomic blocks in the reconstructed karyotype of *B. stricta*[43].

While chromosomes Bs2 to Bs7 have identical structures, chromosome Bs1 differs between the two *B. stricta* accessions. In LTM, the upper arm of Bs1 was restructured via a paracentric inversion of approximately 8.4 Mb (~4% of the genome). To more precisely map the two inversion breakpoints to ~100 kb (approximate size of *Arabidopsis* BAC clones) we carried out BAC fluorescence *in situ* hybridization experiments on pachytene chromosomes of the accessions, SAD12 and LTM. The following *Arabidopsis* BACs were used for mapping of the A1 breakpoint: F21M11 (AC003027), F20D22 (AC002411), F19P19 (AC000104), T1G11 (AC002376), F13M7 (AC004809), T7A14 (AC005322), T25N20 (AC005106), F3F20 (AC007153), T20M3 (AC009999) and T21E18 (AC024174). The A1 breakpoint was mapped between BACs F19P19 (1,125,000–1,230,000 bp) and T1G11 (1,233,000–1,333,000 bp). The following BACs were used to narrow down the C1 breakpoint: F27J15 (AC016041), F27K7 (AC084414), T27P22 (GenBank accession missing), F11I4 (AC073555), F9P7 (AC074308), T1N15 (AC020889), F11A17 (AC007932), F21D18 (AC023673), T2J15 (AC051631) and T6B12 (AC079679); the C1 breakpoint was mapped between BACs: F9P7 (17,975,000–17,987,000 bp) and F11I4 (17,982,000–18,062,000 bp). The *A. thaliana* BAC clone T15P10 (AF167571) bearing 35S rRNA genes was used for chromosomal localization of nucleolar organizer regions. Clone pCT 4.2 (a 500-bp 5S rRNA repeat (M65137)) was used to identify 5S rDNA loci.

Chromosome preparations. Chromosome spreads were prepared according to a previously published protocol[45] with minor modifications. Entire inflorescences were fixed overnight in ethanol:acetic acid (3:1) and then stored in 70% ethanol at 20 °C until further use. Before chromosome spreading, closed flower buds with white (young) anthers were rinsed with distilled water and then with citrate buffer (4 mM citric acid and 6 mM trisodium citrate, pH 4.8) and digested using a pectolytic enzyme mixture (0.3% cellulase, cytohelicase, and pectolyase; all Sigma) in citrate buffer at 37 °C for 3–6 h, and then kept in citrate buffer until used. An individual flower bud was put on a microscope slide and dissected using needles in a drop of citrate buffer to form a fine suspension. Then 15 to 30 μl of 60% acetic acid was pipetted into the cell suspension, which was then spread over the slide. The slide was placed on a hot plate at 50 °C for ~0.5–2.0 min. The spread chromosomes and nuclei were then fixed by pipetting 100 μl of ethanol:acetic acid (3:1) fixative around the suspension drop. The slide was tilted to remove the fixative, dried using a hairdryer and postfixed with 4% formaldehyde in distilled water for 10 min and left to air dry.

Nick translation. DNA probes were labelled by nick translation with biotin-dUTP, digoxigenin-dUTP or Cy3-dUTP using the following mixture: 1 μg of DNA in 29 μl of distilled water, 5 μl of nucleotide mixture (2 mM deoxyadenosine triphosphate, deoxycytidine triphosphate and deoxyguanosine triphosphate, and 400 μM deoxythymidine triphosphate; all Roche), 5 μl of 10× NT buffer (0.5 M Tris–HCl (pH 7.5), 50 mM MgCl₂ and 0.05% bovine serum albumin), 4 μl of 1 mM custom-made *x*-dUTP (in which *x* is biotin, digoxigenin or Cy3), 5 μl of 0.1 M β-mercaptoethanol, 1 μl of DNase I (0.008 U μl⁻¹; Roche), and 1 μl of DNA polymerase I (10 U μl⁻¹; Fermentas). This nick translation mixture was incubated at 15 °C for 90 min (or longer) to obtain fragments of ~200–500 bp. The reaction was stopped by addition of 1 μl of 0.5 M EDTA (pH 8.0) and by incubation at 65 °C for 10 min. The labelled probes were stored at −20 °C until use.

Fluorescence *in situ* hybridization and comparative chromosome painting. Before pipetting the probe onto selected slides, the slides were treated with pepsin (0.1 mg ml⁻¹; Sigma) in 0.01 M HCl for 3–6 min, postfixed in 4% formaldehyde in 2× saline sodium citrate buffer (300 mM sodium chloride and 30 mM trisodium citrate, pH 7.0) for 10 min, dehydrated in an ethanol series (70, 80 and 96%; 3 min each), and air dried. The labelled BAC clones (and rDNA probes) were pipetted together and the DNA was precipitated by adding a 0.1 volume of 3 M sodium acetate (pH 5.2) and a 2.5 volume of ice-cold 96% ethanol, kept at −20 °C for at least 30 min, and centrifuged at 13,000 *g* at 4 °C for 30 min. The pellet was dried using a desiccator and resuspended in 20 μl of hybridization buffer (50% formamide and 10% dextran sulfate in 2× SSC) per slide. The probe and chromosomes were denatured together on a hot plate at 80 °C for 2 min and hybridization was carried out by placing the slides in a moist chamber at 37 °C for 40–63 h. Post-hybridization washing was performed in 20% formamide in 2× SSC at 42 °C. Signal detection and amplification were achieved as follows: biotin-dUTP was detected by avidin–Texas Red (1:1,000; Vector Laboratories) and amplified using goat anti-avidin–biotin (1:200, Vector Laboratories) and avidin–Texas Red; digoxigenin-dUTP was detected by mouse anti-digoxigenin (1:250; Jackson Immuno-Research Laboratories) and goat anti-mouse Alexa Fluor 488 (1:200; Molecular Probes); Cy3-dUTP was observed directly. DNA was counterstained with 4′,6-diamidino-2-phenylindole (2 μg ml⁻¹) in Vectashield (Vector Laboratories). The hybridization signals were analysed using an Olympus BX-61 epifluorescence microscope equipped with fluorochrome-specific excitation and emission filters (AHF Analysentechnik), and a Zeiss Axio-Cam CCD camera. The monochromatic images were pseudo-coloured and processed using Photoshop CS5 (Adobe).

Mapping SNPs to pseudomolecules. In the genetic map, 208 scaffolds were ordered and oriented into 7 LGs. Misjoins were found in the original assembly of Scaffold25219 and Scaffold19424, which were split into four scaffolds, two with the original names (Scaffold25219 and Scaffold19424) and two with new names Scafffold100000 and Scaffold200000. In Scaffold25219, the first 135 bins (bin: 0–134; position: 1–2,700,000) were found to be unlinked with the last three bins (bin: 137–139; position: 2,740,001–2,797,113 bp), while the position of bins 135–136 (position: 2,700,0001–2,740,000) between them could not be determined. Thus, the first 135 bins were retained in Scaffold25219 (a shorter one; position: 1–2,720,000), the last three bins were moved to a new Scaffold100000 (position, 2,740,001–2,797,113 of Scaffold25219 was changed to position 1–57,113 of Scaffold100000), and bins 135–136 were excluded and not used in map. In Scaffold19424, the first 345 bins (bin: 0–344; position: 1–6,900,000) were found to be unlinked to the remaining 47 bins (bin: 345–391; position: 6,900,001–7,828,947), thus bins 0–134 were kept in Scaffold19424 (position: 1–6,900,000) and the rest were moved into a new scaffold200000 (position 6,900,001–7,828,947 of Scaffold19424 was changed to position 1–928,947 on Scaffold200000).

Accordingly, SNP coordinates were extracted from the vcf file (scaffold name and position on scaffold) and examined as to whether they fell in a bin of the linkage map on the basis of its scaffold position. If so, its position on the pseudomolecules (g) was calculated as $g = A - |g - c|$, where A is the end position of

the bin that the SNP fell in, g is the SNP's position on the scaffold, and c is the end coordinate of the bin on the scaffold. For SNPs that did not fall in any bins but were located on regions 2,740,001–2,797,113 of Scaffold25219 or 6,900,001–7,828,947 of Scaffold19424, their positions on Scaffold100000 and Scaffold200000 were calculated as $g = g_1 - 2{,}740{,}000$ or $g = g_1 - 6{,}900{,}000$, respectively (g_1 is the position on the original scaffold and g is the position on the new scaffold). After that, their positions on pseudomolecules (g') were calculated as described above.

Delimiting and genotyping the inversion. After chromosome painting narrowed down possible locations of the inversion breakpoints in *A. thaliana*, we blasted these *Arabidopsis* genes onto the *B. stricta* LTM genome and identified putative scaffolds containing the two breakpoints. To identify the exact breakpoints, we mapped paired-end Illumina reads from SAD12 onto the LTM scaffolds with BWA[38], followed by *de novo* assembly with Velvet[46] for read-pairs having one or both ends near the putative breakpoint region in the LTM genome.

The primers Inv-A, Inv-B and Inv-E were designed to amplify the inversion region using Primer 3 (ref. [47]). Primer Inv-D was used for Sanger sequencing the breakpoints in the two karyotypes for the confirmation of breakpoints. Locations of the breakpoints and related primers are shown in the Dryad Data Archive.

Thermocycling consisted of heating to 94 °C for 3 min, then 33 cycles of 94 °C for 20 s, 55 °C for 1 min, 72 °C for 30 s, followed by 72 °C for 6 min. Amplified fragments were separated by electrophoresis on a 1% agarose gel stained with SYBR Safe (Thermo Fisher). Fragments were ~300 bp in accessions containing the inversion, and ~600 bp in accessions without the inversion.

Differences between subspecies. From our earlier study of ecological divergence between *East* and *West* subspecies[16], we used least-squares means from 24 genotypes (Supplementary Table 1) to quantify differences in flowering time. In the greenhouse, *East* genotypes flower an average of 5.0 days before *West* genotypes.

Here we report on three *East* × *West* crosses, which all gave healthy advanced generation progeny in F2, F4 or F6 generations. More generally, intrinsic reproductive isolation might be present in some instances, especially if complicated by occasional polyploidy or apomixis[48].

Effect of the inversion on NIL flowering time. In the LTM × SAD12 (MT$_{West-inv}$ × CO$_{East-std}$) cross[17], we generated two F7 NIL families from the F6 population to investigate the inversion effect. Two F6 heterogeneous inbred families (HIFs: 116A and 120A)[49] heterozygous for the inversion and mostly homozygous elsewhere in the genome were self-fertilized, and 144 F7 plants from each parent were grown in a completely randomized design in the greenhouse. After three months, plants were vernalized at 4 °C for six weeks, and flowering time was recorded. We also genotyped the inversion status in all individuals using the inversion primers Inv-A, Inv-B and Inv-C. We used fixed effects ANOVA to test the equation, Flowering Time = Family + Inversion + Family × Inversion, and verified compatibility with parametric statistical assumptions.

For families from this MT$_{West-inv}$ × CO$_{East-std}$ cross, we also analysed lifetime fecundity in the field, using Dryad-archived data (http://dx.doi.org/10.5061/dryad.rp3pc) from the Lost Trail field site in the inversion zone[20]. We analysed lifetime fecundity in an *East* × *West* recombinant inbred F6 population grown in nature for three years.

Effects of the inversion in a sympatric cross. The inversion might be favoured if this rearrangement alters genes around the break points and gives it higher fitness over its ancestral standard haplotypes from the same population. To test this, we generated crosses between the reference accession, LTM (inversion), and the accession, SDM, which has the standard haplotype. These genotypes are of the same subspecies (*West*) and were collected from similar habitats ~1.7 km apart. Two crosses were generated, one with LTM as mother (CL9.1) and the other with SDM as mother (CL10.1). The F1 hybrids were self-fertilized, and 1,000 F2 plants were grown in the greenhouse. The inversion was genotyped in all individuals, and we measured 11 life history traits (Survival after vernalization (binary), flowering (binary), width at 4 weeks, width at 10 weeks, flowering time (days after vernalization), flowering width, flowering height, flowering rosette number, flowering leaf number, fruit number, lifetime fitness and inversion genotype) on 994 individuals.

To test the phenotypic effects of the inversion, we permuted the inversion genotypes in the crosses 1,000 times, and performed a MANOVA on the vector of traits using R[53] (v.3.1.2). Because R only reports a type-I sum of squares for MANOVA, we obtained a type III sum of squares by performing several MANOVAs, and reported the effect of each factor when added to the model last. We modelled Multivariate Traits = Cross + Inversion + Cross × Inversion, with cross type (either CL9.1 or CL10.1 as maternal parent), three inversion genotypes (inversion, standard, or heterozygote), and their interaction as fixed-effect predictor variables. *P*-values were obtained by comparing the true Wilks' Lambda statistic to those from 1,000 permutations (Supplementary Table 4). Because MANOVA showed that multivariate means differed between *Bsi1* genotypes, the univariate traits were analysed by ANOVA with a $P = 0.05$ significance threshold[50].

Gene prediction. We obtained the 10 kb region on each side of each breakpoint in the LTM reference genome (inversion) and created pseudomolecules corresponding to the corresponding regions around each breakpoint in the standard haplotype. All four sequences were submitted to the Augustus gene prediction algorithm (http://bioinf.uni-greifswald.de/augustus/), and we found no open reading frames spanning the breakpoints in either the inversion or standard haplotypes. Thus, the chromosomal inversion does not disrupt existing open reading frames or create new ones (Supplementary Fig. 7). Genes predicted by Augustus largely correspond to the transcriptome-assisted annotation in the *B. Stricta* genome v1.2.

Gene expression. To investigate whether the inversion event changed the expression of flanking genes, we compared two nearby populations using F2 progeny from the LTM × SDM cross (representing inversion and standard haplotypes, respectively). On the basis of the inversion genotypes, we chose 11 F2 plants that were homozygous for the inversion and 12 that were homozygous for standard haplotypes, providing sample sizes well above those recommended by power analysis[51]. All 23 plants were from cross CL9.1 where LTM was the mother. These F2 plants were self-fertilized for one generation, and two-month-old rosette leaves from *Bsi1* homozygous F3 plants were used to compare the expression of seven genes flanking the two breakpoints and syntenic to the region in *A. thaliana*. We used a Sigma Spectrum Plant Total RNA kit to extract the RNA, and a Thermo Scientific DyNAmo cDNA Synthesis kit to synthesize the cDNA. We used a Thermo Scientific DyNAmo SYBR Green qPCR kits for quantitative PCR (primers in Dryad Data Archive). As in previous gene expression studies of *B. stricta*[18], we used *ACTIN2* (*ACT2*) as a reference and calculated the expression as $\Delta C_t = C_{t\,ACT2} - C_{t\,gene}$, where C_t is the cycle threshold, $C_{t\,ACT2}$ is the C_t value for *ACT2* and $C_{t\,gene}$ is the C_t value for each gene. Fold gene expression relative to *ACT2* was calculated as $2^{\Delta C_t}$. Since $2^{\Delta C_t}$ has a skewed distribution, we analysed results from both $2^{\Delta C_t}$ and ΔC_t. For each gene separately, ANOVA was performed with inversion status as a fixed effect (Supplementary Table 5). In addition, we examined the statistical power using JMP Pro v.12.0.1 (SAS) and via simulations in R.

Dissecting the inversion region in a collinear cross. The *East* × *West* 'Parker × Ruby' cross used for QTL mapping was developed from two parents in the *East-West* contact zone: one in Parker Meadow (RP105, Parker, *East* subspecies; 44° 37′ N, 114° 31′ W) and one at Ruby Creek (RP109, Ruby, *West* subspecies; 45° 33′ N, 113° 46′ W). On the basis of nucleotide similarity (Supplementary Fig. 8a, left), the *West-std* parent (RP109) used in this cross was very similar to the most recent common ancestor that gave rise to the inversion and its closest *West-std* relatives (IN086 and IN087). The F1 hybrid was self-fertilized to produce F2 plants, and subsequent generations were propagated by self-fertilization and single-seed descent to create 153 independent genetic lines (families). In each line, multiple F4 progeny from the same F3 plant were used in a randomized complete block design totalling 1,714 individuals. The phenotypic least-squares means were calculated in JMP 8 to represent the genotypic value for their F3 parent.

Genotyping by sequencing. We used genotyping by sequencing (GBS)[52] to identify SNPs in this cross. In each family, DNA from each accession was extracted from 0.1 g of young leaf tissue following a dark treatment of about three days. Tissue was stored at −80 °C, flash-frozen in liquid nitrogen and homogenized prior to genomic DNA extraction using Qiagen DNeasy Plant Mini kits. DNA concentration was measured using a Qubit fluorometer (Turner Biosystems, Invitrogen) from at least ten pooled F4 individuals to represent the genotype of their F3 parent. This protocol uses an adaptor design that is compatible with TruSeq adaptors and indexes while allowing paired-end sequencing. The combination of 48 unique barcodes with four different TruSeq indexes allowed multiplexing of 192 samples (153 F3:F4 families, 19 inbred individuals of the Parker parent, and 20 inbred individuals of the Ruby parent). For sequencing, we used an Illumina HiSeq-2000 or HiSeq-2500 at the Duke Genome Sequencing and Analysis Core Resource. From this population we obtained ~249 million reads with unambiguous barcodes (Sequence Read Archive accession, SRP075905). Read pairs were assigned to genotypes and parents by custom Perl code, and low-quality bases in the end of reads were trimmed. SNPs were called using GATK Best Practices (above). In addition, these protocols were used to assay GBS SNPs (Sequence Read Archive, accession SRP075997) in 122 genotypes across the inversion area.

Because the LTM × SAD12 cross does not recombine in the inversion on chromosome 1, we used the Parker × Ruby cross to infer the linkage map for bins within the inversion. The mean sequencing depth of GBS SNPs in the Parker and Ruby parents was ~20×, with a standard deviation of ~40×. We focused on 52,827 bi-allelic SNPs where the two parents had sequencing depths ≥6× and ≤100× (the mean depth plus two standard deviations), and were homozygous for different alleles. We binned the genome into 100 kb windows, and for each individual we counted the number of reads from the two parents, and calculated the proportion of Parker-derived reads. For an individual, if the cumulative depth of all SNPs in a window was less than 20×, it was classified as missing. We removed windows with ≥30% missing data. We used a hard-genotype cutoff based on the proportion of Parker reads: a Parker proportion ≤0.2 was defined as homozygous Ruby, a Parker proportion ≥0.7 was defined as homozygous Parker, and the remaining windows

were classified as heterozygous. These cutoffs gave the correct hard-genotype proportions for an F3 population (about 37.5% of either homozygote and 25% heterozygous genotypes). After determining the 'hard genotype calls', we calculated the proportion of the three genotypes in each window. To remove windows with excessive proportions of heterozygotes or either homozygote, we excluded windows where the proportion of the three genotypes was beyond the upper or lower 5% of their respective whole-genome distribution. Given this was a cross in the F3 generation, which had only experienced two generations of recombination, we regarded it as extremely improbable that two recombination events would occur within 1 Mb of each other. Therefore, in each individual, if two recombination events were identified within 1 Mb, we changed the genotype calls of markers between two breakpoints to missing. The linkage map was constructed with MSTmap. With only three windows (markers) dropped out, MSTmap constructed seven linkage groups, with scaffold–chromosome assignment that was consistent with the LTM×SAD12 cross. The final linkage map was refined in multiple steps: we first removed markers that were >5 cM away from both flanking markers and re-constructed the map. In the new map, we imputed a missing marker genotype if the upstream and downstream markers with data had the same genotype call and were not both >10 cM away from the missing position. Another linkage map was constructed, and the imputation process was repeated since the new map had slightly changed the order of some markers, making some missing genotypes now imputable. The final linkage map was then constructed with 1,010 markers. Thus, we used the Parker×Ruby cross to infer the linkage map for bins within the inversion and we inferred the position of polymorphisms within the inversion from their positions in each ordered and oriented contig.

This collinear *East×West* 'Parker×Ruby' greenhouse experiment consisted of 12 blocks, each with one F4 plant from each of the 153 lines and multiple Parker and Ruby individuals ($N = 1,714$). Seeds were stratified in 4 °C for four weeks and planted in 'Cone-tainers' (Ray Leach SC10; Stuewe and Sons), with soil composition and greenhouse conditions as previously described[16]. When rosettes were 11 weeks old, all leaves from three blocks of plants were harvested for rosette- and leaf-morphology measurements as described[16]. At 12 weeks of age, the remaining nine blocks were vernalized at 4 °C for 6 weeks, then returned to the same greenhouse conditions for phenology measurements. All traits were measured as previous described[16], except that (1) no physiological traits were measured and (2) leaf width/length ratio was used instead of leaf shape morphometrics because the leaf-shape landscape points were highly correlated[16] (Supplementary Table 7).

For QTL mapping, all individual-level measurements were transformed to family-level least-squares means in JMP ($N = 153$ families). Due to the skewed distribution of most phenotypes, all traits (except binomial traits or plant stages) were log-transformed at the individual level. For greenhouse measurements, Block and Genotype were considered as random effects. We used the R libraries qvalue and MASS for MANOVA, false discovery rate and discriminant function analysis. For each marker across the inversion interval and adjacent contigs, we fitted a MANOVA model: Multivariate Phenology Traits = Block + Single Marker Genotype. At each marker we computed the Wilks' Lambda, permuted the vector of phenotypes with respect to marker genotypes 1,000 times, and finally corrected for false discovery rates across the inversion region[54].

DFA was performed in R for a marker at the peak of each multivariate QTL (bins: Scaffold-25219_50000, Scaffold13671_1150000, and Scaffold26675_450000), defining the direction of trait variation along the axis of greatest differentiation among the multivariate means of these marker genotypes. Subsequently, at these three markers we used the eigenvector for the first DFA axis to compute the phenotypic projection of each genotype on the trait axis identified by DFA. These new trait values were used for univariate QTL mapping across the inversion region, with correction for false discovery rates across the inversion region.

Finally, we annotated SNPs using snpEff[55] v.4.2 and found 17 genes that are orthologues of the flowering-time genes of *A. thaliana*[56] within the inversion interval. See Supplementary Table 8 for the candidate genes and the Dryad Data Archive for the annotated SNPs.

Population genetics. *SNP filtering.* SNPs called by GATK were filtered in each of three groups: sets consisting of 35 inversion genotypes (G_{INV}) and 87 standard genotypes (G_{STD}) from the inversion zone, and a 'Reference Population' set of 83 genotypes from across the species range, outside the inversion zone (G_{RP}). Next, data from these single groups were pooled together into two datasets, the complete dataset ($G_{INV}+G_{STD}+G_{RP}$) and the inversion zone genotypes ($G_{INV}+G_{STD}$). In each dataset, genotypes supported by less than two reads were assigned as missing, and SNPs were discarded if they met the following criteria in any of the groups: (1) detected in fewer than 50% of individuals; (2) mean depth >20; (3) more than one variant allele was observed; (4) sites where the proportion of heterozygous genotypes was >15% (*B. stricta* is predominantly inbred[17], hence high heterozygosity may indicate paralogous loci); (5) reference or variant alleles were indels. After filtering, 75,737 and 43,722 SNPs were retained for the $G_{INV}+G_{STD}+G_{RP}$ and $G_{INV}+G_{STD}$ sets, respectively. The following analyses used these SNPs.

Ancestral allelic state. Several population genetic statistics require information on the ancestral state of segregating variants. Therefore, we sequenced a *Boechera holboellii* (*B. holboellii*; reference genotype 'Panther'; location: 45° 18.198′ N, 114° 22.599′ W) individual with a mean depth of ~400× (Joint Genome Institute project ID: 1051698). Short reads were mapped to the *B. stricta* reference genome using BWA, genotypes were called using GATK with default settings, and sites were filtered out if they had a depth of <20 or >800, or if they were heterozygous. A Python (v.2.7.9) script assigned ancestral alleles as the shared allele between *B. holboellii* and *B. stricta*. In total, we obtained information on the ancestral state for 63,182 (83.4%) of the 75,737 $G_{INV}+G_{STD}+G_{RP}$ SNPs and for 35,100 (80.3%) of the 43,722 $G_{INV}+G_{STD}$ SNPs.

Windows. We partitioned the genome into 300 kb windows. The sequence coverage of each window was calculated by counting the number of available sites (both variant and non-variant sites obtained from GATK using the '-allSites' argument) per window passing the following quality filters: detected in at least 50% of individuals, mean depth ≤20, and proportion of heterozygous genotypes ≤15% in each of the three groups (G_{INV}, G_{STD}, and G_{RP}). Windows were distributed on a per-scaffold basis, beginning at position 1 of a scaffold, and were then oriented along pseudomolecules according to the linkage map. After excluding windows with sequence coverage <5 kb, 2,801 windows were retained for downstream analyses.

Phylogenetic relationship and population structure analyses. Python scripts were used to generate alignments from genotypes in the vcf file, with missing and heterozygous loci coded as 'N'. With genome-wide SNPs, we verified the *East-West* population structure using principal component analysis in EIGENSOFT[57] v.6.0.

To investigate the origin of the *Bsi1-inv* haplotype we used data from 901 SNPs in the inversion region of the genome to examine relationships among 35 *Bsi1-inv* and 83 *Bsi1-std* haplotypes. (Four admixed *Bsi1-std* genotypes were excluded.) First, we constructed neighbour-joining trees using MEGA[58] v.6.06 with 1,000 bootstrap steps. Second, maximum-likelihood (ML) trees were constructed using RAxML[59] v.8.0.0. One thousand ML trees were generated to find the best-scoring ML tree, and topological robustness was investigated using 1,000 non-parametric bootstrap replicates. Neighbour-joining and ML trees were displayed using FigTree[60].

Site frequency spectrum and summary statistics. To control for error rates and variable coverage of short-read sequencing, we used a probabilistic method implemented in ANGSD[61] v.0.911 to estimate the site frequency spectrum (SFS) and related population genetic statistics. A number of filtering steps were performed: (1) removed reads with a minimal mapping quality of 30 and bases with a minimal quality score of 30 (-minMapQ and -minQ); (2) removed sites with information from less than 50% of individuals (-minInd); (3) removed sites with a *P*-value higher than 1×10^{-6} (-snp_pval); (4) assigned genotypes as missing if the depth was less than two for an individual; (5) only used genotypes with a posterior probability higher than 0.95; (6) removed sites that did not pass previous filtering criteria (above). Because *B. stricta* is predominantly inbreeding[17], we estimated inbreeding coefficients for each individual in ngsTools[62] v.1.0 and incorporated them into the calculation of SFS in ANGSD. Using genotype likelihoods based on the GATK genotyping model[63], we estimated folded and unfolded SFS and derived a set of population genetic summary statistics in 300 kb windows. For the G_{INV} and G_{STD} groups, we estimated Tajima's *D*[64] on the basis of both folded and unfolded SFS, and calculated Fay and Wu's H^{29} and Fu and Li's *D* and *F*[65] using the unfolded SFS. The inferred ancestral allelic state (based on short reads from the *B. holboellii* genotype 'Panther') aligned to the *B. stricta* reference genome was used to estimate the unfolded SFS. We compared population genetic statistics in the inversion region, on other chromosomes (LG2 to LG7), and in Block D[21], a chromosome region with unusually high polymorphism in related species, perhaps due to clusters of genes for nucleotide-binding site-leucine-rich repeat proteins[66].

F_{ST}, D_{xy}, D_A and π. Genetic differentiation (F_{ST})[67] between G_{INV} and G_{STD} groups was estimated using VCFtools[68] v.0.1.12 and we used Python scripts to estimate nucleotide diversity (π)[69] within each group, pairwise nucleotide divergence (D_{xy})[70] and net pairwise nucleotide divergence (D_A)[70] between groups. All parameters were calculated on a per-site basis, and then averaged to obtain window-based estimates (300 kb windows). The window-based F_{ST} was calculated by averaging per site F_{ST} across all variable loci, and the window-based π, D_{xy} and D_A were estimated by averaging all sites (both variable and monomorphic) passing the initial quality filters for each window (above).

Linkage disequilibrium. The level of linkage disequilibrium (LD) between the inversion and all SNPs on chromosome 1 (including those in the inverted region) was estimated as the mean-squared correlation (r^2) for 118 samples from inversion zone accessions (35 *Bsi1-inv* and 83 *Bsi1-std*) using plink[71] v.1.90 . (Four admixed *Bsi1-std* samples were excluded; see above.) In addition, to compare LD between the inversion and flanking SNPs, we identified ten randomly chosen comparator SNPs from LG2 to LG7, having similar derived allele frequencies (25–35%; close to the 29.7% frequency of *Bsi1-inv* in the inversion zone), and then calculated LD

between them and flanking SNPs along the chromosomes. For LG1 (Fig. 5), strong LD ($r^2 > 0.4$) between the inversion and SNPs within and outside the inverted region extended to the end of the chromosome (more than 10 Mb). In contrast, LD with the ten comparator SNPs from LG2 to LG7 declined quickly to background levels (Supplementary Fig. 8b).

Age of the inversion. To estimate the age of the inversion, we phased the 35 *Bsi1-inv* genotypes using Beagle[72] v.4.1 with default settings, and estimated pairwise genetic distances among haplotypes on the inverted genome region. After that, the divergence time between haplotypes was calculated as $T = d/2\mu$, where d is pairwise genetic distance and μ is the mutation rate. We used a mutation rate of 7×10^{-9} per site per generation in *A. thaliana*[24] with a generation time of two years in *B. stricta*. To avoid biases due to missing data, we only considered 1,675 haplotype pairs with less than 50% missing sites in the inversion region. We approximated the lower bound for the age of the inversion on the basis of the mean pairwise distances[73], and the upper bound from the maximum of all pairwise distances among inversion genotypes.

Divergence time between QTL alleles in the collinear cross. Using the same mutation rate and generation time, we estimated $D_A = D_{xy} = 2\mu T$, and solved for T between these two alleles in 200-kb non-overlapping windows[71]. This window size was chosen to ensure sufficient polymorphic sites within each interval.

Frequency distribution of derived alleles. We tested whether *Bsi1-inv* had unusually high frequency among derived SNP alleles in the inversion zone population. For this analysis, we excluded one admixed, genetically divergent inversion individual (IN019; Supplementary Fig. 9c, right), and analysed *Bsi1-inv* and closely related *Bsi1-std* genotypes from the inversion zone (Supplementary Fig. 9c, left; first principal component < 0.0). These genotypes comprise 54 samples (34 *Bsi1-inv* and 20 *Bsi1-std*), with 2,416 SNPs that, like the inversion, segregate among these genotypes and are monomorphic in the rest of our collection.

Code availability. All code is available in the Dryad Data Archive at http://dx.doi.org/10.5061/dryad.tt743.

References

1. Kirkpatrick, M. How and why chromosome inversions evolve. *PLoS Biol.* **8**, e1000501 (2010).
2. Huber, B. *et al.* Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity* **114**, 515–524 (2015).
3. Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**, e1000500 (2010).
4. Corbett-Detig, R. B. & Hartl, D. L. Population genomics of inversion polymorphisms in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1003056 (2012).
5. Kirkpatrick, M. & Kern, A. Where's the money? Inversions, genes, and the hunt for genomic targets of selection. *Genetics* **190**, 1153–1155 (2012).
6. Guillen, Y. & Ruiz, A. Gene alterations at *Drosophila* inversion breakpoints provide *prima facie* evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genom.* **13**, 53 (2012).
7. Smith, A. C. *et al.* Maternal gametic transmission of translocations or inversions of human chromosome 11p15.5 results in regional DNA hypermethylation and downregulation of *CDKN1C* expression. *Genomics* **99**, 25–35 (2012).
8. Kennington, W. J., Partridge, L. & Hoffmann, A. A. Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster* are indicative of coadaptation. *Genetics* **172**, 1655–1663 (2006).
9. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
10. Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358 (2001).
11. Noor, M. A. F., Grams, K. L., Bertucci, L. A. & Reiland, J. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl Acad. Sci. USA* **98**, 12084–12088 (2001).
12. Navarro, A. & Barton, N. H. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* **57**, 447–459 (2003).
13. Lohse, K., Clarke, M., Ritchie, M. & Etges, W. Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution* **69**, 1178–1190 (2015).
14. Huang, C. H. *et al.* Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Mol. Biol. Evol.* **33**, 394–412 (2016).
15. Lee, C.-R. & Mitchell-Olds, T. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Mol. Ecol.* **20**, 4631–4642 (2011).
16. Lee, C.-R. & Mitchell-Olds, T. Complex trait divergence contributes to environmental niche differentiation in ecological speciation of *Boechera stricta*. *Mol. Ecol.* **22**, 2204–2217 (2013).
17. Anderson, J., Lee, C.-R. & Mitchell-Olds, T. Life history QTLs and natural selection on flowering time in *Boechera stricta*, a perennial relative of *Arabidopsis*. *Evolution* **65**, 771–787 (2010).
18. Prasad, K. *et al.* A gain-of-function polymorphism controlling complex traits and fitness in nature. *Science* **337**, 1081–1084 (2012).
19. Heo, J.-Y. *et al.* Identification of quantitative trait loci and a candidate locus for freezing tolerance in controlled and outdoor environments in the overwintering crucifer *Boechera stricta*. *Plant Cell Environ.* **37**, 2459–2469 (2014).
20. Anderson, J. T., Lee, C.-R. & Mitchell-Olds, T. Strong selection genome-wide enhances fitness tradeoffs across environments and episodes of selection. *Evolution* **68**, 16–31 (2014).
21. Lysak, M. A., Mandakova, T. & Schranz, M. E. Comparative paleogenomics of crucifers: Ancestral genomic blocks revisited. *Curr. Opin. Plant Biol.* **30**, 108–115 (2016).
22. Mehringer, P. J., Arno, S. F. & Petersen, K. L. Postglacial history of Lost Trail Pass Bog, Bitterroot Mountains, Montana. *Arct. Alp. Res.* **9**, 345–368 (1977).
23. Mumma, S. A., Whitlock, C. & Pierce, K. A 28,000 year history of vegetation and climate from Lower Red Rock Lake, Centennial Valley, Southwestern Montana, USA. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **326**, 30–41 (2012).
24. Ossowski, S. *et al.* The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94 (2010).
25. Anderson, J., Lee, C.-R., Rushworth, C., Colautti, R. & Mitchell-Olds, T. Genetic tradeoffs and conditional neutrality contribute to local adaptation. *Mol. Ecol.* **22**, 699–708 (2013).
26. Anderson, J. T., Inouye, D. W., McKinney, A. M., Colautti, R. I. & Mitchell-Olds, T. Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proc. R. Soc. B* **279**, 3843–3852 (2012).
27. Smadja, C. M. & Butlin, R. K. A framework for comparing processes of speciation in the presence of gene flow. *Mol. Ecol.* **20**, 5123–5140 (2011).
28. Guerrero, R. F., Rousset, F. & Kirkpatrick, M. Coalescent patterns for chromosomal inversions in divergent populations. *Phil. Trans. R. Soc. B* **367**, 430–438 (2012).
29. Fay, J. C. & Wu, C.-I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
30. Naseeb, S. *et al.* Widespread impact of chromosomal inversions on gene expression uncovers robustness via phenotypic buffering. *Mol. Biol. Evol.* **33**, 1679–1696 (2016).
31. Chapman, J. A. *et al.* Meraculous: *de novo* genome assembly with short paired-end reads. *PLoS ONE* **6**, e23501 (2011).
32. van Oeveren, J. *et al.* Sequence-based physical mapping of complex genomes by whole genome profiling. *Genome Res.* **21**, 618–625 (2011).
33. Soderlund, C., Humphray, S., Dunham, A. & French, L. Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res.* **10**, 1772–1787 (2000).
34. Nelson, W. & Soderlund, C. Integrating sequence with FPC fingerprint maps. *Nucleic Acids Res.* **37**, e36 (2009).
35. Soderlund, C., Bomhoff, M. & Nelson, W. M. SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res.* **39**, e68 (2011).
36. Slotte, T. *et al.* The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* **45**, 831–835 (2013).
37. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
39. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

40. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at http://arxiv.org/abs/1303.3997 (2013).

41. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

42. Wu, Y., Bhat, P. R., Close, T. J. & Lonardi, S. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet.* **4**, e1000212 (2008).

43. Mandáková, T., Schranz, M. E., Sharbel, T. F., de Jong, H. & Lysak, M. A. Karyotype evolution in apomictic *Boechera* and the origin of the aberrant chromosomes. *Plant J.* **82**, 785–793 (2015).

44. Mandáková, T. & Lysak, M. A. Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (Brassicaceae). *Plant Cell* **20**, 2559–2570 (2008).

45. Lysak, M. A. & Mandáková, T. *Analysis of Plant Meiotic Chromosomes by Chromosome Painting* (Humana Press, 2013).

46. Zerbino, D. R. & Birney, E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).

47. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).

48. Rushworth, C. A., Song, B. H., Lee, C. R. & Mitchell-Olds, T. *Boechera*, a model system for ecological genomics. *Mol. Ecol.* **20**, 4843–4857 (2011).

49. Tuinstra, R. M., Ejeta, G. & Goldsbrough, B. P. Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theor. Appl. Genet.* **95**, 1005–1011 (1997).

50. Scheiner, S. M. in *Design and Analysis of Ecological Experiments* (eds Scheiner, S. M. & Gurevitch, J.) 99–115 (Chapman and Hall, 2001).

51. Schurch, N. *et al.* How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* **22**, 839–851 (2016).

52. Cande, J., Andolfatto, P., Prud'homme, B., Stern, D. L. & Gompel, N. Evolution of multiple additive loci caused divergence between *Drosophila yakuba* and *D. santomea* in wing rowing during male courtship. *PLoS ONE* **7**, e43888 (2012).

53. R Core Team. *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2013).

54. Benjamini, Y. & Yekutieli, D. Quantitative trait loci analysis using the false discovery rate. *Genetics* **171**, 783–790 (2005).

55. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. *Fly* **6**, 80–92 (2012).

56. Bouche, F., Lobet, G., Tocquin, P. & Perilleux, C. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res.* **44**, D1167–D1171 (2016).

57. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

58. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).

59. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

60. Rambaut, A. FigTree (2016); http://tree.bio.ed.ac.uk/software/figtree

61. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

62. Fumagalli, M., Vieira, F. G., Linderoth, T. & Nielsen, R. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486–1487 (2014).

63. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

64. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).

65. Fu, Y.-X. & Li, W.-H. Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).

66. Clark, R. M. *et al.* Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338–342 (2007).

67. Weir, B. & Cockerham, C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).

68. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

69. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. USA* **76**, 5269–5273 (1979).

70. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, 1987).

71. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

72. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).

73. Long, Q. *et al.* Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* **45**, 884–890 (2013).

Acknowledgements

Author contributions

C.-R.L. and T.M.-O. conceived this work. C.-R.L., T.M.-O., Y.Y., D.K., J.Z. and J.L.G. designed the study. T.M. and M.A.L. performed the chromosome painting. C.-R.L., K.V.S.K.P., N.P., H.N.H., D.K., Y.Y., J.T., W.G., J.Z., J.G. and J.J. worked on the molecular biology and sequencing. S.F., U.H., J.W.J., J.S., D.S.R., R.W. and K.B.; and J.L.G., D.K., Y.Y., J.Z., J.T., W.G. and R.W. planned and analysed the genomic and physical mapping experiments, respectively. C.-R.L., B.W., J.P.M., U.H., J.L.G. and T.M.-O. performed bioinformatic and evolutionary analyses. K.G., C.-R.L. and N.P. performed the experiments with plants. C.-R.L., B.W., J.P.M., M.E.S. and T.M.-O. drafted the manuscript. All authors read, revised and approved the manuscript.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.-R.L. or T.M.-O.

How to cite this article: Lee, C.-R. *et al.* Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat. Ecol. Evol.* **1**, 0119 (2017).

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

Competing interests

The authors declare no competing financial interests.

Erratum: Young inversion with multiple linked QTLs under selection in a hybrid zone

Cheng-Ruei Lee, Baosheng Wang, Julius P. Mojica, Terezie Mandáková, Kasavajhala V. S. K. Prasad, Jose Luis Goicoechea, Nadeesha Perera, Uffe Hellsten, Hope N. Hundley, Jenifer Johnson, Jane Grimwood, Kerrie Barry, Stephen Fairclough, Jerry W. Jenkins, Yeisoo Yu, Dave Kudrna, Jianwei Zhang, Jayson Talag, Wolfgang Golser, Kathryn Ghattas, M. Eric Schranz, Rod Wing, Martin A. Lysak, Jeremy Schmutz, Daniel S. Rokhsar and Thomas Mitchell-Olds

This Article should have been published under a Creative Commons licence according to the Nature policy on publishing the primary sequence of an organism's genome for the first time. The manuscript is now open access and published under a CC-BY licence. All versions of the Article have been modified accordingly.

Publisher correction: Young inversion with multiple linked QTLs under selection in a hybrid zone

Cheng-Ruei Lee, Baosheng Wang, Julius P. Mojica, Terezie Mandáková, Kasavajhala V. S. K. Prasad, Jose Luis Goicoechea, Nadeesha Perera, Uffe Hellsten, Hope N. Hundley, Jenifer Johnson, Jane Grimwood, Kerrie Barry, Stephen Fairclough, Jerry W. Jenkins, Yeisoo Yu, Dave Kudrna, Jianwei Zhang, Jayson Talag, Wolfgang Golser, Kathryn Ghattas, M. Eric Schranz, Rod Wing, Martin A. Lysak, Jeremy Schmutz, Daniel S. Rokhsar and Thomas Mitchell-Olds

In Fig. 5 of the version of this Article originally published, the final number on the *x* axes of each panel was incorrectly written as 1.5; it should have read 7.5. This has now been corrected in all versions of the Article.