RESEARCH ARTICLE SUMMARY

PLANT GENETICS

Natural selection drives emergent genetic homogeneity in a century-scale experiment with barley

Jacob B. Landis, Angelica M. Guercio, Keely E. Brown, Christopher J. Fiscus, Peter L. Morrell, Daniel Koenig*

INTRODUCTION: Plant domestication by ancient farmers typically occurred in a limited geographic area, but trade and migration dispersed these early crops across entire continents. Once introduced to a new locale, crop plants were confronted with different environments and farming practices. Genetic variation segregating within each field facilitated rapid adaptation by targeting advantageous alleles under local conditions. Understanding the process of environmental adaptation in crops and the number and types of genes that drive it has important agronomic implications in a rapidly changing world. However, the long generation times of large, multicellular organisms make it difficult to study the process of adaptation directly.

RATIONALE: We characterized the effect of natural selection on genetic diversity during adaptation over the course of a long-term competition experiment, barley composite cross II (CCII). CCII was founded in 1929 from thousands of distinct recombinant genotypes and has been grown in Davis, California, USA, for more than 58 generations. We sequenced the genomes of the parents of CCII and characterized the evolutionary trajectory of millions of genetic variants over the course of the experiment. We explored the origin of the alleles that were favored in the experiment, searched for loci targeted by unusually strong selection, and linked genetic shifts to phenotypic change over time.

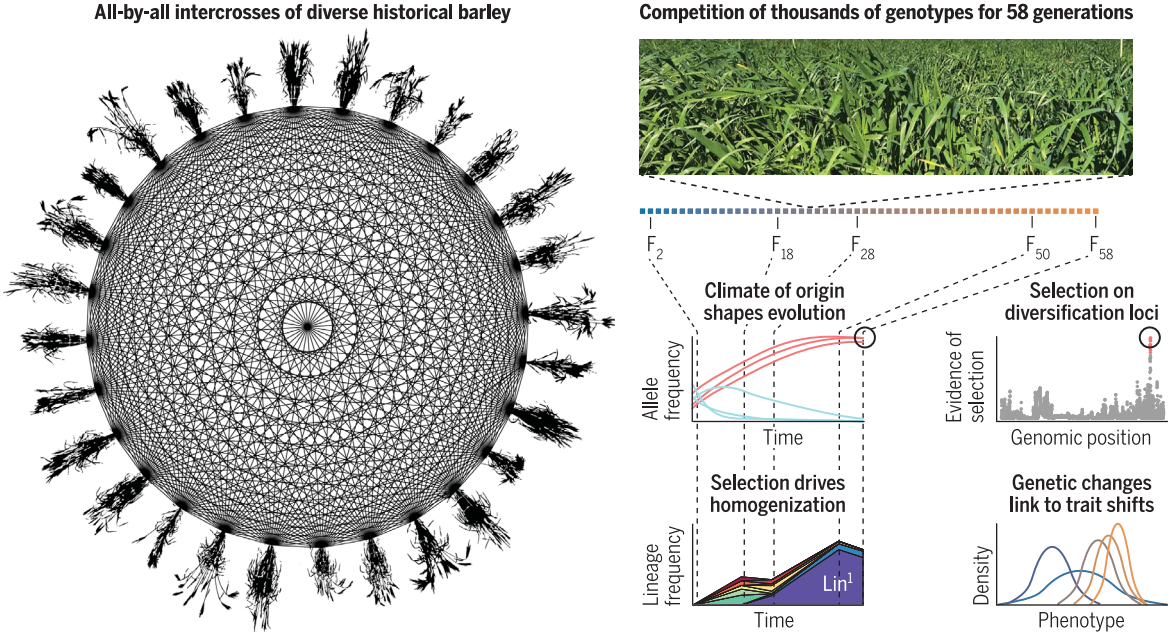
RESULTS: CCII initially segregated for nearly all the common genetic variation in barley. However, CCII lost this variation rapidly, putting the population on a path to near complete homogenization. Genetic diversity has been driven out of CCII by the ascendance of a single dominant lineage that constitutes over half of sampled individuals in later generations. Adaptation to local environmental conditions has played a major role in this process, with surviving alleles being drawn from ancestors that grew in climates similar to Davis. Loci targeted by strong directional selection often played a role in reproductive development and included some of the most well-characterized barley diversification loci: *Vrs1*, *HvCEN*, and *Ppd-H1*. Selection favored alleles that accelerate completion of the life cycle at the latter two loci, presumably to avoid the dry summer season. However, in this study, we show that stabilizing selection shapes reproductive timing on longer timescales, with selection eventually removing the earliest flowering genotypes by targeting another locus, *Vrn-H2*.

CONCLUSION: We found that natural selection has dominated the evolution of CCII. The speed of wholesale genetic restructuring in CCII suggests that the magnitude of selection in a real-world environment is larger than typically assumed. However, the degree to which diversity was lost genome-wide is almost certainly due to linked selection. In CCII, this factor has led to the extinction of all but a small number of lineages. Thus, even in the presence of extremely high founding diversity, populations of self-fertilizing plants may be doomed by selection to genetic homogeneity. Despite the clear signal of environmental adaptation in CCII, yield increases did not keep pace with pedigree-based breeding approaches. An improved understanding of the trade-offs between local adaptation, plant competitiveness, and crop yield could aid in the development of more-resilient crop varieties and improve our understanding of plant evolution. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: dkoenig@ucr.edu
Cite this article as J. B. Landis *et al.*, *Science* **385**, eadl0038 (2024). DOI: 10.1126/science.adl0038

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adl0038

Rapid adaptation in CCII. We used a long-term evolutionary experiment competing thousands of recombinant barley over the course of the last century to understand the genomic and phenotypic basis of adaptation in a real-world environment.



RESEARCH ARTICLE

PLANT GENETICS

Natural selection drives emergent genetic homogeneity in a century-scale experiment with barley

Jacob B. Landis[1]†, Angelica M. Guercio[1], Keely E. Brown[1], Christopher J. Fiscus[1], Peter L. Morrell[3], Daniel Koenig[1,2]*

Direct observation is central to our understanding of adaptation, but evolution is rarely documented in a large, multicellular organism for more than a few generations. In this study, we observed evolution across a century-scale competition experiment, barley composite cross II (CCII). CCII was founded in 1929 in Davis, California, with thousands of genotypes, but we found that natural selection has massively reduced genetic diversity, leading to a single lineage constituting most of the population by generation 50. Selection favored alleles originating from climates similar to that of Davis and targeted loci contributing to reproductive development, including the barley diversification loci *Vrs1*, *HvCEN*, *Ppd-H1*, and *Vrn-H2*. Our findings point to selection as the predominant force shaping genomic variation in one of the world's oldest biological experiments.

The survival of cultivated plants after their dispersal is a classic example of rapid adaptive evolution (*1*). Early farmers brought newly domesticated crops to wildly different environments, subjecting them to strong natural and artificial selective forces. These early crop varieties were phenotypically and genetically diverse, providing ample raw material to facilitate adaptation (*2*). Ultimately, evolution resulted in locally adapted varieties that formed the backbone of early human civilization.

The history of the Neolithic founder crop barley (*Hordeum vulgare*) is a prototypical example of this process. Barley is a self-fertilizing, annual plant domesticated over 10,000 years ago and was dispersed from the Fertile Crescent to become a major source of nutrition for humans and livestock throughout Europe, Asia, and Northern Africa over just a few thousand generations (*1*). Studies of collections of early cultivars have uncovered some of the population genetic history of the crop (*3*–*6*) and mapped loci that contributed to its spread (*7*–*9*). However, precise estimates of the important parameters of genetic adaptation, such as the number and types of genes involved and the magnitude of selection at these loci, are limited without direct observation (*10*).

We used a long-term competition experiment in barley to observe the process of local adaptation over decades. Composite cross II (CCII) is a multigenerational common garden experiment started in 1929 to adapt a genetically diverse population to the environmental conditions of Davis, California, USA (*11*, *12*). Twenty-eight varieties were selected to represent the ecological, phenotypic, and geographical diversity of barley (table S1) and were crossed in all possible combinations, omitting reciprocal crosses (half-diallel) to generate 20,000 recombinant F_2 progeny. Progeny from each cross were combined in equal proportion to found the experiment. Each year the population was propagated, 5,000 to 20,000 seeds from CCII were sown, allowed to compete with minimal human intervention, and harvested in bulk (*13*). The harvested seeds were used in the following year to continue the experiment, and live seeds of early generations were maintained by less-frequent propagation of parallel lineages (Fig. 1 and fig. S1). Barley predominantly reproduces by self-fertilization (98 to 99%) (*14*), meaning that after the early generations, CCII is expected to be primarily composed of genetic lineages equivalent to recombinant inbred lines with a limited number of progeny resulting from recent hybridization.

Comparison of early and late generations of CCII has previously revealed substantial shifts in fitness-associated traits (*12*, *15*–*17*). These include changes in the timing of reproduction, increases in plant biomass, and increases in seed size and number, indicating that natural selection was an important force shaping phenotypic diversity in CCII. In this study, we leveraged modern sequencing technologies to characterize the genetic underpinnings of adaptation over a half-century of evolution.

The genetic diversity of CCII founders

We began by sequencing each of the genomes of the 28 CCII parents to a coverage depth of 8 to 15× and identified 12,922,667 variants that were segregating at the founding of the experiment. Segregating variation in the founders included 64.8% of 1,316,845 single-nucleotide polymorphisms (SNPs) segregating in a global survey of barley genetic diversity (*9*), including 96.4% of common alleles (allele frequency > 0.1, fig. S2); the estimated allele frequencies of genetic variants co-segregating in the parents and global datasets were well correlated (Spearman's $\rho = 0.81$, $P < 2 \times 10^{-16}$, fig. S3). Principal components analysis (PCA) showed that CCII parents are well dispersed among global barley diversity along the first two PCs, representing 20.2% of the variance in the dataset (Fig. 1A and fig. S4). We conclude that the traditional cultivars selected to found CCII in the 1920s reflect the distribution of common genetic variation in barley.

Rapid evolution in CCII

To understand how genetic diversity in CCII changed over time, we traced the evolutionary trajectory of each polymorphism from the founding population to generations F_{18}, F_{28}, and F_{58} (fig. S1) by sequencing pools of 1000 individuals at each time point. Genetic diversity of CCII, as measured by mean expected heterozygosity (H_e) across all sites and the shape of the allele frequency spectrum, changed rapidly throughout the experiment (Fig. 2A and fig. S5). The rate of decrease of H_e was constant [$-0.003\ H_e$ per generation or 1.4% H_e^0 per generation, $P < 0.0012$, coefficient of determination (R^2) = 0.996] with a predicted genome-wide average of 0 by generation 71 at our sequencing depth. Allele frequency in the founding population was a strong predictor of eventual fixation genome-wide (figs. S5 and S6), with 78% of all alleles and 93% of rare alleles [minor allele frequency (MAF) < 0.05] undetectable in F_{58}. Near complete loss of genetic variation ($H_e < 0.01$) was initially restricted to a few regions of the genome (0.7% in F_{18}); however, by generations F_{50} and F_{58}, 29.8% of the genome was near fixed at our sequencing depth. The average genome-wide F_{ST} was 0.18 between the founders and F_{58} (fig. S7). Genome-wide F_{ST} was linearly associated with generational time increasing at a rate of 0.0027 ΔF_{ST} per generation ($P < 0.003$, R^2 = 0.88, fig. S7). CCII lost genetic variation at a rate similar to that of neutral simulations conducted with population sizes two orders of magnitude smaller than the actual size of the experiment (Fig. 2A), much faster than would be expected because of genetic drift alone.

A genome-wide signal of environmental adaptation

We considered two forms of selection that might be responsible for the unusually rapid genome-wide loss of diversity in CCII: selection against generally deleterious variants or

[1]Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA. [2]Institute for Integrative Genome Biology, University of California, Riverside, CA 92521, USA. [3]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA.
*Corresponding author. Email: dkoenig@ucr.edu
†Present address: Section of Plant Biology and the L. H. Bailey Hortorium, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA.

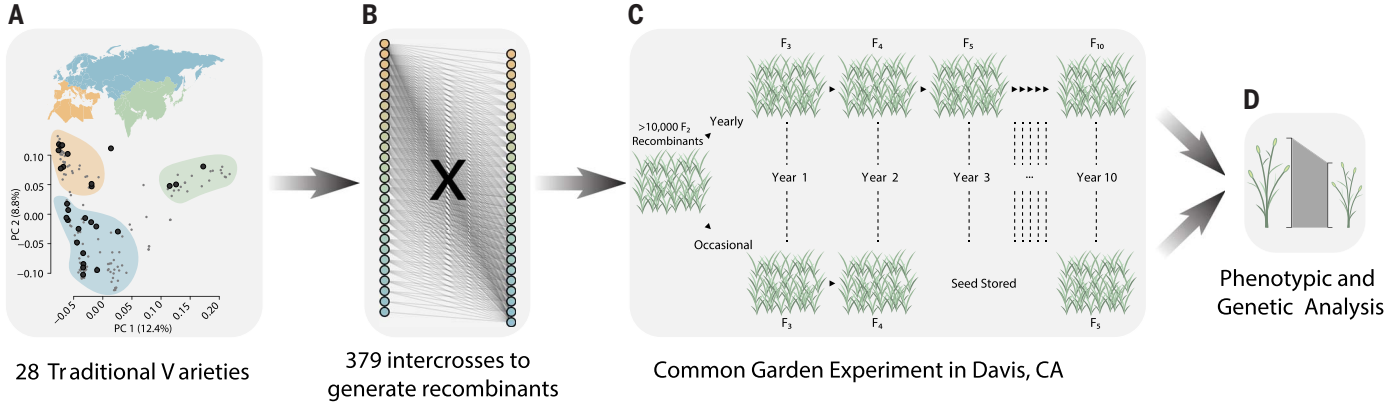


Fig. 1. The design of CCII. (**A**) A PCA of genetic diversity showing the distribution of the 28 CCII parents (dark gray) in a global panel of traditional varieties (light gray). (**B**) The crossing scheme of CCII, in which each of the 28 parents was crossed to each other's parent in a single direction to generate 379 segregating families. (**C**) Illustration of the propagation of CCII with less-frequent grow outs of parallel lineages to retain live seed from earlier generations. (**D**) Seed reserved from previous grow outs can be used to compare changes over evolutionary time in phenotype and genotype.

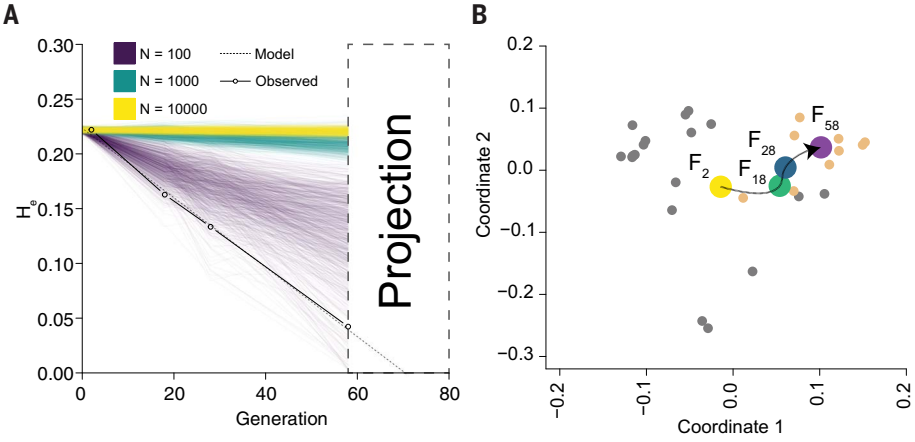


Fig. 2. Rapid adaptive shift in CCII. (**A**) Genome-wide loss of genetic diversity over time in CCII compared with 1000 simulations of neutral evolution in the population, using three different sizes. (**B**) Multidimensional scaling analysis showing the evolution of CCII relative to the parents over time. Orange points indicate parents from Mediterranean climates in North Africa, and gray points indicate all other parents.

positive selection for locally adapted alleles. Whereas deleterious variation is not expected to show strong regional differences among barley varieties, locally adapted alleles would presumably be more common in parental lines from environments similar to Davis.

To distinguish between these possibilities, we examined the evolution of CCII over time by calculating Nei's genetic distance between genome-wide allele frequencies at each time point and the genotypes of each parental accession (fig. S8). We then used a multidimensional scaling procedure [see supplementary materials (SM), materials and methods] to determine the relationship of each CCII generation to the parents (Fig. 2B). CCII quickly evolved toward North African parents that grow in Mediterranean climates similar to that of Davis. Examination of the genetic distances over time revealed this process to be dynamic, with the population most similar to the North African variety Arequipa collected in Peru at time point F_{18}; but in the later stages of the experiment, alleles derived from the variety Atlas were favored (fig. S8). Atlas is a selection from the traditional variety brought by Spanish colonists to California in the late 18th century (*18*).

The success of Mediterranean alleles was also evident for rare alleles that rose in frequency in the population. Minor alleles that increased in frequency were much more likely to be found in North African parents (fig. S8). Alleles identical to Atlas comprised 90.8% of fixed minor alleles (N = 779,883) and 82.2% of fixed private alleles (N = 10,739) in generation F_{58} (table S3).

These results provide a picture of adaptation resembling historical experiments with barley that predate CCII (*19*). These experiments were designed to permit competition among barley varieties, rather than recombinants, at many sites throughout the United States. Maladapted varieties rapidly decreased in number, allowing several varieties to increase in initial generations, but as time passed, competition between the remaining lines led to a new round of extinction. It is notable that alleles from the local variety, Atlas, show evidence for increased fitness in the population even though Atlas was only brought to California around a century and a half before the start of the experiment (*20*).

The emergence of a dominant lineage in CCII

The paucity of genetic diversity found in generation F_{58} suggested that relatively few genetic lineages may have survived into the later stages of CCII. To understand the composition of the genomes of individual progeny, we conducted a genotyping by sequencing (GBS) experiment on CCII parents and 878 total progeny sampled across generations F_{18}, F_{28}, F_{50}, and F_{58}. After alignment, SNP calling, and filtering, we identified 263,238 sites suitable for further analysis.

Consistent with large-scale loss of genetic diversity in CCII over time, we found that the average genome-wide genetic distance between lines fell dramatically in later generations (Fig. 3A and fig. S9). After hierarchical clustering of lines according to genetic distance, we found nearly identical whole-genome haplotypes at all time points that have risen to high enough frequency (from an initial 1/30,000) to be sampled multiple times (Fig. 3B).

Of the 878 sequenced progeny across all four generations, we identified only 261 distinct lineages by these criteria. These lineages are functionally equivalent to recombinant inbred lines. Of these, 177 (68%) occurred just once in our sample. The remainder occurred more than once, with most individuals (445) belonging to just eight clusters. One lineage, Lin^1, was found at exceptionally high frequency, making up 34.9% of all sampled individuals (Fig. 3B).

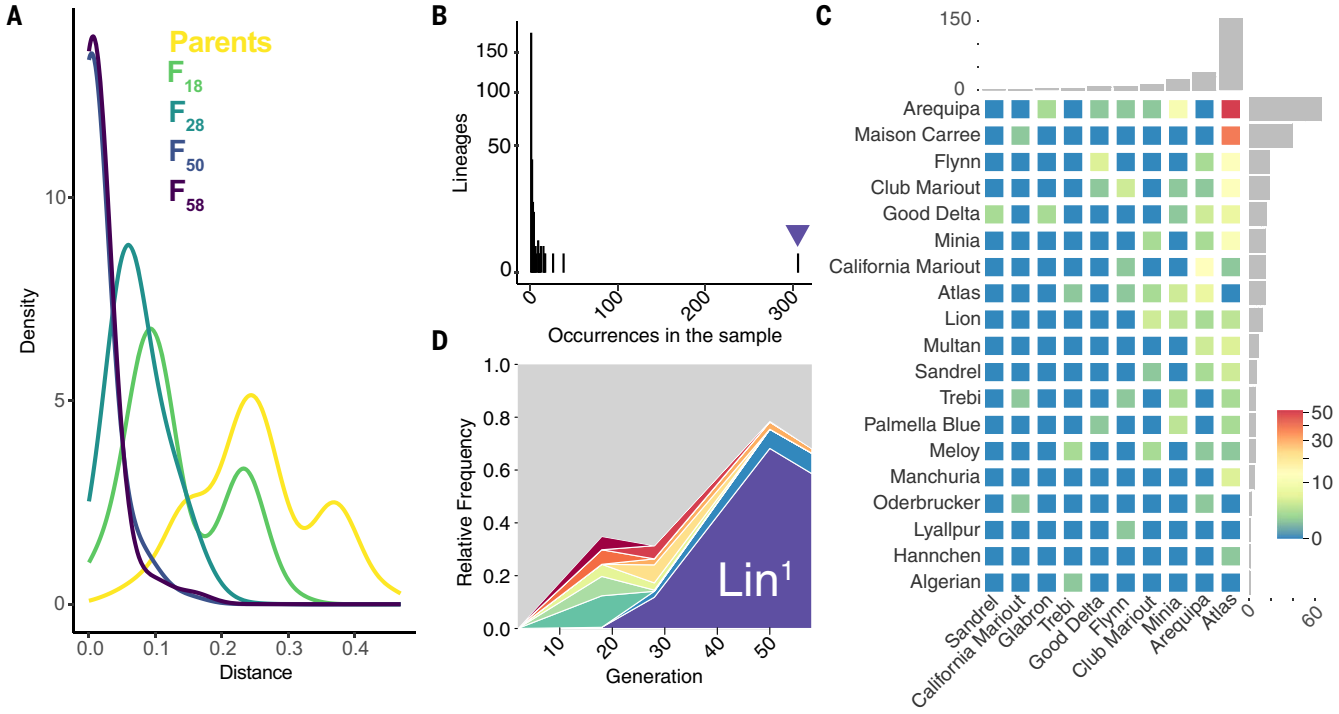


Fig. 3. The rapid rise of select lineages in CCII. (**A**) Density plots of the distribution of pairwise genetic distances within generation across five time points in CCII. (**B**) Nearly identical lineage abundance in CCII combined across all sampled progeny from F_{18}, F_{28}, F_{50}, and F_{58}. The purple arrowhead marks Lin^1. (**C**) Primary (*x* axis) and secondary (*y* axis) ancestry for the CCII lineages across all generations. The adjacent bar plots indicate the sum of individuals that show a primary or secondary relationship with each observed parent, and the colors of the heatmap indicate the number of individuals with ancestry from a particular combination of parents. (**D**) Muller plot showing the proportion of sampled individuals belonging to the top 10 most common lineages over time. The gray portion of the plot represents the sum of all other lineages.

PCA of representative samples from each multilocus haplotype cluster revealed a close relationship between the most successful clusters and North African–derived parents (fig. S9). To understand which specific parents contributed to successful lineages over time, we inferred the dominant ancestry of each multilocus haplotype. We assigned each lineage a primary and secondary parent, defined as those that shared the greatest and second-greatest fraction of the line's genome in near identity [genetic distance (D) < 0.01]. Ten of the original 28 parents were assigned a primary relationship with at least one lineage, and 19 were assigned a secondary relationship (Fig. 3C). Six of the eight primary parents identified in this analysis came from the North African group. All nine North African varieties were identified as secondary parents of at least one lineage. The Californian historical variety, Atlas, was the primary parent relationship in 59.7% of lineages, a massive enrichment from the expected 3.6% (χ^2 test, $P < 1 \times 10^{-46}$). Potential Atlas and Arequipa progeny were the most common (19.5%), with Atlas and Maison Carre being a close second (14.9%). The most abundant lineage, Lin^1, fell into this latter group (fig. S10). These patterns confirm that a relatively limited number of lineages, predominantly derived from Mediterranean-adapted parents, were able to succeed in CCII.

The abundances of each of the 261 lineages were not static over evolutionary time. The number of sampled lineages decreased in later generations (93 in F_{18}, 93 in F_{28}, 38 in F_{50}, and 55 in F_{58}). This drop in diversity was largely the result of a dramatic increase in the abundance of Lin^1 over time, rising from a frequency of 0.005 in F_{18} to a frequency >0.586 in our sample of generations F_{50} and F_{58} (Fig. 3C). A subtle increase in the total number of lineages was seen between F_{50} and F_{58}, but this change was not statistically significant ($\chi^2 = 3.4904$; $P = 0.06172$). The runaway success of Lin^1 appears to be driving out most of the diversity in CCII, putting the population on a path to genetic homogeneity within the coming decades.

The targets of natural selection in CCII

We next sought to pinpoint the genes targeted by directional selection during rapid adaptation in CCII. Strong selection at a locus can leave a footprint of depleted genetic diversity and strong allele frequency change surrounding the targeted site relative to the rest of the genome (*21*). Because genetic variation in CCII had become so depleted by F_{58}, we focused on identifying selected loci in the first phase of the experiment by comparing F_2 and F_{18}.

For 25,844 sliding windows (1000 SNP windows, 500 SNP step size) covering the barley genome, we calculated the combined probability of the observed H_e and mean change in allele frequency (between F_2 and F_{18}) in a set of 100,000 simulations of neutral evolution (see SM, materials and methods). After false discovery rate correction, we identified 58 genomic regions that showed significant evidence for selection (Fig. 4 and figs. S11 to S14). These regions ranged from 162 kb to more than 42 Mb in size, with a median size of just under 1 Mb (928.691 kb, table S2), covering a total of 3.5% of the barley genome.

Several compelling candidates overlapped with regions of the genome targeted by natural selection. *Vrs1*, a homeobox gene that controls a major dimorphism in inflorescence architecture in barley (*22*), overlapped with a 2.06-Mb selected region on the long arm of chromosome 2H (Fig. 4, table S2, and fig. S15). Wild barley and some cultivated barley have inflorescences that produce two rows of seeds (two-rowed), but plants that carry loss-of-function mutations of the *Vrs1* gene produce up to six rows of seeds (six-rowed or intermedium types). Seven of the 28 CCII parents carried two-rowed alleles (table S1; four from

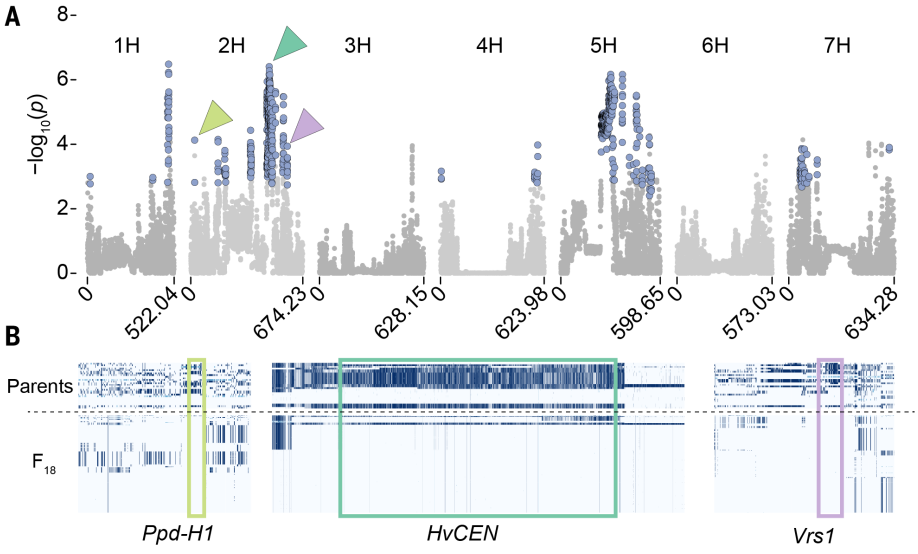


Fig. 4. The footprint of selection in CCII. (**A**) Manhattan plot showing $-\log_{10}(p)$ of significant selected regions (light blue points) over the first 18 generations of CCII. *P* values were calculated in 1000 SNP windows by using the joint probability of the observed H_e and mean absolute change in allele frequency (ΔAF) in 100,000 simulations of neutral evolution for each window. (**B**) Highlighted selected regions with the significant region shown as a box. Parental haplotypes are shown in the top row and F_{18} in the lower row. For visualization purposes, SNPs were polarized against the Atlas genotype across the region so that lighter colors indicate similarity to a North African parent.

Europe, one from North Africa, and two hybrids), but by F_{18} the population is fixed for the six-rowed allele *vrs1.a1* in our pooled-sequencing data. This allele contains a single–base pair deletion that generates a frameshift in the last exon of *Vrs1*. Ultimately, the success of this allele led to the extinction of the two-rowed phenotype in the population (*15*, *16*).

The frequency of six-rowed types in cultivated barley is often attributed to artificial selection by early farmers (*22*). However, these results indicate that natural selection at Davis for the six-rowed type is likely very strong, perhaps because of the larger number of seeds produced per head. Several other genes overlapping selected regions had homology to factors that affect floral development in plants (table S2). These included genes characterized in rice (*OsLa*, *OsMADS22*, *OsLBD12*, *OsLG3*, *MINI ZINC FINGER2*, *Shattering1*, and *ABERRANT PANICLE ORGANIZATION1*) and in the plant model system *A. thaliana* (*AINTEGUMENTA*, *KRYPTONITE2*, and *UNICORN*), pointing to inflorescence development as an important target of selection in the experiment (*23*–*32*). However, these genes have not been characterized in barley, and the evidence that they were directly targeted by selection should be treated with caution.

We also found that the signal of selection overlapped with genes that play a role in the timing of reproduction, a key adaptive trait in barley (*33*). The most well-studied are the pseudoresponse regulator gene *PHOTOPERIOD1* (*Ppd-H1*) and the phosphatidyl ethanolamine–binding protein–encoding gene *HvCENTRORADIALIS* (*HvCEN*). *Ppd-H1* and *HvCEN* regulate the timing of flowering via the day-length sensing and autonomous genetic pathways, respectively (*34*, *35*). Early flowering ancestral alleles were nearly fixed at both loci in F_{18} (Fig. 4B and figs. S16 and S17). The late-flowering phenotype of derived alleles of *HvCEN* is thought to result from a proline-to-alanine change at position 135 in the protein sequence. The late-flowering allele at *HvCEN* was found at intermediate frequencies in the parents (12/28), but pooled-sequencing data found only limited evidence of its continued segregation in the progeny (1/39 in F_{18}, 1/17 in F_{28}, and 0/21 in F_{58}). Similarly, a late-flowering allele at *Ppd-H1* was found in 8/28 parents but was no longer detected in the later generations. It should be noted that an allelic series at *Ppd-H1* regulating flowering time with differing effect sizes has been proposed, although in our dataset it seems that just one allele predominates (*36*). Other selected regions overlapped with homologs of genes involved in flowering time in other species, including *OsLF* and *Hd3a BINDING REPRESSOR FACTOR2* (*HBF2*) (*37*, *38*). These last two genes have yet to be studied in barley, and their involvement in adaptation remains preliminary. Also, even for the best-studied candidates listed above, it is possible that polymorphism in multiple genes within a region might contribute to adaptation.

Two unusually large selected regions were found on chromosome 2H and chromosome 5H (table S2 and fig. S14). The first overlapped with *HvCEN* and a known large-inversion polymorphism that segregates in barley (*39*). The exceptionally large footprint that we identified in our analysis (29.2 Mb) is likely the result of suppressed recombination from the inversion. The largest selected region (42.9 Mb) was found in the pericentromere of chromosome 5H. We did not identify a previously characterized candidate gene in this region, but it colocalizes with a region that appears to have been targeted by strong selection in barley breeding lines over the course of the past century (*40*). It remains to be determined whether the strong signal of local adaptation in these regions might be driven by more than one gene.

The selected regions identified here have become nearly fixed in just a few decades without conscious human-mediated selection. This suggests that adaptive alleles that emerged during early agriculture could dominate a locale during a farmer's lifetime.

The genetic basis of stabilizing selection in CCII

We next asked how genetic shifts in CCII translated into adaptive phenotypic change. We focused on the timing of reproduction because of its general importance in plant fitness (*41*) and because our selection scans uncovered several candidate genes associated with this process.

The appropriate timing of reproduction is a critical contributor to barley local adaptation (*42*). Flowering too early can expose delicate inflorescences to winter frosts and limits the accumulation of vegetative biomass. However, late flowering risks failure to set seeds before the onset of the hot, dry summer season. Previous work with CCII indicated that these combined forces favor intermediate flowering times, a process known as stabilizing selection (*12*, *17*, *43*).

We confirmed these results by observing CCII progeny from F_{18}, F_{28}, F_{50}, and F_{58} alongside the 28 parents in our greenhouses. Median flowering time (as estimated by days to awn emergence) fell modestly from the midparent values by 5.6 days (Mann-Whitney test, $P < 1 \times 10^{-7}$) over the first 18 generations with a notable reduction in variance due to the extinction of late-flowering outliers (*F* test, $F = 1.14177$, $P < 0.001$; Fig. 5 and fig. S19). After generation F_{18}, flowering time slowly rose to just above the founding mean (+1.4 days comparing F_2 with F_{58}, +7 days comparing F_{18} with F_{58}; Mann-Whitney test, $P < 1 \times 10^{-6}$) with greatly reduced variance (*F* test, $P < 1 \times 10^{-13}$; Fig. 5). In our experiments with continued watering throughout the plant's lifetime, flowering time was positively correlated with fitness-related traits, including plant height and fecundity (Fig. 5B and figs. S18 and S20; $P = 3.5 \times 10^{-56}$, adjusted $R^2 = 0.62$). Despite the advantages of late flowering, the population did not favor

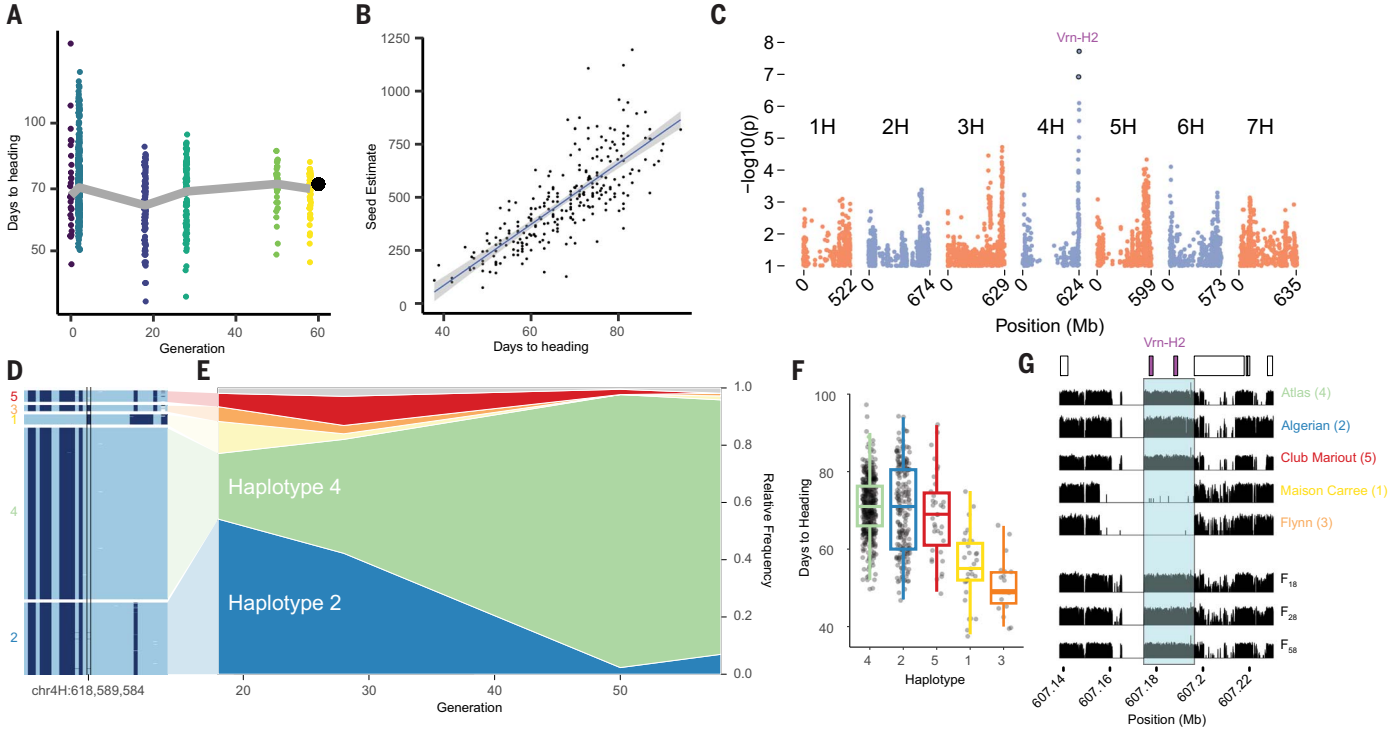


Fig. 5. The dynamic process of purifying selection on reproductive timing. (**A**) Evolution of the time to reproduction as measured by an estimate of heading date (the first appearance of awns in the shoot) in CCII. The F_2 distribution was estimated from midparent means (teal). The black dot shows the value Lin[1]. (**B**) The relationship between days to heading and fecundity based on seed number estimates. (**C**) The GWAS of days to heading shows a single significant peak (Bonferroni corrected $P < 0.05$) on Chr4H. (**D**) Haplotype structure surrounding the *Vrn-H2* locus. The highlighted SNP is the nearest significant SNP to the *Vrn-H2* deletion. (**E**) Muller plot showing haplotype frequency from generation 18 to 58. (**F**) Distribution of days to heading across the five main haplotypes segregating in CCII. (**G**) Coverage of sequencing reads across the *Vrn-H2* locus in the B1K-04-12 *H. vulgare* ssp. *spontaneum* assembly for examples of each haplotype group. The highlighted region (light blue) shows the segregating deletion, which overlaps with two annotated ZCCT genes. The three bottom tracks show coverage for each progeny pool.

the latest-flowering genotypes; instead, the dominant Lin[1] showed an intermediate flowering time (fig. S21). Together, these observations indicate that flowering time in CCII is the target of stabilizing selection, which drives populations toward intermediate phenotypes by removing phenotypic extremes. An initial shift in phenotypic mean is predicted by theory if the trait distribution is skewed (*44*), as is the case here, but stabilizing selection on the timing of reproduction only becomes apparent in CCII on longer timescales (*43*).

The success of alleles derived from Mediterranean-adapted parents at *Ppd-H1* and *HvCEN* is consistent with selection against late-flowering types in CCII (Fig. 4 and figs. S14 and S15). This left us to explore how stabilizing selection could drive a second shift by eliminating the earliest-flowering progeny. We conducted a genome-wide association study (GWAS) to discover loci associated with variation in flowering time. We identified a single significant peak on the long arm of chromosome 4H (Fig. 5C), which we verified using a subset of lines in the following year (fig. S22). The lead SNP in our analysis was 5.42 Mb distal to a well-characterized regulator of adaptive shifts in barley flowering time, *VERNALIZATION2* (*Vrn-H2*). A second significant SNP was 53 kb away from *Vrn-H2* and in linkage disequilibrium (LD) with the lead SNP (R^2 = 0.83). This led us to investigate the possibility that *Vrn-H2* may be driving the shift toward later flowering in the later generations of CCII.

The ancestral barley *Vrn-H2* locus encodes three zinc-finger/CCT domain (ZCCT) transcription factors that repress flowering without prolonged cold treatment, preventing early flowering during the winter (*45*, *46*). A large deletion common in Northern European barley spans all three genes and drives early flowering without extended cold exposure. CCII segregated for eight haplotypes at the *Vrn-H2* locus, five of which were found at an allele frequency >0.05 in at least one of the sampled generations (Fig. 5D). Three North African alleles (haplotypes 2, 4, and 5) conferred late flowering and together were found in 92.4% of the sampled progeny and 81.8% of the F_{18} (Fig. 5, E and F). The two early-flowering alleles, haplotypes 1 and 3, were less common and less geographically restricted, being shared by Mediterranean and Northern European parents (16.4% of F_{18} and 5.9% of all sampled progeny). The late-flowering haplotype 4 from Atlas increased rapidly over time, and by generation F_{58} 88.9% of individuals carried this allele, while early flowering alleles dropped to 2.3% of the sample. The rise of haplotype 4 corresponds to an almost complete loss of the minor allele at the tag SNP in our GWAS.

By realigning the parental whole-genome sequences to a reference that contained a functional ZCCT gene cluster (*H. vulgare* ssp. *spontaneum*, B1K-04-12; fig. S24) (*39*), we found clear evidence for the presence of the ZCCT genes in the three late-flowering haplotypes 2, 4, and 5, and in the pooled-sequencing samples at all time points (Fig. 5G). The two early-flowering alleles lacked coverage across both genes, consistent with a deletion.

Taken together, our results indicate that stabilizing selection tuned flowering time in CCII through a two-step process. First, selection for Mediterranean alleles at *Ppd-H1*, *HvCEN*, and perhaps other loci ensured completion of the life cycle before the onset of the dry season. Second, selection for a functional *Vrn-H2* allele from Atlas eliminated premature flowering in the winter or early spring. The unusual length of the CCII experiment made it possible

to resolve the long-term trend of stabilizing selection that emerges from shorter directional shifts in flowering time.

Discussion

Our analysis of CCII reveals the power of selection to rapidly drive strong directional evolution even in a variable environment, which is consistent with recent results from field experiments with *Drosophila* (*47*). Low genetic diversity in wild populations of self-fertilizing plants is often attributed to founder effects, although linked selection has also been proposed as a potentially important driver (*48*). The latter appears to be the most important force in CCII, with high genetic diversity in the founding population rapidly obscured by the pervasive effects of natural selection. The footprint of early selective events has already begun to decay after only a few dozen generations, suggesting that historical selection may be obscured when divergences are larger than a few tens of generations. However, the length of the CCII experiment allowed us to observe the dynamics of asymmetric stabilizing selection, which would have been disguised as directional selection over shorter periods. These findings highlight the challenge of interpreting patterns of genetic diversity in large multicellular organisms; long-term observation is difficult but crucial. An improved model of the dynamics of adaptive evolution based on observation has great potential to facilitate prediction of patterns of genetic diversity and to aid in the development of strategies for countering the impacts of rapidly changing environments.

CCII was founded to adapt a diverse gene pool to local conditions in Davis (*11*, *13*). Unfit genotypes were expected to go extinct over time, leaving the breeders with useful material to conduct selections for release to farmers. This "evolutionary breeding" approach was also convenient because it did not require breeders to independently maintain many families generated by intercrosses. The adaptation in CCII resembles the process that allowed ancient farmers to adapt early crops to new environments. We found considerable evidence that local adaptation dominates evolution in this experiment. However, despite early, rapid gains in yield in CCII, the evolutionary breeding approach failed to keep pace with the gains observed from pedigree-based breeding methods (*49*). Understanding why the most competitive genotypes produced during local adaptation are not necessarily the highest yielding will be of great interest in the future.

Materials and methods

Plant material

The CCII experiment has been propagated every few years since its inception in population sizes of approximately 5,000 to 25,000 plants. Seed viability for early generations was maintained by propagating less frequently, for example every 5 years rather than every year (Fig. 1). The population was maintained with minimal human intervention excepting planting in the fall and harvest in the spring. The material used for the experiments here was from a grow-out in the 2015 field season. These seed were made available to us by the former curator of the population, Cal Qualset at UC Davis (see SM for further details).

CCII parent whole-genome sequencing

Nuclear genomic DNA was extracted using the CTAB method independently from two plants from each of the 28 parents of CCII (*50*). The first replicate of DNA extractions was used as input for the Illumina TruSeq DNA PCR-Free library preparation kit. Libraries were each sequenced on a single lane of an Illumina HiSeq 4000 instrument on the paired-end 150-bp read setting. To increase the depth of coverage across the parents, libraries were generated from the second replicate using the Nextera DNA Flex library preparation kit. These libraries were pooled and sequenced on an Illumina NovaSeq instrument to generated paired-end 150-bp reads.

Pooled DNA extractions and genome sequencing

One thousand seeds from generations F_{18}, F_{28}, and F_{58} were planted in flats and grown until the two-leaf stage. A hole punch of leaf tissue was taken from each plant and combined across all plants to generate a single pool of tissue that was then ground in liquid nitrogen to homogeneity with a mortar and pestle. A subsample was then taken and used as input for CTAB genomic DNA extraction and sheared to 300-bp fragments on a Covaris S220 Ultrasonicator. A library for each generation was constructed with the NEB Nextseq kit library preparation kit. The three libraries were independently pooled four times and run on four lanes of an Illumina HiSeq 4000. To increase coverage, the parents were resampled and extracted DNA was used to generate sequencing libraries with the Illumina Nextera Flex library preparation kit pooled and sequenced on an Illumina Novaseq instrument.

Sequence alignment and variant calling for CCII parents

The parental and progeny pool DNA sequences were aligned to the Morex 2019 TRITEX barley reference genome assembly (also referred to as v.2) (*51*) using the bwa mem short read alignment algorithm v. 0.7.17-r1188 (*52*) and sorted using samtools v. 1.14 (*53*). Sorted bam files were input into the standard pipeline to identify SNPs using GATK v. 4.1.4.1 (*54*). Initial alignments were screened for PCR duplications using the markduplicates tool in GATK, with the subsequent alignment files being used one chromosome and sample at a time as input for the GATK haplotypecaller tool. The CombineGVCFs tool was used to combine data across parents, and then variant calls were made using the GenotypeGVCFs tool allowing a maximum of two alternate alleles at variant sites. The raw variant dataset from each chromosome was filtered using vcftools v. 0.1.17 (*55*) and bcftools v. 1.15 (*56*) to include only SNPs with no more than four parents having missing data, a minimum genotype quality of five, a variant quality of 30, a maximum of two alleles, no more than six heterozygous calls, and a minimum per sample depth of two. We also required at least one observation of an individual homozygous for the alternate allele. The bcftools concat tool was then used to combine the multiple chromosomes of barley into a single file. Variant sites were annotated using SNPeff (v.5.1) (*57*).

Exome-capture variant calling and analysis with CCII parents

Raw Illumina read data from the barley exome capture panel (*9*) were downloaded from the European Nucleotide Archive and aligned with bwa mem to the Morex 2019 Tritex assembly (*51*). Adapter sequences were trimmed using trim_galore wrapper for cutadapt (v0.6.7) (*58*). Variant calling proceeded according to the standard practices recommended for the GATK tool as described for the whole-genome sequencing for CCII parents. SNP filtering was performed using vcftools v.0.1.16-18 (*55*) with the options –max-missing 0.8 –maf 0.01 –mac 3 –minDP 10 –maxDP 500 –minQ 10 –min-alleles 2 –max-alleles 2.

Genotype calls for each parental accession were made at each polymorphic site in the exome panel using the samtools mpileup and call tools (*56*) and merged with the exome dataset using bcftools merge. The final dataset was filtered to include only domesticated samples from the exome capture dataset, leaving 1,739,465 biallelic SNPs.

PCA of the combined sample was conducted with the software PLINK v1.90b6.25 (*59*) considering the first 25 axes of variation. Plotting and analysis of allele frequencies was performed using the R statistical computing environment (*60*).

Pooled-sequencing allele frequency and diversity estimations

Alignments for pooled-sequencing samples were conducted as described for parental genome sequences. Allele counts at each of the sites polymorphic in the parents were generated for each alignment file using the script extractsite_counts.py which directly extracts the observed nucleotides at each position from the alignment. Count data were then summed across bam files to generate a final allele count dataset for F_{18}, F_{28}, and F_{58}, which was then

combined with allele counts from the parents generated from the filtered VCF file using the script Process_Filtered.py. A final filter was applied to keep sites with a coverage of <300 across all samples and >10 counts for each generation in the R statistical computing environment (*60*).

Allele frequencies p and q in population n at site l were calculated from the count data as follows:

$$p_{n,l} = \frac{cov_{n,l}}{cov_{t,l}}$$

$$q_{n,l} = 1 - p_{n,l}$$

where cov_t is the total coverage of the site, and cov_n is the coverage of the first allele at the site. For genome-wide expected heterozygosity calculation and simulation, only sites with complete data across the parents were considered. Expected heterozygosity for a specific population was calculated at each site as follows:

$$H_l = 2(p_l)(q_l)$$

where H_l is the estimate of expected heterozygosity at site l.

For windowed of genome wide estimates:

$$\bar{H} = \sum_l \frac{H_l}{n}$$

where n is the total number of segregating sites in the window.

Genome-wide pairwise Wright's F_{ST} was calculated as:

$$F_{ST} = \frac{N}{D}$$

where

$$N = p_1(q_2 - q_1) + p_2(q_1 - q_2)$$

$$D = N + p_1q_1 + p_2q_2$$

with n being the specific pool. The mean across sites was used to describe genome-wide F_{ST}.

To understand the relationships between the population and the parents over time, we calculated Nei's genetic distance (*61*). Let X_u be the frequency of allele u in population X and Y_u be the frequency of allele u in population Y at site l. L is the total number of segregating sites in the sample, and J_X and J_Y are the mean of allele frequencies in populations X and Y. The distance D between each population was calculated as follows:

$$D = \ln\left(\frac{J_{XY}}{\sqrt{J_X J_Y}}\right)$$

where

$$J_X = \sum_u \frac{X_u^2}{L}$$

$$J_Y = \sum_u \frac{Y_u^2}{L}$$

$$J_{XY} = \sum_l \sum_u \frac{X_{u,l} Y_{u,l}}{L}$$

Allele frequencies for individual samples were 0, 0.5, and 1 for homozygous reference, heterozygous, and homozygous alternate genotypes. The resulting distance matrix was used to conduct a multidimensional scaling analysis implemented in R (*62*).

Simulations of CCII

Simulations of loss of genome-wide H_e over time were conducted in R using the script Sim_Script.R. The simulations were initialized with N homozygous recombinants each generated from two randomly drawn parental genomes with one randomly positioned crossover per chromosome. Generations F_{n+1} were generated from random draws of F_n with replacement assuming complete selfing. Expected heterozygosity was calculated for generations F_{18}, F_{28}, and F_{58} as above, using the summed allele counts of the remaining genomes at each time point. Each simulation was initialized with a different random subset of 10,000 sites with no missing data across all the parents (1,166,976 sites).

Realignment of the whole-genome sequences against the H. vulgare ssp. spontaneum reference

The Morex reference contains the null, deletion allele at *Vrn-H2* preventing us from determining segregation of functional alleles in the parents. For this purpose, the parental and pooled-sequencing reads were realigned against the *H. vulgare* ssp. *spontaneum* reference B1K-04-12 (*4*), which contains two of the three characterized ZCCT genes. The sequencing depth was calculated per base pair using samtools v. 1.14 for primary read mappings of a minimum mapping quality of 60. The reported coverage data shows 21-bp sliding medians of these values calculated and plotted in R.

Selection scans

We scanned the genome in 1000 SNP windows for signals of selection between generations F0 and F_{18} using the combined signals of change in allele frequency (Δp) and loss of H_e. We began by simulating neutral evolution of each window using the known starting allele frequencies. The simulated population sizes were held at the very conservative size of N = 100 to account for the genome-wide loss of genetic diversity due to LD. Coverage variability was simulated for each site as normally distributed, with a mean of 25 and variance of 7.2. Evolution was then simulated for 16 generations (equivalent to F_{18}) for each window, like the whole genome simulations described above. At the final time point, the mean Δp and H_e were calculated and recorded. 100,000 simulations were conducted to generate the distribution of both statistics expected under a neutral model.

A one-sided P value was then calculated for each statistic in each window based on the likelihood of the observed value in the simulated data, and Fisher's method was used to generate a combined p-value for both tests. Significant regions were defined as those which showed $P < 0.05$ after multiple testing correction using the Benjamini-Hochberg procedure. Regions were additionally required to have $P < 0.05$ for both H_e and Δp tests without multiple testing correction. Adjacent windows were collapsed into single regions for reporting.

To identify candidate genes that might be targeted by selection within the regions we first extracted all genes overlapping each region and searched for those with a previously characterized role in barley diversification. We identified *Vrs1*, *Ppd-h1*, and *HvCEN* and directly scored the frequency of previously reported causal mutations. These were not necessarily included in our parental variant panel because of stringent filtering or because we excluded deletions from this set. We directly observed the frequency of alleles across generations in our alignment files using the Integrative Genomics Viewer (*63*).

For regions which did not overlap with a previously described barley diversification locus, we searched for genes that have been studied in other plant species. We list these genes in table S2 because of their potential interest to the barley research community, but the evidence that they have been directly targeted by natural selection is weak and they should be treated with caution.

Genotype by sequencing dataset

A random selection of 220 seeds from each of F_{18}, F_{28}, F_{50}, and F_{58} and the parents were surface sterilized and stratified for 1 week on wet paper towels before germination on the bench top. Individual seedlings were transplanted into pots in the greenhouse within a week of germination. Young leaf tissue from each progeny and three replicates of each parent was harvested and transferred into 96 well plates. CTAB genomic DNA extractions were performed for each sample, followed by DNA quantification and quality assessment on agarose gels and using the Quant-It PicoGreen assay (Thermo Fisher Scientific, P11496). Sequencing library construction followed (*64*). Sequencing was performed on an Illumina HiSeq 4000 instrument using single-end PE 150-bp run mode.

Reads for the parents were parsed by barcode, trimmed for adaptor sequences and low-quality ends using the software package trimmomatic v.0.39 (*65*), and aligned to the TRITEX barley reference genome (*51*) using the minimap2 software v.2.1 (*66*). Alignments were sorted using samtools v.1.17 (*53*). Polymorphic sites were identified using bcftools v.1.17 (*56*) using the mpileup and call tools with the flags -c -v. bcftools was then used to filter the raw calls to include only SNPs with

<4 heterozygous calls, a minimum coverage of three reads in each parent, a total depth <100 but less than 1000 across all parents, a quality of more than 500 and no more than 1 alternate allele.

Trimming and alignment was conducted for the progeny as described for the parents. SNP calling at each parental polymorphism in the progeny was conducted using the approach implemented in STITCH (*67*) using a K = 28 and nGen = 60. Imputed genotype data were then merged across chromosomes and filtered with bcftools for sites with missing data <0.05 and <5 heterozygous calls. Two lines with very low coverage were removed from the analysis.

Genetic distance calculation and principal component analysis between progeny and parent datasets were performed using PLINK v.1.90b6.25 (*59*). Genetic distance (D) in PLINK is calculated as the distance in allele counts as follows:

$$D = \frac{\sum_{i-1}^{n} |p_i - q_i|}{n}$$

where p is the allele count for the first individual at SNP i, q is the allele count for the second individual at SNP i, and n is the total number of sites.

Genetic distances of progeny lines were used as input for hierarchical clustering with average linkage and a tree cut with a 0.001 threshold. This approach was used to select a random member of each cluster for GWASs. A consensus haplotype was also generated for ancestry estimation based on the majority rule at each site. The consensus was then used as input for ancestry calculations.

Assignment of primary and secondary ancestry was accomplished using a windowed analysis. For each progeny, the total number of 1-Mb genomic windows falling below a genetic distance threshold of 0.01 from a parent was recorded. The parent with the largest fraction of windows falling below this threshold was assigned as the primary parent. The secondary parent was assigned as the parent with the largest fraction of remaining windows falling under the 0.01 threshold. This approach is approximate because of allele sharing between the parents, segregating variation within the parent accession not captured in our resequencing samples, and several near-identical parents.

Greenhouse flowering-time experiment

In the fall of 2016 and 2017, seeds from the parents and progeny were surface sterilized using 15% bleach, washed with distilled water, and placed on a damp paper towel in a plastic container. Each container was stratified at 4°C for 1 week in the dark, and then seeds were placed at room temperature on the benchtop for 4 days to allow germination. Individual seedlings were transplanted into 1L pots containing UC Soil mix 3 in a greenhouse in Riverside, CA, USA set to maintain temperatures above 20°C and with artificial light set for 14:10 long day conditions in a completely randomized design. In 2016, four replicates of each parent were transplanted, with one individual each of the progeny genotypes. In 2017, a subset of progeny genotypes was selected based on pruning of nearly identical samples and two replicates per line were planted in a completely randomized design. For both experiments, measurements of individuals deemed mix ups in our genetic analysis were excluded. Flowering time (heading date) was estimated based on the emergence of awns (awn tipping) from the first inflorescence. Plant height was measured after plant dry down during harvest. An estimate of the total seed number, $\hat{S}$, was generated with the following formula:

$$\hat{S} = 100 * \left(\frac{w_t}{w_{100}}\right)$$

Where w_t is total seed weight and w_{100} is hundred seed weight.

When available, replicate data were summarized for both parents and progeny as the mean of observed values.

GWAS

The down-sampled progeny dataset, which contained a single member drawn at random from each of the near-clonal genetic lineages, was used to conduct a GWAS on heading date. GWASs were conducted using the software gemma v.0.98.1 (*68*) with cryptic relatedness correction using a kinship matrix calculated from the subsampled genetic dataset.

REFERENCES AND NOTES

1. D. Zohary, M. Hopf, E. Weiss, *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin* (Oxford Univ. Press, 2012). doi: 10.1093/acprof:osobl/9780199549061.001.0001
2. R. W. Allard, A. L. Kahler, "Patterns of molecular variation in plant populations" in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Darwinian, Neo-Darwinian, and Non-Darwinian Evolution*, L.M. Le Cam, J. Neyman, E. L. Scott, Eds., Univ. of California Press, 1972; https://projecteuclid.org/euclid.bsmsp/1200514600.
3. P. L. Morrell, M. T. Clegg, Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3289–3294 (2007). doi: 10.1073/pnas.0611377104; pmid: 17360640
4. A. Pankin, J. Altmüller, C. Becker, M. von Korff, Targeted resequencing reveals genomic signatures of barley domestication. *New Phytol.* **218**, 1247–1259 (2018). doi: 10.1111/nph.15077; pmid: 29528492
5. S. G. Milner *et al*., Genebank genomics highlights the diversity of a global barley collection. *Nat. Genet.* **51**, 319–326 (2019). doi: 10.1038/s41588-018-0266-x; pmid: 30420647
6. A. M. Poets, Z. Fang, M. T. Clegg, P. L. Morrell, Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biol.* **16**, 173 (2015). doi: 10.1186/s13059-015-0712-3; pmid: 26293830
7. L. Lei *et al*., Environmental Association Identifies Candidates for Tolerance to Low Temperature and Drought. *G3* **9**, 3423–3438 (2019). doi: 10.1534/g3.119.400401; pmid: 31439717
8. D. Bustos-Korts *et al*., Exome sequences and multi-environment field trials elucidate the genetic basis of adaptation in barley. *Plant J.* **99**, 1172–1191 (2019). doi: 10.1111/tpj.14414; pmid: 31108005
9. J. Russell *et al*., Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat. Genet.* **48**, 1024–1030 (2016). doi: 10.1038/ng.3612; pmid: 27428750
10. K. E. Brown, D. Koenig, On the hidden temporal dynamics of plant adaptation. *Curr. Opin. Plant Biol.* **70**, 102298 (2022). doi: 10.1016/j.pbi.2022.102298; pmid: 36126489
11. H. V. Harlan, M. L. Martini, A Composite Hybrid Mixture. *Agron. J.* **21**, 487–490 (1929). doi: 10.2134/agronj1929.00021962002100040014x
12. R. W. Allard, The Wilhelmine E. Key 1987 invitational lecture. Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. *J. Hered.* **79**, 225–238 (1988). doi: 10.1093/oxfordjournals.jhered.a110503; pmid: 3166481
13. C. A. Suneson, An evolutionary plant breeding method. *Agron. J.* **48**, 188–191 (1956). doi: 10.2134/agronj1956.00021962004800040012x
14. S. K. Jain, R. W. Allard, Population studies in predominantly self-pollinated species, I. Evidence for heterozygote advantage in a closed population of barley. *Proc. Natl. Acad. Sci. U.S.A.* **46**, 1371–1377 (1960). doi: 10.1073/pnas.46.10.1371; pmid: 16590760
15. B. S. Bal, C. A. Suneson, R. T. Ramage, Genetic shift during 30 generations of natural selection in barley. *Agron. J.* **51**, 555–557 (1959). doi: 10.2134/agronj1959.00021962005100090014x
16. H. V. Harlan, M. L. Martini, H. Stevens, "A study of methods in barley breeding" (Technical Bulletin 720, United States Department of Agriculture, 1940).
17. R. W. Allard, S. K. Jain, Population Studies in Predominantly Self-Pollinated Species. II. Analysis of Quantitative Genetic Changes in a Bulk-Hybrid Population of Barley. *Evolution* **16**, 90–101 (1962). doi: 10.1111/j.1558-5646.1962.tb03201.x
18. G. W. Hendry, M. P. Kelly, The plant content of adobe bricks: With a note on adobe brick making. *Calif. Hist. Soc. Q.* **4**, 361–373 (1925). doi: 10.2307/25177783
19. H. V. Harlan, M. L. Martini, The effect of natural selection in a mixture of barley varieties. *J. Agric. Res.* **57**, 189–199 (1938).
20. G. W. Hendry, The Adobe Brick as a Historical Source; Reporting Further Studies in Adobe Brick Analyses. *Agric. Hist.* **5**, 110 (1931).
21. S. P. Otto, Detecting the form of selection from DNA sequence data. *Trends Genet.* **16**, 526–529 (2000). doi: 10.1016/S0168-9525(00)02141-7; pmid: 11102697
22. T. Komatsuda *et al*., Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1424–1429 (2007). doi: 10.1073/pnas.0608580104; pmid: 17220272
23. H. Guo *et al*., The RNA binding protein OsLa influences grain and anther development in rice. *Plant J.* **110**, 1397–1414 (2022). doi: 10.1111/tpj.15746; pmid: 35322500
24. N. Sentoku, H. Kato, H. Kitano, R. Imai, OsMADS22, an STMADS11-like MADS-box gene of rice, is expressed in non-vegetative tissues and its ectopic expression induces spikelet meristem indeterminacy. *Mol. Genet. Genomics* **273**, 1–9 (2005). doi: 10.1007/s00438-004-1093-6; pmid: 15682279
25. W. Ma *et al*., The LBD12-1 Transcription Factor Suppresses Apical Meristem Size by Repressing Argonaute 10 Expression. *Plant Physiol.* **173**, 801–811 (2017). doi: 10.1104/pp.16.01699; pmid: 27895202
26. J. Yu *et al*., OsLG3 contributing to rice grain length and yield was mined by Ho-LAMap. *BMC Biol.* **15**, 28 (2017). doi: 10.1186/s12915-017-0365-7; pmid: 28385155
27. N. Bollier *et al*., At-MINI ZINC FINGER2 and SI-INHIBITOR OF MERISTEM ACTIVITY, a Conserved Missing Link in the Regulation of Floral Meristem Termination in Arabidopsis and Tomato. *Plant Cell* **30**, 83–100 (2018). doi: 10.1105/tpc.17.00653; pmid: 29298836
28. Z. Lin *et al*., Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* **44**, 720–724 (2012). doi: 10.1038/ng.2281; pmid: 22581231
29. R. C. Elliott *et al*., AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* **8**, 155–168 (1996). doi: 10.1105/tpc.8.2.155; pmid: 8742707

30. K. Schiessl, J. M. Muiño, R. Sablowski, Arabidopsis JAGGED links floral organ patterning to tissue growth by repressing Kip-related cell cycle inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2830–2835 (2014). doi: 10.1073/pnas.1320457111; pmid: 24497510
31. S. Scholz *et al*., The AGC protein kinase UNICORN controls planar growth by attenuating PDK1 in Arabidopsis thaliana. *PLOS Genet.* **15**, e1007927 (2019). doi: 10.1371/journal.pgen.1007927; pmid: 30742613
32. K. Ikeda, M. Ito, N. Nagasawa, J. Kyozuka, Y. Nagato, Rice ABERRANT PANICLE ORGANIZATION 1, encoding an F-box protein, regulates meristem fate. *Plant J.* **51**, 1030–1040 (2007). doi: 10.1111/j.1365-313X.2007.03200.x; pmid: 17666027
33. M. Fernández-Calleja, A. M. Casas, E. Igartua, Major flowering time genes of barley: Allelic diversity, effects, and comparison with wheat. *Theor. Appl. Genet.* **134**, 1867–1897 (2021). doi: 10.1007/s00122-021-03824-z; pmid: 33969431
34. A. Turner, J. Beales, S. Faure, R. P. Dunford, D. A. Laurie, The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* **310**, 1031–1034 (2005). doi: 10.1126/science.1117619; pmid: 16284181
35. J. Comadran *et al*., Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.* **44**, 1388–1392 (2012). doi: 10.1038/ng.2447; pmid: 23160098
36. A. Hemshrot *et al*., Development of a Multiparent Population for Genetic Mapping and Allele Discovery in Six-Row Barley. *Genetics* **213**, 595–613 (2019). doi: 10.1534/genetics.119.302046; pmid: 31358533
37. V. Brambilla *et al*., Antagonistic Transcription Factor Complexes Modulate the Floral Transition in Rice. *Plant Cell* **29**, 2801–2816 (2017). doi: 10.1105/tpc.17.00645; pmid: 29042404
38. X.-L. Zhao, Z.-Y. Shi, L.-T. Peng, G.-Z. Shen, J.-L. Zhang, An atypical HLH protein OsLF in rice regulates flowering time and interacts with OsPIL13 and OsPIL15. *N. Biotechnol.* **28**, 788–797 (2011). doi: 10.1016/j.nbt.2011.04.006; pmid: 21549224
39. M. Jayakodi *et al*., The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* **588**, 284–289 (2020). doi: 10.1038/s41586-020-2947-8; pmid: 33239781
40. R. Wonneberger *et al*., Major chromosome 5H haplotype switch structures the European two-rowed spring barley germplasm of the past 190 years. *Theor. Appl. Genet.* **136**, 174 (2023). doi: 10.1007/s00122-023-04418-7; pmid: 37477711
41. S. J. Franks, S. Sim, A. E. Weis, Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1278–1282 (2007). doi: 10.1073/pnas.0608379104; pmid: 17220273
42. J. Cockram *et al*., Control of flowering time in temperate cereals: Genes, domestication, and sustainable productivity. *J. Exp. Bot.* **58**, 1231–1244 (2007). doi: 10.1093/jxb/erm042; pmid: 17420173
43. S. K. Jain, Stabilizing Selection for Heading Time in a Bulk-Hybrid Population of Barley. *Nature* **191**, 1123–1124 (1961). doi: 10.1038/1911123b0
44. B. Walsh, M. Lynch, *Evolution and Selection of Quantitative Traits* (, 2018).doi: 10.1093/oso/9780198830870.001.0001
45. J. von Zitzewitz *et al*., Molecular and structural characterization of barley vernalization genes. *Plant Mol. Biol.* **59**, 449–467 (2005). doi: 10.1007/s11103-005-0351-2; pmid: 16235110
46. L. Yan *et al*., Positional cloning of the wheat vernalization gene VRN1. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6263–6268 (2003). doi: 10.1073/pnas.0937399100; pmid: 12730378
47. S. M. Rudman *et al*., Direct observation of adaptive tracking on ecological time scales in *Drosophila*. *Science* **375**, eabj7484 (2022). doi: 10.1126/science.abj7484; pmid: 35298245
48. B. Charlesworth, Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009). doi: 10.1038/nrg2526; pmid: 19204717
49. E. A. Hockett, R. F. Eslick, C. O. Qualset, A. L. Dubbs, V. R. Stewart, Effects of Natural Selection in Advanced Generations of Barley Composite Cross II1. *Crop Sci.* **23**, 752–756 (1983). doi: 10.2135/cropsci1983.0011183X002300040036x
50. J. J. Doyle, J. L. Doyle, A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).
51. C. Monat *et al*., TRITEX: Chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol.* **20**, 284 (2019). doi: 10.1186/s13059-019-1899-5; pmid: 31849336
52. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: 10.1093/bioinformatics/btp324; pmid: 19451168
53. H. Li *et al*., The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
54. R. Poplin *et al*., Scaling accurate genetic variant discovery to tens of thousands of samples, bioRxiv 201178 [Preprint] (2018) https://doi.org/10.1101/201178.
55. P. Danecek *et al*., The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011). doi: 10.1093/bioinformatics/btr330; pmid: 21653522
56. P. Danecek *et al*., Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021). doi: 10.1093/gigascience/giab008; pmid: 33590861
57. P. Cingolani *et al*., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012). doi: 10.4161/fly.19695; pmid: 22728672
58. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011). doi: 10.14806/ej.17.1.200
59. C. C. Chang *et al*., Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015). doi: 10.1186/s13742-015-0047-8; pmid: 25722852
60. R Core Team, R: A language and environment for statistical computing. (R Foundation for Statistical Computing, 2013); https://www.R-project.org/.
61. M. Nei, Genetic Distance between Populations. *Am. Nat.* **106**, 283–292 (1972). doi: 10.1086/282771
62. J. C. Gower, Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338 (1966). doi: 10.1093/biomet/53.3-4.325
63. J. T. Robinson *et al*., Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011). doi: 10.1038/nbt.1754; pmid: 21221095
64. B. A. Rowan, D. K. Seymour, E. Chae, D. S. Lundberg, D. Weigel, "Methods for Genotyping-by-Sequencing" in *Genotyping: Methods and Protocols*, S. J. White, S. Cantsilieris, Eds. (Springer, 2017), pp. 221–242.
65. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). doi: 10.1093/bioinformatics/btu170; pmid: 24695404
66. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). doi: 10.1093/bioinformatics/bty191; pmid: 29750242
67. R. W. Davies, J. Flint, S. Myers, R. Mott, Rapid genotype imputation from sequence without reference panels. *Nat. Genet.* **48**, 965–969 (2016). doi: 10.1038/ng.3594; pmid: 27376236
68. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012). doi: 10.1038/ng.2310; pmid: 22706312
69. J. B. Landis *et al*., dpkoenig/Landis2024: v1.0.1, Zenodo (2024); https://doi.org/10.5281/zenodo.11225490.
70. J. B. Landis *et al*., Data From Natural selection drives emergent genetic homogeneity in a century-scale experiment with barley, Dryad (2024); https://doi.org/10.5061/dryad.z34tmpgm8.

ACKNOWLEDGMENTS

We thank the leadership and staff at the University of California, Riverside (UCR) Agricultural Experimental Station, the leadership and staff of the UCR High-Performance Computing Center, and H. Bockelman at the USDA Small Grains and Potato Germplasm Research program for supporting the infrastructure and resources necessary to complete this project. We thank C. Qualset for providing the composite cross germplasm and for comments on the manuscript. We also thank members of the Koenig and Seymour labs for their comments and suggestions during this research. We acknowledge the assistance of F. Reyes Gaibor in the initial curation of prior publications on CCII populations and the current availability of parental seed stock. Lastly, we thank the researchers, including H. Harlan, M. Martini, C. Suneson, and G. Wiebe, who worked to maintain composite cross experiments over the past 100 years. **Funding:** This work was supported by NSF CAREER grant IOS-2046256 (D.K.), NSF grant IOS-13393939 (P.L.M.), and NSF Plant Genome Research Program Postdoctoral Research Fellowships in Biology grants IOS-1711807 (J.B.L.) and IOS-1907061 (K.E.B.). **Author contributions:** Conceptualization: D.K. Methodology: D.K., J.B.L., and P.L.M. Formal analysis: D.K., J.B.L., K.E.B., and C.J.F. Investigation: J.B.L. and A.M.G. Funding acquisition: D.K., J.B.L., and K.E.B. Project administration: D.K. and J.B.L. Supervision: D.K. Writing – original draft: D.K. Writing – review and editing: D.K., J.B.L., K.E.B., C.J.F., A.M.G., and P.L.M. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Genotype by sequencing datasets have been deposited at the National Center for Biotechnology Information (NCBI) under BioProject accession no. PRJNA575963. Whole-genome sequencing datasets have been deposited at the National Center for Biotechnology Information (NCBI) under BioProject accession no. PRJNA1018527. Scripts for simulations and major datasets are found at Zenodo (69) and Data Dryad (70).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adl0038
Supplementary Text
Figs. S1 to S25
Tables S1 to S7

Submitted 22 September 2023; accepted 21 May 2024
10.1126/science.adl0038