TITLE: Local adaptation to climate facilitates a global invasion

AUTHORS: Diana Gamba[1]*, Megan L. Vahsen[2], Toby M. Maxwell[3], Nikki Pirtel[2], Seth Romero[4], Justin J. Van Ee[5], Amanda Penn[1], Aayudh Das[1], Rotem Ben-Zeev[1], Owen Baughman[6], C. Sean Blaney[7], Randy Bodkins[8], Shanta Budha-Magar[9], Stella M. Copeland[10], Shannon L. Davis-Foust[11], Alvin Diamond[12], Ryan C. Donnelly[13], Peter W. Dunwiddie[14], David J. Ensing[15], Thomas A. Everest[16], Holly Hoitink[8], Martin C. Holdrege[17], Ruth A. Hufbauer[5], Sigitas Juzėnas[18], Jesse M. Kalwij[19, 20], Ekaterina Kashirina[21], Sangtae Kim[22], Marcin Klisz[23], Alina Klyueva[24], Michel Langeveld[8], Samuel Lutfy[25], Daniel Martin[8], Christopher L. Merkord[26], John W. Morgan[27], Dávid U. Nagy[28], Jacqueline P. Ott[29], Radoslaw Puchalka[30], Lysandra A. Pyle[31], Leonid Rasran[32], Brian G. Rector[33], Christoph Rosche[34], Marina Sadykova[8], Robert K. Shriver[35], Alexandr Stanislavschi[36], Brian M. Starzomski[37], Rachel L. Stone[38], Kathryn G. Turner[39], Alexandra K. Urza[40], Acer VanWallendael[41], Carl-Adam Wegenschimmel[42], Justin Zweck[43], Cynthia S. Brown[5], Elizabeth A. Leger[44], Dana M. Blumenthal[4], Matthew J. Germino[45], Lauren M. Porensky[4], Mevin B. Hooten[46], Peter B. Adler[2], Jesse R. Lasky[1]

[1] Department of Biology, Pennsylvania State University; University Park, PA, USA.

[2] Department of Wildland Resources and the Ecology Center, Utah State University; Logan, UT, USA.

[3] Department of Biological Sciences, Boise State University; Boise, ID, USA.

[4] US Department of Agriculture, Agricultural Research Service, Rangeland Resources and Systems Research Unit; Fort Collins, CO, USA.

[5] Department of Agricultural Biology, Colorado State University; Fort Collins, CO, USA.

[6] The Nature Conservancy; Burns, OR, USA.

[7] Atlantic Canada Conservation Data Centre; Sackville, NB, Canada.

[8] Not affiliated.

[9] Northtec; Whangarei, New Zealand.

[10] US Department of Agriculture, Agricultural Research Service, Eastern Oregon Agricultural Research Center; Burns, OR, USA.

[11] Biology Department, University of Wisconsin Oshkosh; Oshkosh, WI, USA.

[12] Department of Biological and Environmental Sciences, Troy University; Troy, Alabama, USA.

[13] Division of Biology, Kansas State University; Manhattan, KS, USA.

[14] Department of Biology, University of Washington; Seattle, WA, USA.

[15] Summerland Research and Development Centre, Agriculture and Agri-Food Canada; Summerland, BC, Canada.

[16] Department of Biology, New Mexico State University; Las Cruces, NM, USA.

[17] Northern Arizona University, Center for Adaptable Western Landscapes; Flagstaff, AZ, USA.

[18] Department of Botany and Genetics, Institute of Biosciences, Life Sciences Center, Vilnius University; Vilnius, Lithuania.

[19] Centre for Ecological Genomics & Wildlife Conservation, Department of Zoology, University of Johannesburg; Johannesburg, South Africa.

[20] Institute of Geography and Geoecology, Karlsruhe Institute of Technology; Karlsruhe, Germany.

[21] A.O. Kovalevsky Institute of Biology of the Southern Seas of RAS, Sevastopol, Russian Federation.

[22] Department of Biology, Sungshin Women's University; Seoul, Republic of Korea.

[23] Department of Silviculture and Genetics of Forest Trees, Forest Research Institute; Raszyn, Poland.

[24] Bryansk State University Academician I. G. Petrovsky; Bryansk, Russian Federation.

[25] Caesar Kleberg Wildlife Research Institute, Texas A&M University - Kingsville; Kingsville, TX, USA.

[26] Biosciences Department, Minnesota State University Moorhead; Moorhead, MN, USA.

[27] Department of Environment and Genetics, La Trobe University; Bundoora, Victoria, Australia.

[28] Faculty of Biological Sciences, Goethe University Frankfurt; Frankfurt am Main, Germany.

[29] US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Rapid City, SD, USA.

[30] Department of Ecology and Biogeography, Nicolaus Copernicus University; Torun, Poland.

[31] Alberta Biodiversity Monitoring Institute; Edmonton, AB, Canada.

[32] BOKU University Vienna; Vienna, Austria.

[33] US Department of Agriculture, Agricultural Research Service, Invasive Species and Pollinator Health Research Unit; Albany, CA, USA.

[34] Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg; Halle, Saale, Germany.

[35] Department of Natural Resources and Environmental Science, University of Nevada; Reno, NV, USA.

[36] Department of Organic Biochemical and Food Engineering, Gheorghe Asachi Technical University of Iasi; Iasi, Romania.

[37] School of Environmental Studies, University of Victoria; Victoria, BC, Canada.

[38] Department of Biology, Case Western Reserve University; Cleveland, OH, USA.

[39] Department of Biological Sciences, Idaho State University; Pocatello, ID, USA.

[40] US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Reno, NV, USA.

[41] Department of Horticultural Science, North Carolina State University; Raleigh, NC, USA.

[42] Terrastory Environmental Consulting Inc.; Hamilton, ON, Canada.

[43] Department of Ecosystem Science and Management, Pennsylvania State University; University Park, PA, USA.

[44] Department of Biology, University of Nevada; Reno, NV, USA.

[45] US Geological Survey, Forest and Rangeland Ecosystem Science Center; Boise, Idaho, USA.

[46] Department of Statistics and Data Sciences, The University of Texas at Austin; Austin, TX, USA.

* Email address of corresponding author: dgamba333@gmail.com

ABSTRACT

Local adaptation may facilitate range expansion during invasions, but the mechanisms underlying successful invasions remain unclear. Cheatgrass (*Bromus tectorum*), native to Eurasia and Africa, has invaded globally, with severe impacts in western North America. We aimed to identify mechanisms and consequences of local adaptation in the North American cheatgrass invasion. We sequenced 307 range-wide genotypes and conducted controlled experiments. We found that diverse lineages invaded North America, where long-distance gene flow is common. Nearly half of North American cheatgrass comprises a mosaic of ~19 locally adapted, near-clonal genotypes, each seemingly very successful in a different part of North America. Additionally, ancestry, phenotype, and allele frequency-environment clines in the native range predicted those in the invaded range, indicating pre-adapted genotypes colonized different regions. Common gardens showed directional selection on flowering time that reversed between warm and cold sites, potentially maintaining clines. In the USA Great Basin, genomic predictions of strong local adaptation identified sites where cheatgrass is most dominant. Our results indicate that multiple introductions and migration within the invaded range fueled local adaptation and success of cheatgrass in western North America. Understanding how environment and gene flow shape adaptation and invasion is critical for managing ongoing invasions.

INTRODUCTION

Biological invasions are a major cause of global biodiversity decline and ecosystem disruption, but the mechanisms driving ongoing invasions remain poorly understood[1,2]. In particular, the role of adaptive evolution in enabling invasive species to succeed is poorly understood. Is invasion success primarily determined by the susceptibility of invaded ecosystems[3], or are the worst invaders adapted to spread and dominate[4]? For example, local adaptation to invaded environments could increase fitness and abundance[5], while ongoing gene flow into the invaded range could swamp or reshape local adaptation[6,7]. If colonizing propagules are diverse, new populations may quickly adapt to local environments, facilitating invasive spread[4,8]. However, colonizing genotypes may reach new environments to which they are maladapted, swamping local adaptation and potentially hindering further spread[6,7,9,10]. Furthermore, if the diversity of colonizing propagules is low, new populations may be unable to adapt to new conditions[11], phenotypic plasticity of invasive genotypes may counteract local adaptation[1], and/or colonization bottlenecks may increase the frequency of deleterious mutations[12]. Testing these hypotheses requires a rare combination of genomic, fitness, and abundance data[13].

Multiple mechanisms could contribute to local adaptation during invasions, generating distinct patterns of genomic and phenotypic variation[14,15]. In general, selection may change along environmental gradients and promote genotypic and phenotypic clines[16]. If environmental gradients are similar in native versus invaded regions, clines may be similar between native and invaded regions, indicating niche conservatism of different lineages[17–19]. Alternatively, if selective pressures are novel in the invaded range, clines may be distinct between native and invaded regions, suggesting niche shift of some lineages[19–22]. Furthermore, invasive genotypes may closely match the genetic diversity of the native range[23], represent newly admixed

144 populations[24], or form novel genotypes via introgression from congeners[25]. Understanding
145 successful invasions thus requires dissecting global patterns of genomic and phenotypic
146 variation, which has seldom been accomplished. Although some studies have examined genomic
147 and phenotypic differences between native and invasive populations, sampled populations
148 between ranges are hard to compare because they often differ in spatial scale and/or do not
149 incorporate enough environmental variation[26].
150
151 *Bromus tectorum* L. (cheatgrass) is a grass native to Eurasia and northern Africa that spread
152 across North America by the 1890s[27,28], heavily influencing ecological dynamics of arid and
153 semi-arid ecosystems of the North American Intermountain West[29]. At least some introductions
154 likely came via contamination in grain shipments[27,28]. Cheatgrass occurs in high abundance
155 across an estimated 31% (210,000 km^2) of this region[30], displacing native perennials via rapid
156 reproduction and shortened fire return intervals[29], reducing biodiversity and degrading wildlife
157 habitat[31]. It is highly selfing, typically winter annual, with a high-quality reference genome (~2.5
158 Gb)[28,32]. Existing genetic studies, while limited to small numbers of markers or populations,
159 suggest that multiple introductions from different regions in Europe might have occurred in
160 North America[28,33–38]. Studies have shown evidence for local adaptation in phenology at small
161 scales[39–43], and substantial genetic differentiation in phenology between populations from
162 different regions[32,44], although range-wide patterns of local adaptation remain elusive. Due to
163 cheatgrass's high rate of selfing, novel recombinant genotypes are expected to be rare, limiting
164 novel genomic diversity in the invaded range. However, repeated introductions into North
165 America and post-introduction dispersal could have promoted the adaptative potential and
166 invasive spread of cheatgrass populations.
167

168 Here, we aim to identify mechanisms and consequences of local adaptation in the North
169 American cheatgrass invasion. We hypothesize that for cheatgrass, as a self-pollinating species
170 with few constraints on dispersal, local adaptation in the invaded range would be likely due to
171 rapid spread of pre-adapted genotypes to suitable environments (as opposed to adaptation by *de*
172 *novo* mutation or novel admixtures), provided there were multiple diverse introductions. We
173 sequence whole genomes of a global panel of 307 genotypes from the native and invaded ranges
174 and measure phenotypes and performance in one growth chamber experiment and two field
175 common gardens. We ask whether there were multiple and diverse introductions to North
176 America and examine genetic consequences of the invasion. We evaluate how geography,
177 environment, and phenotype shape genomic diversity. We test whether ancestry, trait, and allele
178 frequency-environment clines were repeated in native and invasive genotypes, and if selection
179 maintains clines. Finally, we integrate field surveys of cheatgrass abundance in the USA Great
180 Basin[30] to assess whether genomic matching to local climates facilitated invasive dominance.
181 Our results reveal that multiple introductions and migration within the invaded range fueled local
182 adaptation and success of cheatgrass in western North America.
183

184 **RESULTS AND DISCUSSION**

185 **Diverse native range ancestries invaded North America**

186 Cheatgrass populations in North America stem from multiple, diverse introductions. Using
187 ~267k unlinked single-nucleotide polymorphisms (SNPs), different clustering analyses of global
188 genomic variation showed that population genetic structure largely followed geography in the
189 native range and to a lesser degree in North America (Fig. 1 showing K=4 ancestral genetic
190 clusters/ancestries, Supplementary Fig. 1). In the native range, west Asian, Mediterranean, and
191 Atlantic genotypes primarily fell in a single ancestry, while central and eastern European

genotypes were mostly assigned to two ancestries differentiated by latitude and were overall more intermediate (*i.e.*, composed of multiple ancestries). In the invaded range, genotypes were assigned to all four ancestries in western North America (WNA, west of the Rocky Mountains), but only to two ancestries in eastern North America (ENA, east of the Rocky Mountains) (Fig. 1a–c, Supplementary Fig. 2). The majority of invasive genotypes were similar to genotypes from north, central, or eastern Europe (Fig. 1d–f). In WNA, however, warm desert genotypes in southern California and Nevada were similar to genotypes from Iran and Afghanistan. The warm Mojave and the cool Pacific Northwest also harbored genotypes similar to those from the western Mediterranean (Fig. 1e,f). For regions with less extensive invasions, results showed that: Argentines are similar to Spanish genotypes, an Australian is similar to western Mediterranean genotypes and a widespread lineage from WNA, New Zealand genotypes are similar to northeastern European genotypes, and a Korean genotype is similar to central eastern European genotypes and a widespread lineage from ENA.

The diversity of genotypes found in WNA reflects colonization by propagules from different native regions, while patterns in ENA reflect reduced genetic diversity. Accordingly, population-specific F_{ST} (*i.e.*, the degree of relatedness among individuals) was higher in the invaded compared to the native range (0.2 and 0.03, respectively) especially for ENA (0.39) compared to WNA (0.18). Pairwise F_{ST} values were lowest between European and North American genotypes, while other pairs of regions were more diverged (e.g., those involving the Mediterranean and west Asia, Fig. 1d). The native and invaded range were moderately genetically differentiated ($F_{ST} = 0.11$), comparable to the differentiation between genotypes from ENA and WNA ($F_{ST} = 0.12$). In the native range, pairwise F_{ST} values were larger (Fig. 1d), showing strong divergence between European and west Asian genotypes ($F_{ST} = 0.25$).

216

217 **WNA and ENA: different patterns of diversity**

218 Much of North America harbors great genomic diversity with little evidence of elevated genetic
219 load and inbreeding compared to the native range (Supplementary Figs. 3–5 using ~15.1M
220 SNPs). In WNA, nucleotide diversity (π, Supplementary Fig. 3a,b) was comparable to the most
221 diverse native region, north-central-eastern Europe ($0.0016 \pm 4.5\times10^{-6}$ se vs. $0.0018 \pm 4.6\times10^{-6}$
222 se, respectively), followed by the Mediterranean ($0.0015 \pm 3.8\times10^{-6}$ se) and west Asia ($0.0011 \pm$
223 3.1×10^{-6} se). Nucleotide diversity was much lower in ENA ($0.0009 \pm 4.5\times10^{-6}$ se). In WNA, the
224 skew in the site frequency spectrum (Tajima's D, Supplementary Fig. 3c,d) was positively
225 shifted (mean=2.8 ± 0.006 se), indicating an excess of intermediate-frequency SNPs, consistent
226 with strong population structure and heterogeneous ancestry across the region (see also[38]). In
227 ENA, Tajima's D was low (mean=0.5 ± 0.009 se), indicating more rare variants and suggesting
228 recent population expansion. In the Mediterranean and north-central-eastern Europe, Tajima's D
229 was positively shifted (Mediterranean mean=1.6 ± 0.005, north-central-eastern Europe mean=2.2
230 ± 0.007), reflecting substantial population structure within these regions. The Mediterranean
231 comprises multiple distinct eastern and western lineages (see also[45]), while north-central-eastern
232 Europe comprises multiple lineages with some intermediate genotypes. In contrast, west Asian
233 genotypes appeared more closely related to each other (Tajima's D mean=0.3 ± 0.006).

234

235 To understand the effects of potential bottlenecks and drift in North America, we first examined
236 deleterious mutation load using ~15.1M SNPs, under the hypothesis that most protein changing
237 mutations are deleterious. Estimated mutation load was not different between native versus
238 invaded range genotypes of the same ancestry (two-way ANOVA: range $F_{(1,290)}=57.8$, $p=4\times10^{-}$
239 ^{13}, ancestry $F_{(4,290)}=46.7$, $p<2\times10^{-16}$, interaction $F_{(3,290)}=4.4$, $p=0.005$; Tukey HSD range $p=0.2$,

240 Supplementary Fig. 4). The central-eastern European ancestry (teal in Supplementary Fig. 4),
241 widespread in North America, showed the lowest load in both ranges, suggesting large effective
242 population size at some point in the past. In contrast, the west Asian and Mediterranean ancestry
243 (pink in Supplementary Fig. 4) was associated with higher load in both ranges.
244
245 Next we examined runs of homozygosity (ROH[46]) using a panel of 101 closely related native and
246 invasive genotypes sequenced directly from field collections (Supplementary Fig. 5). Native and
247 invasive genotypes (grouped by range) had similar Tajima's D, thus similar skew in the site
248 frequency spectrum. The native group, however, had lower counts of ROH and a much higher
249 FROH (the proportion of the genome with ROH), resulting in inference of a strong selfing rate
250 (Supplementary Fig. 5a–c). Selfing rates were significantly different between the native and
251 invasive groups (two-tailed t-test t=3.8437, df=11.896, p=0.002), though both were >0.9
252 (Supplementary Fig. 5c). This could reflect relaxed selection for reproductive assurance in
253 invasive genotypes from specific environments. For example, although selfing is more common,
254 some highly inbred desert lineages appeared overrepresented among parents of heterozygotes in
255 a previous common garden experiment[47]. Our results suggest that the North American cheatgrass
256 invasion is not associated with higher inbreeding due to selfing compared to the native range
257 (Supplementary Fig. 5c,d).
258
259 Taken together, the high diversity in WNA indicates great potential for adaptation in this heavily
260 invaded region. In contrast, the lower diversity in ENA reflects colonization by a few closely
261 related lineages (see also[34]) that persist as ruderal plants in urban and agricultural environments.
262
263 **Strong isolation-by-environment in North America**

264 Both geography and environment shape genomic diversity in the native range, but geography
265 plays a weak role in North America. Isolation-by-distance (based on ~267k SNPs) was strong in
266 the native range (geographic vs. genetic distance Mantel $p=10^{-4}$, Fig. 2a) but very weak in North
267 America (Mantel $p=0.06$, Fig. 2b). At 0–100 km distance, pairs of distantly related genotypes
268 were common in WNA, but not in ENA or the native range (Supplementary Fig. 6a,b). Even at
269 the smallest scales (0–25 km), isolation-by-distance appeared weaker in the invaded compared to
270 the native range (Supplementary Fig. 6c,d). Moreover, several groups in North America ("1–19"
271 in Fig. 1f) composed of 2–14 near-clonal genotypes (>98% SNPs identity) were found across
272 distances of >3000 km (Fig. 2b, Supplementary Fig. 6a, 7a). In contrast, such widely distributed,
273 near-clonal genotypes were absent in the native range (Fig. 2a, Supplementary Fig. 6b). These
274 patterns suggest long-distance dispersal within North America by lineages descended from
275 distinct native range populations. Furthermore, groups of near-clonal genotypes occupied
276 significantly different environments (PERMANOVA of multivariate environment predicted by
277 clonal group: $p=0.0001$, $R^2=0.65$) that together encompass the extent of climate space in North
278 America (Supplementary Fig. 7b). This suggests that although genotypes might not be dispersal
279 limited in North America, their spread may be limited by different environmental constraints,
280 suggesting local adaptation. The weak spatial patterns in North America may also reflect
281 genotype sorting along the steep, heterogeneous climatic gradients that are common in WNA.
282 Pairwise climatic distance (based on Euclidian distance in Supplementary Fig. 6e) significantly
283 increased with spatial distance in both native and invaded ranges (Mantel $p=10^{-4}$ and Mantel
284 Pearson correlation=0.6 in both ranges; Supplementary Fig. 6f,g), but this relationship was weak
285 in WNA (Mantel Pearson correlation=0.3 WNA vs. 0.7 ENA), reflecting the fine-scale climatic
286 heterogeneity of this region.
287

288 To further examine genomic differentiation along climate gradients, we performed redundancy
289 analysis (RDA) with variance partitioning, comparing the role of climate and spatial variables in
290 explaining genomic variation. SNP variation was better explained by these predictors in the
291 native than in the invaded range (native R^2_{adj}=0.25, invaded R^2_{adj}=0.10; Fig. 2c,d). Spatial
292 variables explained little SNP variation in North America (native R^2_{adj}=0.07, invaded
293 R^2_{adj}=0.005), confirming low isolation-by-distance. In both ranges the abiotic environment
294 explained the largest portion of SNP variation (native R^2_{adj}=0.13, invaded R^2_{adj}=0.09; Fig. 2e,f),
295 highlighting the importance of isolation-by-environment in both the native and invaded range
296 and consistent with invasive local adaptation via native pre-adaptation[19].
297
298 **Repeated ancestry-climate clines**
299 Ancestry-environment clines were remarkably similar in the native and invaded ranges,
300 suggesting environmental filtering of pre-adapted genotypes that could disperse long distances or
301 via directed gene flow (as opposed to local adaptation by novel genotypes). We focused on
302 aridity and temperature gradients representative of global climatic variation in the cheatgrass
303 range (see Supplementary Fig. 6e) and used generalized-additive-models (GAMs) to detect
304 significant ancestry-climate trends between ranges (Supplementary Fig. 8). In native and
305 invasive genotypes, the west Asian and Mediterranean genetic cluster (pink) was more frequent
306 in drier regions (GAM p=0.0004, pseudo-R^2=0.5), the northern Europe cluster (blue) was more
307 frequent in humid regions (GAM p=0.007, pseudo-R^2=0.08), the central Europe cluster (teal)
308 was more frequent in regions with little precipitation seasonality (GAM $p=10^{-5}$, pseudo-R^2=0.2),
309 and the presumably northeast Europe ancestry (green) was more frequent in regions with colder
310 winters (GAM p=0.002, pseudo-R^2=0.1).
311

312 **Repeated phenotype-climate clines**

313 Consistent with the hypothesis that pre-adaptation to local climate facilitated the cheatgrass
314 invasion, we found similar phenotype-environment clines in the invaded and native ranges. A
315 principal components (PC) analysis on genetic variation among 169 native and invasive
316 genotypes for eleven growth chamber phenotypes (Supplementary Data 1, Supplementary Table
317 1) detected multi-trait axes of variation (Fig. 3a). PC1 explained 35.4% of the variation and
318 suggested a life history axis of delayed flowering and high vegetative investment (more tillers
319 and leaves) versus rapid flowering and high reproductive investment (taller, more fecund
320 inflorescences). PC2 explained 22.2% of the variation and indicated an axis associated with
321 larger plants with greater growth after vernalization versus shorter plants with little growth after
322 vernalization. Native genotypes had on average earlier flowering (two-tailed t-test t=4.09,
323 df=41.36, p=0.0002) and higher reproductive investment (two-tailed t-test t= –1.92, df=43.69,
324 p=0.06) than invasive genotypes (PC1 two-tailed t-test t=3.02, df=38.14, p=0.004) which may be
325 due to different ancestry proportions in the native range. We found no significant native versus
326 invasive trait differences after accounting for relatedness (p>0.4), thus no evidence for evolution
327 of increased competitive ability by invasive cheatgrass[48]. Additionally, near-clonal groups were
328 significantly different in multivariate phenotypes (PERMANOVA of multivariate phenotype
329 predicted by clonal group: p=0.03, R^2=0.39), yet the remaining non-clonal genotypes were also
330 diverse (Supplementary Fig. 7c). These results highlight how North America hosts diverse life
331 histories.

332

333 Multiple trait-climate clines potentially maintained by selection were mirrored between the
334 native and invaded range (Fig. 3b), indicating sorting of genotypes along humidity and
335 temperature gradients. We focused on two climate variables that we hypothesized would capture

336 distinct climatic stressors: maximum vapor pressure deficit (Pa), for drought adaptation, and
337 mean winter temperature (ºC), for cold adaptation (Supplementary Fig. 6e). To test for evidence
338 of selection maintaining clines, we used linear-mixed models that accounted for genomic
339 similarity (lmkin below), similar to Q_{ST}–F_{ST} tests[49]. When significant, these models suggest
340 selection is driving trait-climate clines, because the cline is stronger than expected by the
341 genome-wide patterns of variation. In native and invasive genotypes, earlier flowering was
342 associated with higher aridity (native lmkin-p=2e^{-8}, R^2=0.5; invaded lmkin-p=0.03, R^2=0.3),
343 suggesting a locally adaptive cline of rapid phenology/early reproductive investment in arid
344 regions versus delayed phenology/early vegetative investment in humid regions. Also, clines
345 showed evidence of selection specifically within WNA, but not in ENA (Supplementary Fig. 9).
346 These patterns suggest local adaptation via life history clines in WNA. In contrast, life history in
347 ENA is not associated with climate, potentially because a single generalist ruderal strategy is
348 adaptive throughout ENA.
349
350 **Selection on flowering time along a temperature gradient in WNA**
351 In field common gardens, selection on flowering time changed direction between sites that
352 differed in temperature. To test whether phenotypic clines in WNA were promoted by selection,
353 we conducted two common garden experiments in different climates in Idaho with fall plantings
354 across two years (2021 and 2022). One site was cooler (Sheep Station, ID, USA 44.2456ºN,
355 112.2144ºW, annual mean temperature 6ºC) and the other warmer (Wildcat, ID, USA
356 43.4744ºN, 116.9018ºW, annual mean temperature 12ºC). We planted 95 diverse genotypes from
357 across WNA, for a total of 14,800 plants[50]. We measured flowering time, survival, and fecundity.
358 In both years (Fig. 3c), selection favored later flowering at the cool site, with late flowering
359 genotypes having ~3× the fitness of earlier flowering genotypes (~300 vs. ~100 seeds produced

360 per original sown seed). By contrast, at the warm site, selection favored earlier flowering, with
361 the earliest flowering genotypes having ~10× the fitness of the later flowering genotypes (~170
362 vs. ~17 seeds produced per original sown seed). This suggests that late flowering genotypes have
363 an extreme disadvantage in warm climates. This strong selection is consistent with our finding
364 that the hottest sites in WNA were almost exclusively comprised of west Asian-like genotypes.
365 We saw no clear admixture from distantly related, but geographically proximate European-like
366 genotypes inhabiting cooler and wetter higher elevations in WNA (Fig. 1a), suggesting a barrier
367 to dispersal of maladapted genotypes. Thus, climate gradients in WNA appear to impose changes
368 in selection maintaining a strong phenotypic cline.
369
370 **Repeated allele frequency-climate clines**
371 Putative quantitative trait loci (QTL) for traits under selection showed similar allele frequency-
372 climate clines in the invaded and native ranges (Fig. 4, Supplementary Figs. 10–11). Using
373 genome-wide association studies (GWAS), we identified several genetic loci associated with
374 variation in flowering time and number of tillers (Supplementary Data 2). We implemented two
375 methods that accounted for population genetic structure: univariate mixed model LMM and
376 multilocus mixed model MLMM. Of the 400 top GWAS SNPs (100 top SNPs × 2 phenotypes ×
377 2 GWAS methods), just one SNP was segregating exclusively within the invaded range. The
378 other 399 SNPs segregated in both native and invaded ranges, supporting the hypothesis that *de*
379 *novo* mutations have not been major drivers of local adaptation in North America.
380
381 We annotated SNPs based on genome-wide linkage-disequilibrium (LD, Supplementary Fig.
382 12a). We found that at 194.5 kb, LD decayed to ~80% of the background LD (here taken as 5
383 Mb). A similar LD decay pattern was observed across chromosomes (Supplementary Fig. 12b),

384 while between chromosomes average R^2 ~0.1. Below, we thus highlight QTL based on the
385 position of the closest gene within a 200 kb window centered at the GWAS SNP. We focus on
386 flowering time because this was the only trait for which we found gene functions clearly related
387 to phenotype.
388
389 The top flowering time QTL (detected with LMM) contained multiple SNPs along a haploblock
390 of ~28 Mb (chromosome 1: 56–84 Mb, allele frequency (AF)~0.9) containing ~64 genes with
391 annotations based on homology to *Oryza sativa* and *Arabidopsis thaliana*. These genes were
392 enriched for gene ontology terms describing developmental processes involving reproductive
393 structure/system, embryo, embryo ending in seed dormancy, post-embryonic, fruit, and seed (8
394 *O. sativa* genes and 14 *A. thaliana* genes, $p<0.0003$, FDR=0.01). Such a large haploblock could
395 indicate a structural variant, a potential driver of local adaptation[51,52]. This locus thus merits
396 further investigation.
397
398 The top SNP of the haploblock (chromosome 1: 71007448 bp, AF=0.91; top in LMM and 2nd top
399 in MLMM) was 25 kb downstream of a *O. sativa* homolog, the DnaJ protein Erdj3b. Expression
400 of Erdj3b in *O. sativa* is critical for heat stress tolerance during seed development[53]. Late
401 flowering alleles were more frequent in humid/colder regions of the native (two-tailed t-test t= –
402 3.66, df=83.61, p=0.0004) and invaded range (two-tailed t-test t= –4.31, df=26.41, p=0.0002)
403 (Fig. 4a–c), suggesting cheatgrass adaptation to temperature gradients may be linked to seed
404 sensitivity to temperature stress.
405
406 The fourth top flowering time QTL (only found with LMM) comprised three SNPs (chromosome
407 1: 236616590 bp, 236616999 bp, 236617691 bp, AF=0.82) 0.5 kb upstream (putative promoter

408 region) of the *A. thaliana* homolog *ATE1* (AT5G05700). *ATE1* regulates seed maturation,
409 seedling metabolism, and abscisic acid germination sensitivity[54]. Early flowering alleles were
410 found in drier regions of the native (two-tailed t-test t=6.73, df=42.19, p=3.4e^{-8}) and invaded
411 range (two-tailed t-test t= 4.82, df=11.67, p=0.0004), specifically the Mediterranean and west
412 Asia in the native range, and the Mojave and Lahontan Basin in the invaded range, but also
413 reaching Mediterranean climates of coastal WNA (Fig. 4d–f). These patterns suggest that even
414 the specific QTL underlying local adaptation in the native range have been similarly reused for
415 local adaptation in the invaded range.
416
417 We compared our GWAS results with a published study that performed a GWAS for flowering
418 time using a smaller and much less diverse genotype panel[32]. There was no overlap among the
419 200 top GWAS SNPs we found (100 top SNPs × 2 GWAS methods) and the SNPs detected in
420 that study (Table 2 in[32]), likely due to our larger and more diverse panel of genotypes.
421
422 **Cheatgrass dominates where local adaptation is predicted to be stronger**
423 Whole genome-environment associations in the native range predicted local adaptation in the
424 invaded range, especially where cheatgrass is most dominant. To further evaluate whether
425 invasive genotypes matched local climates as in the native range, we used a predictive genome-
426 environment model. Using the native range RDA model of genotype as a function of climate
427 (Fig. 2e), we first predicted invasive genotypes for locations of our sequenced samples. Next, we
428 calculated the genetic distance between predicted and observed genotypes, similar to metrics
429 sometimes referred to as 'genomic offset'[55]. Genotype-environment matching (*i.e.*, low genetic
430 distance, or offset, between predicted and observed genotypes) was strongest at northern
431 latitudes across North America, particularly in WNA. Putative maladaptation (*i.e.*, high genetic

distance between predicted and observed genotypes) was strongest in the southeast USA (Fig. 5a). By comparing mean genetic distance to means of 1000 null permutations, we found the mean genetic distance was significantly lower than the null expectation in WNA ($p<0.002$), but not in ENA ($p=0.5$, Fig. 5b). This finding is consistent with the hypothesis that local adaptation to climate in WNA reflects the patterns observed in the native range, while cheatgrass in ENA has a novel strategy or is not locally adapted. Unlike WNA, cheatgrass populations in ENA are more restricted to highly disturbed urban and agricultural sites, rarely forming large monospecific stands[56].

To assess whether matching of specific genotypes to local environments promotes cheatgrass invasion, we compared the strength of genotype-environment correlations with variation in cheatgrass abundance from 11,307 field surveys across the Great Basin (Fig. 5a), the region where the invasion has its worst impacts[30]. Locations where cheatgrass occurs in high abundance showed significantly high genotype-environment matching based on the native range model compared to sites where cheatgrass does not dominate (n=55, two-tailed t-test t=2.89, df=52.3, $p=0.006$, Fig. 5c), suggesting local adaptation promotes cheatgrass dominance. This pattern was consistent when comparing genotype-environment matching of high-abundance sites to 1000 null permutations of genotypes within the Great Basin ($p=0.01$), evidence that this pattern was not merely due to environmental characteristics of the low-abundance sites but reflects the match of genotypes to their local environments.

Synthesis

Biological invasions pose a major environmental threat, but the roles of genomic diversity, repeated introductions, and adaptation are poorly understood. Our results have important

456 implications for understanding the evolution of local adaptation in invasive species that have
457 greatly expanded their range. Past studies suggest that selfing species with large native ranges –
458 like cheatgrass– are more likely to establish self-sustaining populations in new regions[23,57,58], but
459 the mechanisms promoting successful establishment have been less explored.
460
461 We show that multiple diverse introductions and long-distance dispersal post-introduction likely
462 increased the chances of cheatgrass genotypes arriving to favorable North American
463 environments. Across Eurasia, cheatgrass shows clines indicative of local adaptation and
464 continuous isolation by distance, likely shaped by multiple migrations and subsequent isolation
465 after the Last Glacial Maximum[45]. In the invaded range, many genotypes represent a mosaic of
466 near-clonal lineages across sometimes widely spread locations but similar environments, and
467 non-clonal genotypes sorted along the steep climate gradients of western North America. Despite
468 differences in genomic diversity and likely demographic history between the native and invaded
469 ranges, western North American genotypes closely matched the native signatures of local
470 adaptation along temperature/aridity gradients. Thus, in North America environmental filtering
471 of pre-adapted genotypes likely led to reuse of native diversity and facilitated range
472 expansion[18,19,59]. Accordingly, common gardens revealed that changing selection likely
473 maintains a major life history cline in the USA Intermountain West. Furthermore, genomic
474 signatures of local adaptation also predicted cheatgrass ecological dominance across the USA
475 Great Basin, indicating that factors supporting local adaptation, such as high genetic diversity
476 from repeated introductions, fuel the invasion.
477
478 Our findings emphasize that any sources of genetic diversity could continue to reshape
479 adaptation in established invasive species[60]. With high genomic diversity and no dispersal

480 limitation, range-wide adaptation could persist over time even under shifting environments.
481 Limiting ongoing introductions and intra-continental dispersal of genotypes (*e.g.*, by limiting
482 seed contaminants in grains) could likely help minimize rapid adaptation of invasive plants. For
483 annual selfers like cheatgrass, this strategy might limit local adaptation via pre-adaptation, but
484 also via *de novo* variation from uncommon but potentially important outcrossing events[28,39,47].
485

486 **METHODS**

487 **Plant material.** Natural inbred lines of *Bromus tectorum* were obtained from 1) the Genome
488 Resources Information Network (GRIN), 2) Greenhouse inbred/selfed plants (S1 or S2) of field
489 samples in western North America collected 2019–2020, 3) field samples from the native and
490 invaded range collected 2020–2022, and 4) DNA extractions of frozen seedlings (contributed by
491 Brian Rector, USDA–ARS) for 29, 111, 155, and 12 samples, respectively (Supplementary Data
492 1). With this panel of genotypes, we targeted sites with distinct environmental conditions,
493 favoring environmental variation over intra-population sampling. Sites were ~1–6600 km apart
494 within the native or invaded ranges: 194 North American, 105 Eurasian, and 8 from regions with
495 less extensive invasions: 2 from Argentina, 1 from Australia, 3 from New Zealand, and 2 from
496 South Korea. Genotypes with available seeds (295) were germinated in a growth chamber at
497 20ºC (80% humidity, 12 h light/12 h dark, 200 μmol m^2 s^{-1} light intensity) to increase seed
498 (Supplementary Methods 1), verify identification, and obtain tissue for whole genome
499 sequencing (Supplementary Methods 2–4) and genotyping (Supplementary Methods 5–6).
500 Native genotypes were assigned to central-north-east Europe, Mediterranean, or west Asia based
501 on geographic location (Supplementary Data 1). North American genotypes were assigned to
502 eastern North America (ENA) or western North America (WNA) based on ecological region at
503 location of origin (Supplementary Methods 7). WNA: marine west coast forest, Mediterranean

504 California, North American deserts, northwestern forested mountains, and temperate Sierras.
505 ENA: eastern temperate forests, Great Plains, and northern forests (Supplementary Fig. 2).
506
507 **Population genetic structure.** With a dataset of 266,504 unlinked sites, we used multiple
508 methods to infer population genetic structure that we interpret collectively (Supplementary
509 Methods 8–9). We estimated individual admixture proportions in NGSadmix (v.33)[61] and
510 inferred population genetic structure with PCA in PCAngsd (v.1.11)[62], which work directly with
511 genotype likelihoods that contain all relevant information of unobserved genotypes. Individual
512 admixture proportions were estimated with maximum likelihood in 12 replicates, for K=2–12
513 genetic clusters, on sites with minor allele frequency (MAF) >0.05, at least in 50% individuals,
514 and <75% missing data. Cross-validation of number of clusters was determined from log-
515 likelihoods of the NGSadmix output across all replicates[63].
516
517 We also computed an unrooted phylogenetic tree with the Neighbor-Joining (NJ) algorithm[64] and
518 calculated Nei's[65] pairwise F_{ST} and Weir & Goudet's population-specific F_{ST}[66] based on genetic
519 dissimilarity estimated with SNPRelate (v.0.9.19)[67]. Pairwise F_{ST} measures population genetic
520 differentiation between regions, whereas population-specific F_{ST} measures regional deviations
521 from the ancestral population. High values of population specific-F_{ST} indicate high within-group
522 allele sharing and potentially greater divergence from ancestral populations, while low values
523 indicate possible ancestral populations.
524
525 **Genomic diversity and Tajima's D.** We compared nucleotide diversity (π[68]) between ranges
526 and regions in the native and invaded range using a dataset of 15,101,725 sites (Supplementary
527 Methods 6). Regional .vcf datasets containing all sites were generated with BCFtools (v.1.18)[69]

528 (view -S), and genome-wide estimates of π were obtained with VCFtools (v.0.1.15)[70] using 50
529 kb sampling windows. Deviations from neutral evolution between geographic regions were
530 examined with the Tajima's D statistic[71], which compares the mean number of pairwise
531 differences against the number of segregating sites observed in a set of sequences. Tajima's D
532 for each region was calculated in 50 kb sampling windows for shared SNPs in VCFtools using --
533 TajimaD.
534
535 **Genetic load.** Genotype mutation load (under the hypothesis that most protein changing
536 mutations are deleterious) was estimated separately from the high impact and missense variants,
537 both normalized by (divided by) the number of synonymous variants (Supplementary Methods
538 10). We used 2-way ANOVA and Tukey HSD tests to examine differences in genetic loads
539 between native and invaded range genotypes of the same ancestry. Genotypes were assigned to a
540 cluster based on having >0.55 ancestry proportion for the NGSadmix K=4 ancestral genetic
541 clusters. If no K=4 ancestry was >0.55, genotypes were designated as intermediate.
542
543 **Self-fertilization rates.** We examined the causes of inbreeding between native and invaded
544 genotypes with several statistics. We used a subset of 101 closely related native and invasive
545 genotypes that were sequenced from seedlings of plants collected directly from the field (as
546 opposed to a greenhouse bulking or GRIN). Based on the ~15M SNPs dataset, we first computed
547 runs of homozygosity[46,72] (with BCFtools roh), Tajima's D (with VCFtools --TajimaD in 100 kb
548 windows), and heterozygosity (with PLINK (v.1.9)[73] --het) for one lineage represented in the
549 native and North American invaded range (native n=27, invaded n=74). Then we implemented
550 random forests in R to analyze these genetic statistics together and estimate selfing rates for each

551 group using a recently published model ("sequential model"[46]). We compared group means with
552 a two-tailed t-test.
553
554 **Isolation-by-distance.** We examined isolation-by-distance in native and invasive genotypes, and
555 in genotypes from WNA and ENA, with the same LD-filtered SNP dataset. Genome-wide
556 pairwise genetic dissimilarity matrices were obtained as described above. Pairwise geographic
557 distances were calculated in kilometers from genotype coordinates with the spDists function in
558 the R package sp (v.2.1-3)[74], using the WGS84 ellipsoid projection. Simple Mantel tests[75] were
559 used to test if the natural logarithm of geographic distance predicts genetic distance with the
560 function mantel.rtest in the R package ade4 (v.1.7-22)[76], using 999 permutations. We conducted
561 linear regressions (with the R function lm) to assess the proportion of genetic variance explained
562 by geographic distance. We also used the mantel.rtest R function to assess how climatic distance
563 changed with geographic distance.
564
565 **Environmental differentiation between groups and across space.** We examined if the 19
566 genetically different groups of 2–14 nearly clonal genotypes (detected based on >99% SNP
567 similarity) were differentiated by environment using climate data at their location of origin. We
568 used clonal group identity as the predictor of 52 CHELSA climate variables[77,78] (Supplementary
569 Methods 7, Supplementary Data 1, Supplementary Fig. 7b) with PERMANOVA using function
570 adonis2 (9999 permutations, Euclidian distances) in the R package vegan (v.2.6-4)[79]. To examine
571 spatial environmental heterogeneity, climatic distance was estimated for each region (native,
572 ENA, WNA) based on pairwise Euclidean distances in the environmental PCA produced above,
573 using the function vegdist in vegan.
574

575 **Variance partitioning of genomic diversity.** Redundancy analyses (RDA) were used to model
576 how sets of variables explained SNP variation and for identifying abiotic gradients explaining
577 the most genome-wide SNP variation[80]. To model geographic patterns in the RDA, a distance
578 matrix obtained from coordinates was converted into a spatial weighting matrix to get a reduced-
579 dimension set of orthogonal variables (Moran's eigenvector maps, MEMs[81]). MEMs are
580 eigenvectors of the pairwise spatial weighting matrix among samples (Supplementary Methods
581 11). Then, RDA was conducted with variance partitioning[80] to quantify proportion of genome-
582 wide SNP variation explained by each of two categories of covariates: Abiotic variables and
583 geographic MEMs. We selected abiotic variables that were informative and non-colinear based
584 on the PCA explaining range-wide environmental variation (Supplementary Fig. 6e). Variance
585 partitioning estimates proportion of SNP variation that is explained by the collection of variables
586 in each category and by collinearity among variables. To identify environmental gradients
587 associated with genome-wide divergence, RDA was also conducted using only abiotic variables
588 for native and invasive genotypes. We computed RDA and performed variance partitioning with
589 functions rda and varpart, respectively, in vegan.
590
591 **Ancestry-environment associations.** To assess environmental filtering of pre-adapted
592 genotypes in North America, we examined ancestry-climate associations in invasive versus
593 native genotypes using generalized additive models (GAMs). GAMs allow us to account for
594 nonlinear patterns between predictors and the response variable[82]. Environmental predictors were
595 the same aridity and temperature gradients used for trait-environment clines, in addition to
596 precipitation seasonality (all representative of climatic variation in cheatgrass genotypes;
597 Supplementary Fig. 6e). For each NGSadmix ancestral cluster (K=4), GAMs were implemented
598 with the function gam in the R package mgcv (v.1.9-1)[83] with a logit link function and beta-

599 distributed residuals. Genotypes were assigned to a cluster based on having >0.55 ancestry
600 proportion for the NGSadmix K=4 ancestral genetic clusters. Intermediate genotypes *(i.e.*,
601 composed of multiple ancestries) were excluded from this analysis.
602
603 **Phenotypes.** During the grow out in 2020, we measured phenotypes on up to 184 genotypes with
604 2–3 replicates that emerged within ~9–18 days of planting and survived until harvesting. Eleven
605 phenotypes were recorded: seedling and adult (*i.e.*, reproductive) height (used to get spring
606 growth), number of leaves, number of tillers, days to flower, inflorescence height, dry biomass,
607 total seed mass (*i.e.*, fecundity), individual seed mass (*i.e.*, seed mass), total seed length, and awn
608 length (Supplementary Methods 12).
609
610 After quality/error checking, the best linear unbiased estimate (BLUE) of phenotypes was
611 calculated per genotype with the BLUE function in the R package polyqtlR (v.0.1.1)[84], using
612 genotype as the predictor of trait measurements across 2–3 replicates and tray as a random effect.
613 We then calculated broad sense heritability (H^2) of traits as the proportion of phenotypic variance
614 explained by genotype in a linear model. The total set of phenotyped genotypes included: 184
615 with vegetative height/growth/count data, 173 with flowering/inflorescence data, 178 with dry
616 biomass data, and 182 with seed data, for a total of 169 genotypes with no missing phenotypes.
617
618 **Trait variation and environmental associations.** To detect axes of life history variation, we
619 summarized the natural genetic variation in our growth chamber phenotypes with PCA (function
620 prcomp, variables scaled and centered) in R. PC1, PC2, and flowering time, were then used as
621 response variables for investigating trait differences between ranges and environmental gradients
622 in phenotypes. To assess differences in trait means between native and invasive genotypes we

623 implemented two-tailed t-tests as well as linear mixed models that accounted for kinship between
624 genotypes (see below). To assess phenotypic differentiation between groups of nearly clonal
625 genotypes, we performed PERMANOVA with group identity as the predictor and eleven
626 phenotypes as response with function adonis2 in the R package vegan (9999 permutations,
627 Euclidian distances). To assess trait-environment clines we used kinship linear-mixed models,
628 which when significant (*i.e.*, $p \leq 0.05$), they provide evidence of selection (vs. population genetic
629 structure/drift) explaining variation, similar to Q_{ST}–F_{ST} tests[49]. A kinship matrix was estimated
630 with identity-by-state (IBS, *i.e.*, allele sharing between pairs of genotypes) using the dataset of
631 15,101,725 SNPs and function snpgdsIBS in the R package SNPRelate. This kinship matrix was
632 used to fit linear mixed models with random genotype effects using function lmekin in the R
633 package coxme (v.2.2-20)[85]. We focused on maximum monthly vapor pressure deficit (Pa),
634 describing aridity, and mean air temperature of the coldest quarter (ºC), describing winter
635 temperature. These two climate variables showed the highest loads on PC1 and PC2,
636 respectively, on a PCA of 52 climate variables for the 307 native and invasive genotypes
637 (Supplementary Fig. 6e). To test if clines were repeated, absent, or shifted in the invaded relative
638 to the native range, we also tested for an interaction between ranges in the models.
639
640 **Field common gardens.** We conducted a replicated common garden experiment with cheatgrass
641 genotypes in the 2022 and 2023 growing seasons, across two sites in the Intermountain West that
642 varied in their regional climatic conditions: a cool site with little temperature seasonality (Sheep
643 Station, ID [44.2456ºN, 112.2144ºW]) and a warm site with pronounced temperature seasonality
644 (Wildcat, ID [43.4744ºN, 116.9018ºW]). We grew replicates of 95 genotypes from Fall 2021 to
645 Spring 2022 and 93 genotypes from Fall 2022 to Spring 2023 at two different densities (low=100

646 seeds/1 m^2; high=100 seeds/0.04 m^2) and under two different temperature treatments (low=white
647 gravel; high=black gravel) in a factorial design at both sites[50] (Supplementary Methods 13).
648
649 We compared the direction and magnitude of selection on flowering phenology between the
650 cold, less seasonal site and the warm, seasonal site for both the 2022 and 2023 growing seasons.
651 For each year and common garden site combination, we calculated the average fitness
652 (2022=seed count; 2023=reproductive biomass) and average first flowering day for each
653 genotype, across all treatments. In the calculations of average first flowering day, if a plant did
654 not flower (*i.e.*, fitness=0), it was assigned the average first flowering day for all plants of that
655 genotype that did flower at some point during the growing season. For each growing season year,
656 we regressed the mean fitness data on the mean flowering time data across genotypes and
657 compared the slopes between the cold, less seasonal site and the warm, seasonal site. Positive
658 slopes on this graph indicate that flowering later is selected, while negative slopes indicate that
659 flowering earlier is selected. Models that included a random intercept for each genotype, with a
660 correlation structure specified by a kinship matrix, also provided evidence of selection. To
661 confirm that our larger dataset on flowering time from a growth chamber was consistent with
662 field measurements, we compared these values and found them correlated with Pearson r=0.6.
663
664 **Genome wide association studies (GWAS).** To identify QTL for growth chamber phenotypes,
665 we implemented GWAS that controlled for kinship on our BLUEs dataset (n=173–184
666 genotypes, 14.6–14.7M SNPs excluding SNPs with MAF<0.05) using two methods: a univariate
667 linear mixed model (LMM) and a multilocus mixed model (MLMM). SNP genotype data were
668 generated with function snpgdsGetGeno in the R package SNPRelate. The univariate LMM was
669 fit with gemma (v.0.98.5)[86] with an IBS matrix as the random effect accounting for relatedness

670 between genotypes. The MLMM was implemented with FarmCPUpp[87], which computes a
671 restricted kinship matrix based on pseudo-quantitative trait nucleotides (QTN) selected from a
672 preliminary GWAS step. FarmCPUpp also takes principal components of genome-wide SNP
673 variation as covariates (here PC1–PC3), providing a stronger control of population genetic
674 structure. Principal components were obtained with function snpgdsPCA in the R package
675 SNPRelate. Moreover, while gemma tests for an association with each SNP individually,
676 FarmCPUpp performs additional steps that detect pseudo-QTNs based on LD and uses model
677 selection and multiple regression to retain the best set of pseudo-QTNs. Thus, while gemma
678 might reveal large blocks of significantly associated SNPs (which might correlate with
679 chromosomal rearrangements), this signal should be lost with FarmCPUpp (which might be
680 better at detecting causal SNPs). To detect statistical significance of GWAS SNPs, we used a
681 false-discovery-rate (FDR) threshold of 0.05 on output *p*-values. Associations were inspected
682 with Manhattan plots and model fits were assessed with quantile-quantile (Q-Q) plots.
683
684 **Linkage disequilibrium (LD) and SNP annotation.** To assign GWAS SNPs to annotated
685 cheatgrass genes, we first investigated linkage disequilibrium (LD) decay with genomic distance
686 in our sequenced panel of 307 genotypes. We used PopLDdecay[88], which takes a .vcf with
687 samples and computes the square of Pearson correlations (R^2) between pairs of SNPs genome-
688 wide (using the ~15M SNPs dataset) or per chromosome. We excluded SNPs with MAF<0.05
689 and >5% heterozygote individuals and calculated R^2 between pairs of SNPs at a minimum of 10
690 bp apart and a maximum of 5 Mb apart. We detected the pattern of LD decay by plotting mean
691 LD values in 100 bp bins from 0–5Mb genomic distance. We observed substantial long-range
692 LD with mean R^2~0.3 even at 5Mb (Supplementary Fig. 12a), in line with strong population
693 structure. However, there was a clear decay in LD with genomic distance. The initial mean LD

694 (R^2=0.45) at 10 bp decayed halfway (R^2=0.376) of the minimum observed LD (R^2=0.3) by 194.5
695 kb. Because mean LD was high even at 5Mb, we also assessed inter-chromosomal LD
696 (Supplementary Methods 14).
697
698 **QTL-environment clines and enrichment analysis.** Environmental variation of allele
699 frequency in QTL detected with GWAS was examined with two-tailed t-tests and kinship linear-
700 mixed models. QTL-environment tests were performed separately for the native and invaded
701 range. To find significantly over-represented GO terms or parents of these terms in the
702 haploblock detected for flowering time, we uploaded a gene set to the PlantRegMap GO Term
703 Enrichment tool[89], based on *A. thaliana* and *O. sativa* homologs obtained from CoGe
704 annotations.
705
706 **Genome-environment matching of invasive genotypes.** We further tested for evidence of pre-
707 adaptation by quantifying how well a native range genotype-environment association (GEA)
708 model predicted genetic composition in the invaded range. Our approach is like the genomic
709 offset statistics used for predicting climate change impacts on maladaptation[55,80]. We used the
710 RDA-generated native range GEA to predict maladaptation of invaded range genotypes,
711 separately for WNA and ENA, with the function 'predict' in the R package vegan. The Euclidian
712 distance between predicted and observed allele frequencies for each invasive genotype was then
713 estimated, representing the genetic maladaptation to the invaded range site. To evaluate if the
714 mean genetic maladaptation WNA and ENA was different from random, predictions and genetic
715 (*i.e.*, Euclidean) distances were recalculated in 1000 reshuffled environments. Genetic
716 maladaptation was considered significantly lower than expected by chance if it fell in the lower
717 0.025 tail of the random distribution of mean genomic distances in a two-tailed test.

718

719 **Genome-environment matching and invasive spread.** We examined if cheatgrass dominance
720 was correlated with the strength of local adaptation using available cheatgrass abundance data.
721 Using random forest models, Bradley *et al.*[30] created a regional classification of cheatgrass
722 presence across the Great Basin based on 11,307 surveyed sites. Their classification
723 differentiates between cheatgrass present at high abundance (≥ 15%) and cheatgrass absent or at
724 low abundance (< 15%). Genetic maladaptation/offset was compared between sites where
725 cheatgrass is in high abundance to sites where cheatgrass is in low abundance (high vs. low in
726 Fig. 5c, respectively) with a two-tailed t-test (n=54 sites or Great Basin genotypes). To assess if
727 this pattern was driven by the match of local genotypes to local environments (as opposed to the
728 environmental characteristics of the low abundance cheatgrass sites), we recalculated genomic
729 offset in 1000 reshuffled Great Basin environments and compared offset of regions where
730 cheatgrass dominates to the 1000 null permutations of genotypes within the Great Basin.

731

732 **Data availability**

733 Seeds of *Bromus tectorum* genotypes used in this study are available from the lead contact upon
734 request and will be deposited to GRIN. Herbarium vouchers of genotypes will be deposited at
735 The Pennsylvania State University Herbarium (PAC). Whole-genome sequences for 303
736 genotypes are being deposited in the European Nucleotide Archive (ENA;
737 https://www.ebi.ac.uk/ena). Whole-genome sequences for four genotypes sequenced by the DOE
738 Joint Genome Institute are publicly available at their website under proposal ID: 506608. The
739 beagle (genotype-likelihoods), vcf (SNP calls), and raw phenotype data will be publicly available
740 at Figshare[90]. Genotype-level geographic information, climate data, growth chamber phenotypes,

741 ancestry proportions, GWAS SNPs, and field common gardens data are in Supplementary Data
742 1. Gene annotations of GWAS results (100 top SNPs) are in Supplementary Data 2.
743

744 **Code availability**

745 Code used to analyze NGS data will be publicly available at Figshare[90].
746

747 **References**

748 1. Daly, E. Z. et al. A synthesis of biological invasion hypotheses associated with the
749 introduction–naturalization–invasion continuum. *Oikos* **2023**, e09645 (2023).
750 2. Catford, J. A., Jansson, R. & Nilsson, C. Reducing redundancy in invasion ecology by
751 integrating hypotheses into a single theoretical framework. *Diversity and Distributions* **15**,
752 22–40 (2009).
753 3. Liu, D. et al. Regional invasion history and land use shape the prevalence of non-native
754 species in local assemblages. *Glob. Chang. Biol.* **30**, e17426 (2024).
755 4. Colautti, R. I. & Barrett, S. C. H. Rapid adaptation to climate facilitates range expansion of
756 an invasive plant. *Science* **342**, 364–366 (2013).
757 5. Urban M. C. et al. Evolutionary origins for ecological patterns in space. *Proc. Natl. Acad.*
758 *Sci.* **117**, 17482–17490 (2020).
759 6. Tigano, A. & Friesen, V. L. Genomics of local adaptation with gene flow. *Mol. Ecol.* **25**,
760 2144-2164 (2016).
761 7. Bridle, J. R. & Vines, T. H. Limits to evolution at range margins: When and why does
762 adaptation fail? *Trends Ecol. Evol*. **22**, 140–147 (2007).

763 8. Simón-Porcar, V. I., Silva, J. L. & Vallejo-Marín, M. Rapid local adaptation in both sexual
764 and asexual invasive populations of monkeyflowers (*Mimulus* spp.). *Ann. Bot.* **127**, 655–
765 668 (2021).
766 9. Kirkpatrick, M. & Barton, N. H. Evolution of a species' range. *Am. Nat.* **150**, 1–23 (1997).
767 10. Eckert, C. G., Samis, K. E. & Lougheed, S. C. Genetic variation across species'
768 geographical ranges: the central–marginal hypothesis and beyond. *Mol. Ecol.* **17**, 1170–
769 1188 (2008).
770 11. Mayr E. *Animal Species and Evolution* (Harvard University Press, Cambridge, 1963)
771 12. Robinson, J., Kyriazis, C. C., Yuan, S. C. & Lohmueller, K. E. Deleterious variation in
772 natural populations and implications for conservation genetics. *Annu. Rev. Anim. Biosci*. **11**,
773 93–114 (2023).
774 13. Bay, R. A. et al. Genomic signals of selection predict climate-driven population declines in
775 a migratory bird. *Science* **359**, 83–86 (2018).
776 14. Nota, A., Bertolino, S., Tiralongo, F. & Santovito, A. Adaptation to bioinvasions: When
777 does it occur? *Glob. Chang. Biol.* **30**, e17362 (2024).
778 15. Rosche, C. et al., Climate outweighs native vs. nonnative range-effects for genetics and
779 common garden performance of a cosmopolitan weed. *Ecol. Monogr.* **89**, e01386 (2019).
780 16. Santangelo, J. S. et al., Global urban environmental change drives adaptation in white
781 clover. *Science* **375**,1275–1281 (2022).
782 17. Leger, E. A. & Rice, K. J. Assessing the speed and predictability of local adaptation in
783 invasive California poppies (*Eschscholzia californica*). *J. Evol. Biol.* **20**, 1090–1103 (2007).
784 18. Liu, C., Wolter, C., Xian, W. & Jeschke, J. M. Most invasive species largely conserve their
785 climatic niche. *Proc. Natl. Acad. Sci.* **117**, 23643–23651 (2020).

786 19. Sherpa, S. & Deprés, L. The evolutionary dynamics of biological invasions: A multi-
787 approach Perspective. *Evolutionary Applications* **14**, 1463–1484 (2021).
788 20. Turner, K. G., Ostevik, K. L., Grassa, C. J. & Rieseberg, L. H. Genomic analyses of
789 phenotypic differences between native and invasive populations of diffuse knapweed
790 (*Centaurea diffusa*). *Front. Ecol. Evol.* **8**, 577635 (2021).
791 21. Kreiner, J. M., Caballero, A., Wright, S. I. & Stinchcombe, J. R. Selective ancestral sorting
792 and de novo evolution in the agricultural invasion of *Amaranthus tuberculatus*. *Evolution*
793 **76**, 70–85 (2022).
794 22. Endriss, S. B., Alba, C., Norton, A. P., Pyšek, P. & Hufbauer, R. A. Breakdown of a
795 geographic cline explains high performance of introduced populations of a weedy invader.
796 *J. Ecol.* **106**, 699–713 (2018).
797 23. Gioria, M., Hulme, P. E., Richardson, D. M. & Pyšek, P. Why are invasive plants
798 successful? *Annu. Rev. Plant Biol.* **74**, 635-670 (2023).
799 24. Vallejo-Marín, M. et al. Population genomic and historical analysis suggests a global
800 invasion by bridgehead processes in *Mimulus guttatus*. *Commun. Biol.* **4**, 327 (2021).
801 25. Bieker, V. C. et al. Uncovering the genomic basis of an extraordinary plant invasion. *Sci.*
802 *Adv.* **8**, eabo5115 (2022).
803 26. Lucas, M. S. et al. Re-focusing sampling, design and experimental methods to assess rapid
804 evolution by non-native plant species. *Biol. Invasions* **26**, 1327–1343 (2024).
805 27. Mack, R. N. Invasion of *Bromus tectorum* L. into western North America: An ecological
806 chronicle. *Agro-Ecosyst.* **7**, 145–165 (1981).
807 28. Novak, S. J. & Mack, R. N. "Chapter 4. Mating system, introduction and genetic diversity
808 of *Bromus tectorum* in North America, the most notorious product of evolution within
809 *Bromus* section Genea" in *Exotic Brome-Grasses in Arid and Semiarid Ecosystems of the*

Western US: Causes, Consequences, and Management Implications, Germino, M. J., Chambers, J. C. & Brown, C. S. Eds. (Springer International Publishing, Cham, 2016), pp. 99–132.

29. Germino, M. J., Belnap, J., Stark, J. M., Allen, E. B. & Rau, B. M. "Chapter 3. Ecosystem impacts of exotic annual invaders in the genus *Bromus*" in *Exotic Brome-Grasses in Arid and Semiarid Ecosystems of the Western US: Causes, Consequences, and Management Implications*, Germino, M. J., Chambers, J. C. & Brown, C. S. Eds. (Springer International Publishing, Cham, 2016), pp. 61–95.
30. Bradley, B. A. et al. Cheatgrass (*Bromus tectorum*) distribution in the intermountain western United States and its relationship to fire frequency, seasonality, and ignitions. *Biol. Invasions* **20**, 1493–1506 (2018).
31. Porensky, L. M. & Blumenthal, D. M. Historical wildfires do not promote cheatgrass invasion in a western Great Plains steppe. *Biol. Invasions* **18**, 3333–3349 (2016).
32. Revolinski, S. R., Maughan, P. J., Coleman, C. E. & Burke, I. C. Preadapted to adapt: Underpinnings of adaptive plasticity revealed by the downy brome genome. *Commun. Biol.* **6**, 326 (2023).
33. Novak, S. J. & Mack, R. N. Genetic variation in *Bromus tectorum* (Poaceae): comparison between native and introduced populations. *Heredity* **71**, 167–176 (1993).
34. Barlett, E., Novak, S. J. & Mack, R. N. Genetic variation in *Bromus tectorum* (Poaceae): differentiation in the eastern United States. *Am. J. Bot.* **89**, 602–612 (2002).
35. Valliant, M. T., Mack, R. N. & Novak, S. J. Introduction history and population genetics of the invasive grass *Bromus tectorum* (Poaceae) in Canada. *Am. J. Bot.* **94**, 1156–1169 (2007).

833 36. Schachner, L. J., Mack, R. N. & Novak, S. J. *Bromus tectorum* (Poaceae) in midcontinental
834 United States: Population genetic analysis of an ongoing invasion. *Am. J. Bot.* **95**, 1584–
835 1595 (2008).
836 37. Huttanus, T. D., Mack, R. N. & Novak, S. J. Propagule pressure and introduction pathways
837 of *Bromus tectorum* (cheatgrass; Poaceae) in the central United States. *Int. J. Plant Sci.*
838 **172**, 783–794 (2011).
839 38. Pawlak, A. R., Mack, R. N., Busch, J. W. & Novak, S. J. Invasion of *Bromus tectorum* (L.)
840 into California and the American southwest: Rapid, multi-directional and genetically
841 diverse. *Biol. Invasions* **17**, 287–306 (2015).
842 39. Leger, E. A., Espeland, E. K., Merrill, K. R. & Meyer, S. E. Genetic variation and local
843 adaptation at a cheatgrass (*Bromus tectorum*) invasion edge in western Nevada. *Mol. Ecol.*
844 **18**, 4366–4379 (2009).
845 40. Hufft, R. A. & Zelikova, T. J. "Chapter 5. Ecological genetics, local adaptation, and
846 phenotypic plasticity in *Bromus tectorum* in the context of a changing climate" in *Exotic*
847 *Brome-Grasses in Arid and Semiarid Ecosystems of the Western US: Causes,*
848 *Consequences, and Management Implications*, Germino, M. J., Chambers, J. C. & Brown,
849 C. S. Eds. (Springer International Publishing, Cham, 2016), pp. 133–154.
850 41. Meyer, S. E. & Allen, P. S. Ecological genetics of seed germination regulation in *Bromus*
851 *tectorum* L. I. Phenotypic variance among and within populations. *Oecologia* **120**, 27–34
852 (1999).
853 42 Meyer, S. E. & Allen, P. S. Ecological genetics of seed germination regulation in *Bromus*
854 *tectorum* L. II. Reaction norms in response to a water stress gradient imposed during seed
855 maturation. *Oecologia* **120**, 35–43 (1999).

856 43. Meyer, S. E., Nelson, D. L. & Carlson, S. L. Ecological genetics of vernalization response
857 in *Bromus tectorum* L. (Poaceae). *Ann. Bot.* **93**, 653–663 (2004).
858 44. Meyer, S. E., Leger, E. A., Eldon, D. R. & Craig, C. E. Strong genetic differentiation in the
859 invasive annual grass *Bromus tectorum* across the Mojave–Great Basin ecological transition
860 zone. *Biol. Invasions* **18**, 1611–1628 (2016).
861 45. Kelly, L. J., Mack, R. N. & Novak, S. J. Genetic analysis of *Bromus tectorum* (Poaceae) in
862 the Mediterranean region: biogeographical pattern of native populations. *Heredity* **126**,
863 178–193 (2021).
864 46. Zeitler, L. & Gilbert, K. J. Using runs of homozygosity and machine learning to disentangle
865 sources of inbreeding and infer self-fertilization rates. *Genome Biol. Evol.* **16**, evae139
866 (2024).
867 47. Meyer, S. E., Ghimire, S., Decker, S., Merrill, K. R. & Coleman, C. E. The ghost of
868 outcrossing past in downy brome, an inbreeding annual grass. *J. Hered.* **104**, 476–490
869 (2013).
870 48. Hierro, J. L., Eren, Ö., Čuda, J. & Meyerson, L. A. Evolution of increased competitive
871 ability may explain dominance of introduced species in ruderal communities. *Ecol.*
872 *Monogr.* **92**, e1524 (2022).
873 49. McKay, J. K. & Latta, R. G. Adaptive population divergence: Markers, QTL and traits.
874 *Trends Ecol. Evol.* **17**, 285–291 (2002).
875 50. Vahsen, M. L. et al. Phenological sensitivity of *Bromus tectorum* genotypes depends on
876 current and source environments. *Ecology* **106**, e70025 (2025).
877 51. Todesco, M. et al. Massive haplotypes underlie ecotypic differentiation in sunflowers.
878 *Nature* **584**, 602–607 (2020).

879 52. Battlay, P. et al. Large haploblocks underlie rapid adaptation in the invasive weed *Ambrosia*
880 *artemisiifolia*. *Nat. Commun.* **14**, 1717 (2023).
881 53. Resentini, F., Orozco-Arroyo, G., Cucinotta, M. & Mendes, M. A. The impact of heat stress
882 in plant reproduction. *Front. Plant Sci.* **14**, 1271644 (2023).
883 54. Holman, T. J. et al. The N-end rule pathway promotes seed germination and establishment
884 through removal of ABA sensitivity in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **106**, 4549–4554
885 (2009).
886 55. Gain, C. et al. A quantitative theory for genomic offset statistics. *Mol. Biol. Evol.* **40**,
887 msad140 (2023).
888 56. Morrow, L. A. & Stahlman, P. W. The history and distribution of Downy Brome (*Bromus*
889 *tectorum*) in North America. *Weed Sci.* **32**, 2–6 (1984).
890 57. Razanajatovo, M. et al. Plants capable of selfing are more likely to become naturalized. *Nat.*
891 *Commun.* **7**, 13313 (2016).
892 58. van Kleunen, M., Manning, J. C., Pasqualetto, V. & Johnson, S. D. Phylogenetically
893 independent associations between autonomous self-fertilization and plant invasiveness. *Am.*
894 *Nat.* **171**, 195–201 (2008).
895 59. NA, A. et al. Niche shift in invasive species: is it a case of "home away from home" or
896 finding a "new home"? *Biodivers. Conserv.* **31**, 2625–2638 (2022).
897 60. Smith, A. L. et al. Global gene flow releases invasive plants from environmental constraints
898 on genetic diversity. *Proc. Natl. Acad. Sci.* **117**, 4218–4227 (2020).
899 61. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture
900 proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
901 62. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in
902 low-depth NGS data. *Genetics* **210**, 719–731 (2018).

903 63. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using
904 the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
905 64. Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing
906 phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
907 65. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci.* **70**,
908 3321–3323 (1973).
909 66. Weir, B. S. & Goudet, J. A unified characterization of population structure and relatedness.
910 *Genetics* **206**, 2085–2103 (2017).
911 67. Zheng, X. et al. A high-performance computing toolset for relatedness and principal
912 component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
913 68. Nei, M. & Li, W.-H. Mathematical model for studying genetic variation in terms of
914 restriction endonucleases. *Proc. Natl. Acad. Sci.* **76**, 5269–5273 (1979).
915 69. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008
916 (2021).
917 70. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158
918 (2011).
919 71. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA
920 polymorphism. *Genetics* **123**, 585–595 (1989).
921 72. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of
922 homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet*. **19**,
923 220–234 (2018).
924 73. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based
925 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

74. Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V. & Pebesma, E. J. *Applied Spatial Data Analysis with R* (Springer, New York City, 2008).

75. Diniz-Filho, J. A. F. et al. Mantel test in population genetics. *Genet. Mol. Biol.* **36**, 475–485 (2013).

76. Dray, S. & Dufour, A.-B. The ade4 package: Implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).

77. Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L. & Karger, D. N. Global climate-related predictors at kilometer resolution for the past and future. *Earth Syst. Sci. Data* **14**, 5573–5603 (2022).

78. Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L. & Karger, D. N. Data from: CHELSA-BIOCLIM+ A novel set of global climate-related predictors at kilometer-resolution. EnviDat. (2022); https://doi.org/10.16904/envidat.332

79. Oksanen, J. et al. vegan: Community ecology package, version 2.6-4 (2022); https://CRAN.R-project.org/package=vegan.

80. Capblancq, T. & Forester, B. R. Redundancy analysis: A Swiss army knife for landscape genomics. *Methods Ecol. Evol.* **12**, 2298–2309 (2021).

81. Dray, S., Legendre, P. & Peres-Neto, P. R. Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Model.* **196**, 483–493 (2006).

82. Hastie, T. J. & Tibshirani, R. J. *Generalized Additive Models*. (Chapman and Hall, New York City, 1990).

83. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 3–36 (2011).

950 84. Bourke, P. M. et al. Detecting quantitative trait loci and exploring chromosomal pairing in
951 autopolyploids using polyqtlR. *Bioinformatics* **37**, 3822–3829 (2021).
952 85. Therneau, T. M. coxme: Mixed effects cox models, version 2.2-20 (2022);
953 https://CRAN.R-project.org/package=coxme.
954 86. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association
955 studies. *Nat. Genet.* **44**, 821–824 (2012).
956 87. Kusmec, A. & Schnable, P. S. FarmCPUpp: Efficient large-scale genomewide association
957 studies. *Plant Direct*. **2**, e00053 (2018).
958 88. Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M. & Yang, T.-L. PopLDdecay: a fast and
959 effective tool for linkage disequilibrium decay analysis based on variant call format files.
960 *Bioinformatics* **35**, 1786–1788 (2019).
961 89. Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J. & Gao, G. PlantRegMap: Charting functional
962 regulatory maps in plants. *Nucleic Acids Res.* **2019**, gkz1020 (2019).
963 90. Figshare dataset and code for this study with repository DOI will be available here.
964

965 **Acknowledgments**: Any opinions, findings, conclusions, or recommendations expressed in the
966 material are those of the authors and should not be construed to represent any official USDA
967 determination or policy. Any use of trade, firm, or product names is for descriptive purposes only
968 and does not imply endorsement by the U.S. Government. Xavier Mack, Carlos Rodríguez-
969 Gonzalez, Yuxin Luo, and Katherine Blocklove helped with measuring phenotypes, harvesting,
970 and performing DNA extractions. Ian Burke, Samuel Revolinski, Peter Maughan, and Craig
971 Coleman granted us access to the reference genome prior to publication. The USGS FIREss team
972 contributed to common garden establishment and data collection on the Wildcat site. iNaturalist
973 was a key resource for identifying seed collectors, among those were also Dave Barnett, David

Board, Chalon Boesel, John Bradford, Jaime Braschi, Howard Bruner, Victoria Bustamante, Charles Campbell, Jeanne Chambers, Mike Chen, Mark Chynoweth, Jason Cooper, Massimo Cristofaro, Melodie Cunningham, Kirk Davies, Janelle Downs, Torsten Eriksson, Maggie Eshleman, Erica Fleishman, Carol Kadonsky & Bill Foreman, Berit Gehrke, Tom Getts, Richard Gill, Dana Hartel, Nate Hartley, Patricia Hollins, Alex Hood, Tayla Hook, Parker Hopkins, Becky Hufft, Jennifer Kalt, Lorri Kendrick, Molly Ladd, Matt Lavin, Steven Lee, Jonathan Levine, Marisa Mancillas, John Maron, Grace McCartha, Randal Mindell, Chandra Moffat, Brooke Moore, James Nagler, Yael Orgad, Matthew Pedrotti, David Pyke, Sasha Reed, Matt Rinella, Viktoria Ropak, Håkan Rydin, Geno Schupp, Adam Searcy, Tim Springer, Amy Symstad, Tracy Thomas, Trudy Trevarthen, Samantha vanDeurs, Biljana Vidovic, Viktoria Wagner, Gretchen Whetham, David Wilderman, Eileen Wyza, and Pauline & Jon Zweck.

Funding: National Science Foundation grant DEB-1927282 (PBA), DEB-1927009 (JRL), DEB-1927177 (MBH); Joint Genome Institute of the U.S. Department of Energy grant New Investigator Award-506608 (JRL); National Institutes of Health grant R35GM138300 (JRL).

Author contributions

Conceptualization: DG, CSB, EAL, DMB, MJG, LMP, MBH, PBA, JRL

Methodology: DG, TMM, NP, SR, AP, CSB, EAL, DMB, MJG, LMP, MBH, PBA, JRL

Investigation: DG, MLV, AD, RB-Z, EAL, DMB, MJG, LMP, PBA, JRL, OB, CSB, RB, SB-M, SMC, SLD-F, AD, RCD, PWD, DJE, TAV, HH, MCH, RAH, SJ, JMK, EK, SK, MK, AK, ML, SL, DM, CLM, JWM, DUN, JPO, RP, LAP, LR, BGR, CR, MS, RKS, AS, BMS, RLS, KGT, AKU, AVW, C-AW, JZ.

Funding acquisition: PBA, MBH, JRL, MJG

Project administration: PBA, MBH, CSB, DMB, MJG, LMP, JRL

Writing – original draft: DG, JRL

Writing – review & editing: all authors

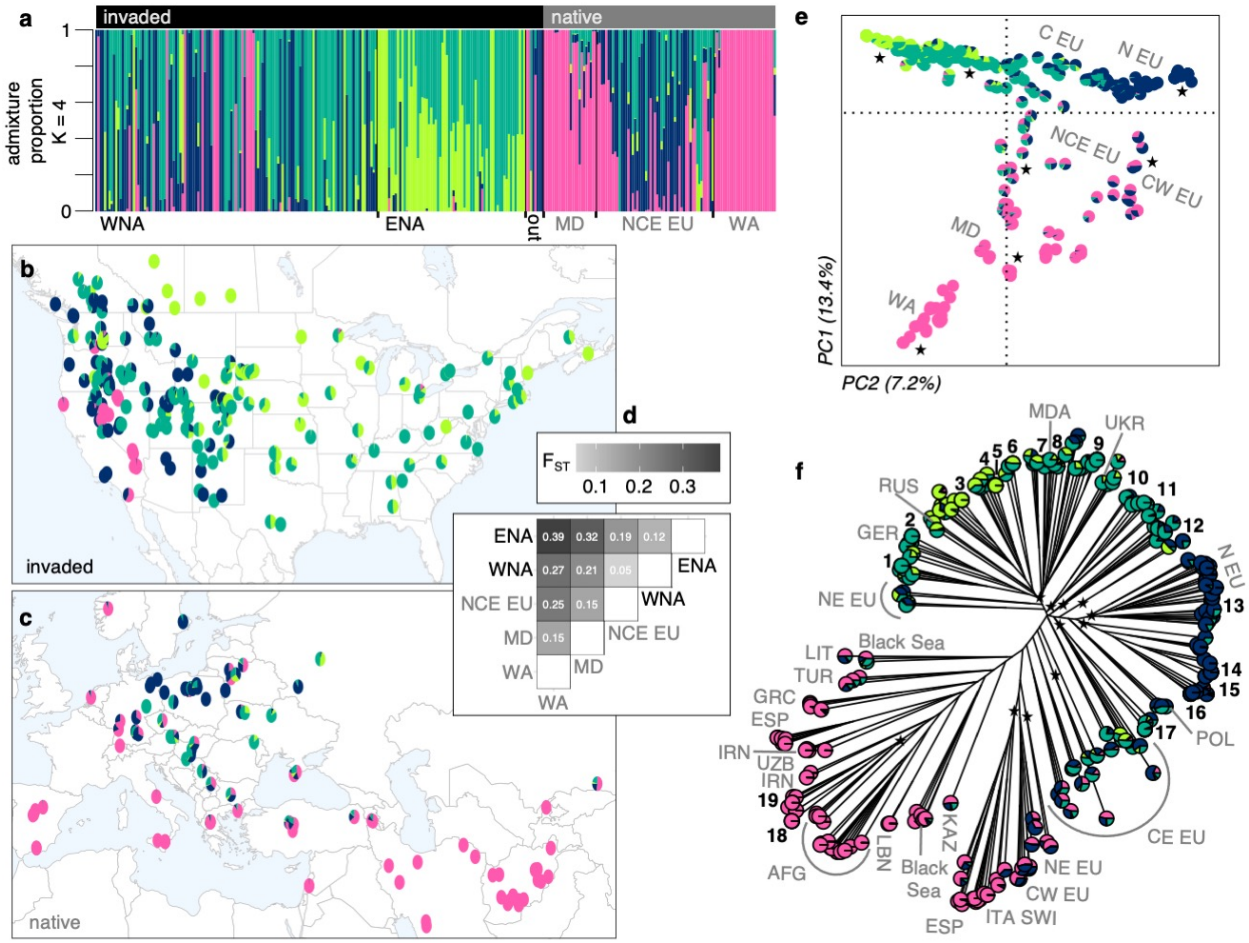
Competing interests: Authors declare that they have no competing interests.

Additional information

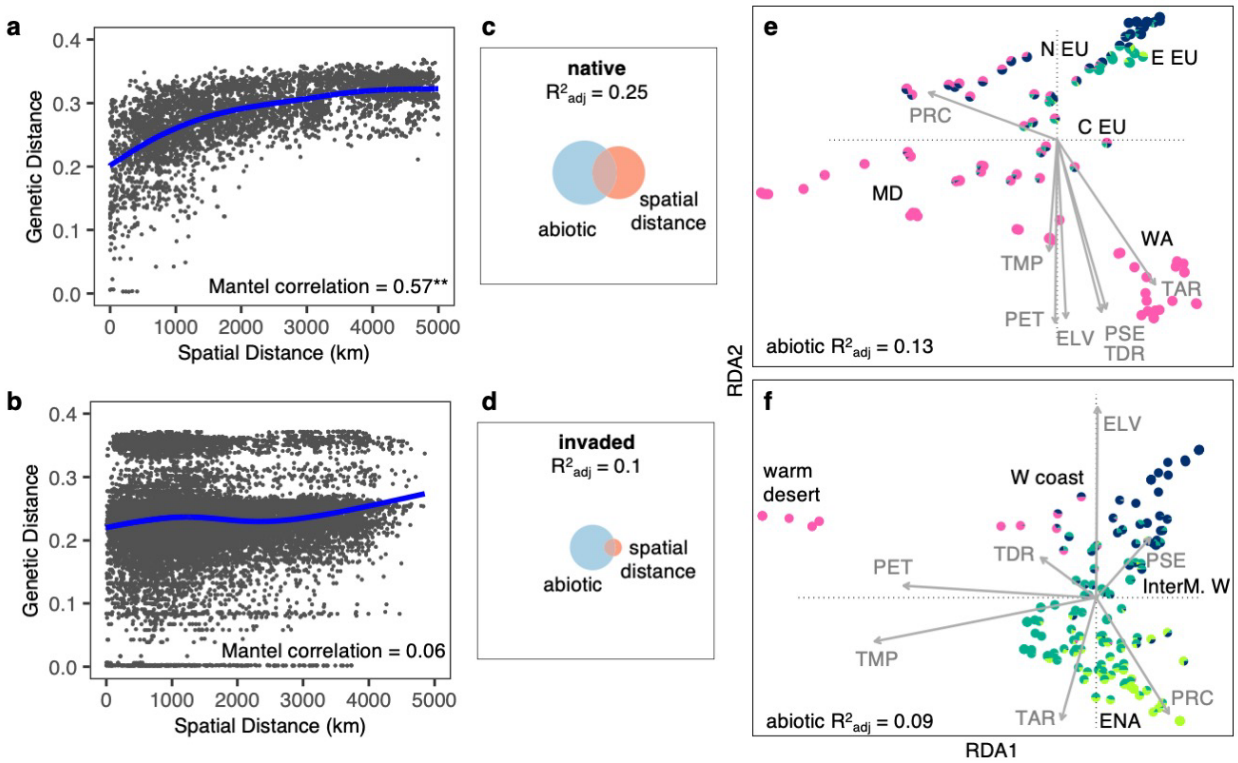
Supplementary information is available for this paper.

Correspondence and requests for materials should be addressed to D.G.

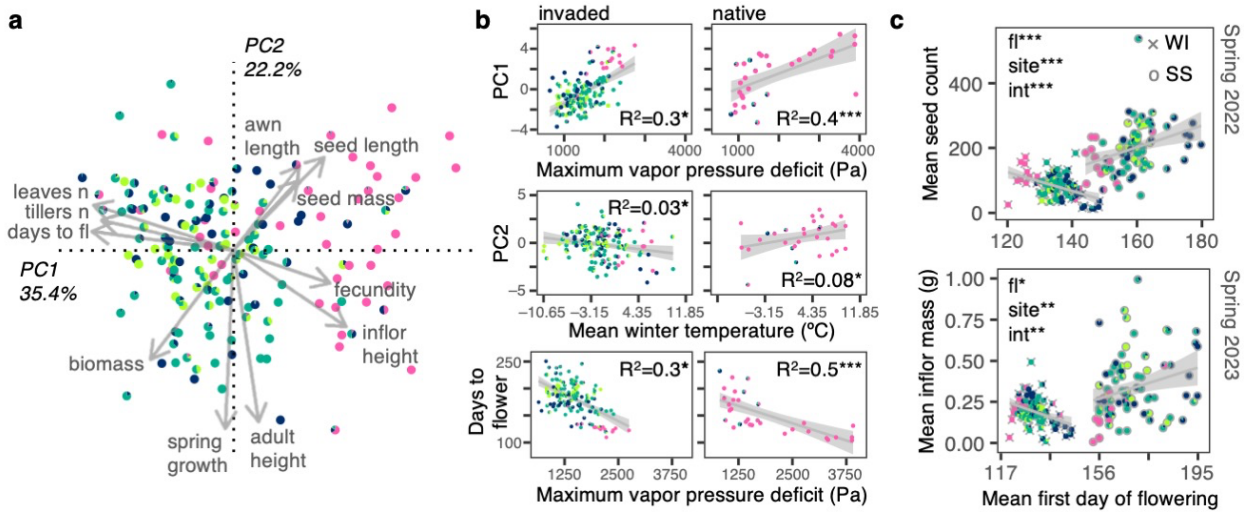
Reprints and permission information is available at http://www.nature.com/reprints


a
invaded
native
admixture proportion K = 4
1
0
WNA
ENA
out
MD
NCE EU
WA
b
invaded
c
native
d
F_{ST}
0.1 0.2 0.3
ENA 0.39 0.32 0.19 0.12
WNA 0.27 0.21 0.05
NCE EU 0.25 0.15
MD 0.15
WA
e
C EU
N EU
NCE EU
CW EU
MD
WA
PC1 (13.4%)
PC2 (7.2%)
f
MDA
UKR
RUS
GER
NE EU
N EU
Black Sea
LIT
TUR
GRC
ESP
IRN
UZB
IRN
POL
CE EU
AFG
LBN
KAZ
Black Sea
NE EU
CW EU
ITA SWI
ESP

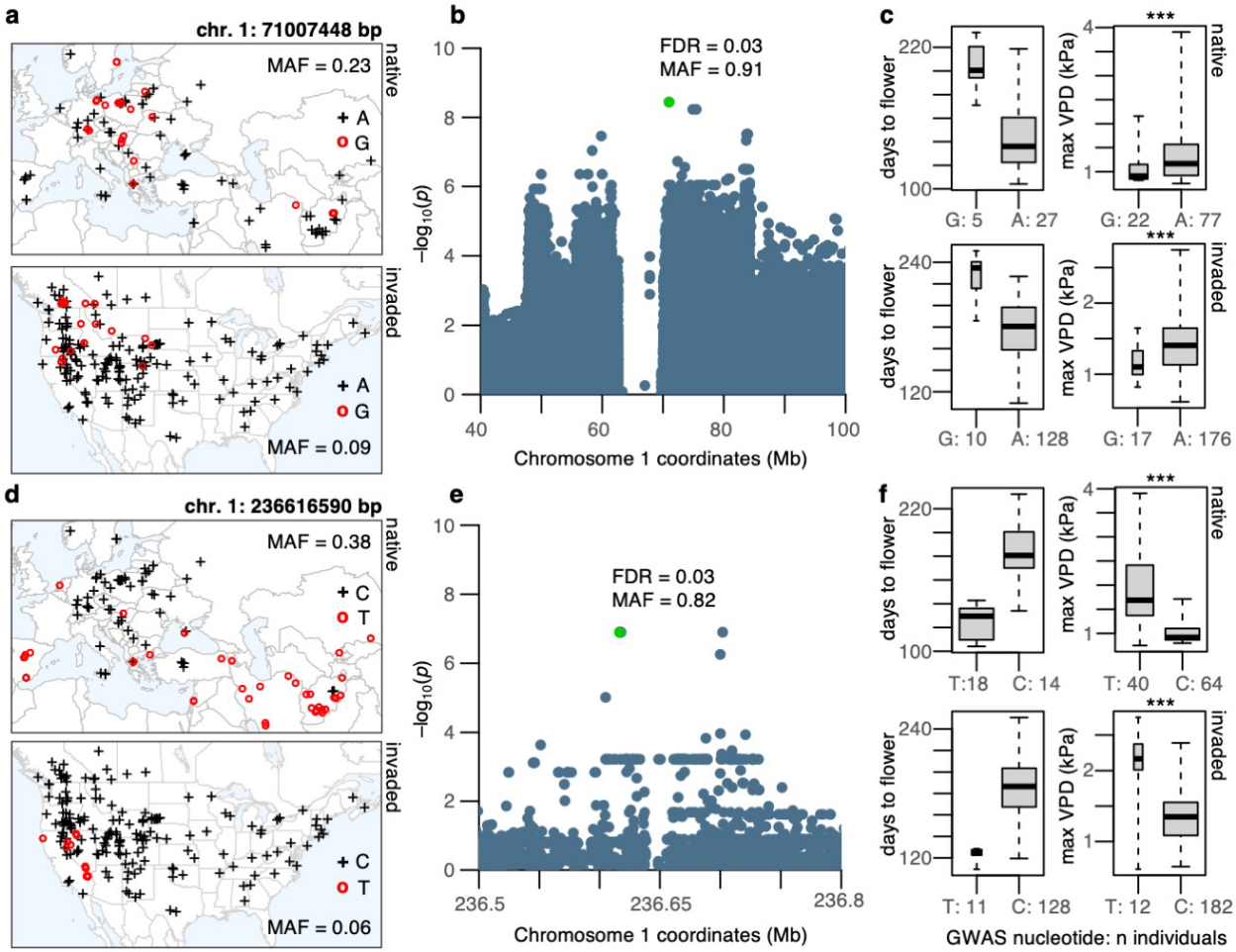
1033 **Fig. 1: The cheatgrass invasion involved multiple diverse introductions from the native**
1034 **range to North America** (**a**) Admixture proportions for K=4 ancestral genetic clusters (colors)
1035 for invasive and native genotypes in different regions; WNA: western North America (n=107),
1036 ENA: eastern North America (n=67), out: not in North America (n=8), MD: Mediterranean
1037 (n=24), NCE EU: north-central-east Europe (n=53), WA: west Asia (n=28). Geographic
1038 distribution of (**b**) invasive (n=194, North American only) and (**c**) native (n=105) genotypes. (**d**)
1039 Genetic differentiation (F_{ST}) between native and invaded regions, with notations following panel
1040 a. (**e**) Principal components analysis showing PC1 (y-axis) and PC2 (x-axis) explaining 20.6% of
1041 genomic variation. Axes are shifted to better reflect the latitudinal distribution of genotypes.
1042 Gray letters denote geographic origin in the native range and stars represent genotypes in the
1043 invaded range. (**f**) Neighbor-joining tree annotated with native (gray letters) and invaded
1044 locations (black numbers and stars). Native notations follow the ISO alpha-3 country code or
1045 their cardinal direction in Europe (EU). Black numbers mark groups of 2–14 near-clonal, and
1046 often widely distributed, invasive genotypes. Stars mark branches with invasive genotypes.
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056


a
Genetic Distance
Spatial Distance (km)
Mantel correlation = 0.57**
b
Genetic Distance
Spatial Distance (km)
Mantel correlation = 0.06
c
native
R^2_{adj} = 0.25
abiotic
spatial distance
d
invaded
R^2_{adj} = 0.1
abiotic
spatial distance
e
N EU
E EU
C EU
PRC
MD
WA
TMP
TAR
PET
ELV
PSE
TDR
abiotic R^2_{adj} = 0.13
f
ELV
warm desert
W coast
TDR
PSE
PET
InterM. W
TMP
PRC
TAR
ENA
abiotic R^2_{adj} = 0.09
RDA1
RDA2

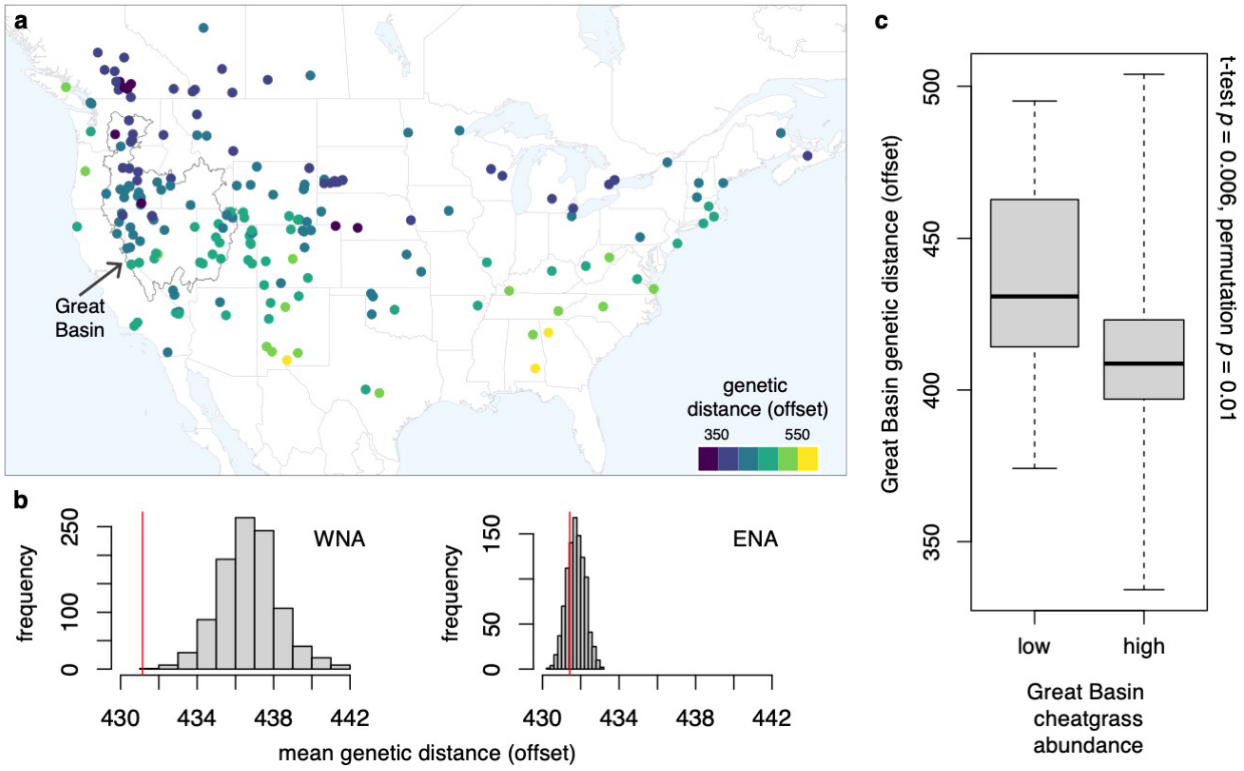
1070 **Fig. 2: Genomic variation is structured by environment in the native and invaded ranges**.
1071 Strong isolation-by-distance in the (**a**) native (**Mantel p=10^{-4}) but not in the (**b**) invaded range;
1072 plots show raw pair-wise data with a spline. Euler Plots show genomic variation is best explained
1073 by both the abiotic environment and spatial distance in (**c**) the native range, but only by the
1074 abiotic environment in (**d**) the invaded range. Fields of squares represent total genomic variation,
1075 circles represent genomic variation explained by a particular group of variables calculated using
1076 variance partitioning with RDA ordination (native n=105, invaded n=194). (**e**) Native and (**f**)
1077 invasive genotypes projected on the first two canonical axes of RDA (x-axis: RDA1 y-axis:
1078 RDA2). Arrows represent environmental predictors that strongly correlate with a maximal
1079 proportion of variation in linear combinations of SNPs. ELV: elevation, PET: potential
1080 evapotranspiration, PRC: total annual precipitation, PSE: precipitation seasonality, TAR:
1081 temperature annual range, TDR: temperature diurnal range, TMP: annual mean temperature.
1082 Colors are K=4 ancestral clusters. Geographic annotations are depicted in bolded black; N EU:
1083 north Europe, E EU: east Europe, C EU: central Europe, MD: Mediterranean, WA: west Asia, W
1084 coast: west coast, InterM. W: intermountain west, ENA: eastern North America.
1085
1086
1087
1088
1089
1090
1091
1092
1093


a
PC2
22.2%
PC1
35.4%
awn
length
seed length
seed mass
leaves n
tillers n
days to fl
fecundity
inflor
height
biomass
spring
growth
adult
height
b
invaded
native
PC1
R²=0.3*
R²=0.4***
Maximum vapor pressure deficit (Pa)
PC2
R²=0.03*
R²=0.08*
Mean winter temperature (°C)
Days to
flower
R²=0.3*
R²=0.5***
Maximum vapor pressure deficit (Pa)
c
fl***
site***
int***
WI
SS
Spring 2022
Mean seed count
fl*
site**
int**
Spring 2023
Mean inflor mass (g)
Mean first day of flowering

1111 **Fig. 3: Selection along aridity and temperature gradients shapes flowering phenology.** (**a**)
1112 Eigenvector plot with loadings of eleven phenotypes onto PC1 (x-axis) and PC2 (y-axis)
1113 describing axes of life history variation of 169 genotypes in a growth chamber; fl: Flowering, n:
1114 Number, inflor: Inflorescence. (**b**) Growth chamber phenotype-environment associations for
1115 invasive (left; n=138–145) and native genotypes (right; n=31–36). Coefficients of determination
1116 (R^2), trends (gray lines), and 95% confidence intervals (gray shades) come from linear
1117 regressions. Significance comes from linear-mixed kinship models that accounted for relatedness
1118 among genotypes: *p<0.05, ***p<0.0001. (**c**) Fitness advantage of early flowering genotypes at
1119 a warm site/common garden (WI: Wild Cat, gray crosses, n=93) and of late flowering genotypes
1120 at a cool site/common garden (SS: Sheep Station, gray open circles, n=82) in two consecutive
1121 years (top: 2022 Spring harvest and bottom: 2023 Spring harvest). Trends (gray lines) and 95%
1122 confidence intervals (gray shades) come from linear regressions. Significance comes from linear-
1123 mixed kinship models of fitness (seed count for 2022 and inflorescence mass for 2023) in
1124 response to mean first day of flowering (fl), site, and their interaction (int): *p=0.006,
1125 **p<0.0006, ***p<0.00001. In all panels colors are K=4 ancestral clusters.
1126


a
chr. 1: 71007448 bp
MAF = 0.23
native
A
G
invaded
A
G
MAF = 0.09
b
FDR = 0.03
MAF = 0.91
−log10(p)
Chromosome 1 coordinates (Mb)
c
days to flower
max VPD (kPa)
***
native
G: 5
A: 27
G: 22
A: 77
invaded
G: 10
A: 128
G: 17
A: 176
d
chr. 1: 236616590 bp
MAF = 0.38
native
C
T
invaded
C
T
MAF = 0.06
e
FDR = 0.03
MAF = 0.82
−log10(p)
Chromosome 1 coordinates (Mb)
f
T:18
C: 14
T: 40
C: 64
T: 11
C: 128
T: 12
C: 182
GWAS nucleotide: n individuals

1137 **Fig. 4: Environmental trends of two flowering time QTL are mirrored between native and**
1138 **invasive genotypes.** (**a** and **d**) Geographic distribution of QTL SNP alleles in the native (top)
1139 and invaded (bottom) range; crosses represent the reference/major allele, and open circles the
1140 alternate/minor allele. (**b** and **e**) Zoomed-in Manhattan plots showing Wald-test *p*-values (plotted
1141 as $-\log_{10}$) from GWAS and genomic location of top SNP (marked in green), with respective
1142 false-discovery-rate (FDR) and minor allele frequency (MAF). (**c** and **f**) Phenotypic (boxplots to
1143 the left) and environmental variation (boxplots to the right) of flowering time QTL SNP alleles
1144 identified with GWAS. Boxplots indicate median (middle line), 25th, 75th percentile (box), and
1145 whiskers cover the data extent. ***$p<0.0008$ from two-tailed t-tests, but kinship linear-mixed
1146 models showed no significant differences. Max VPD: Maximum vapor pressure deficit in kPa.
1147 A, G, T, C on maps and x-axis of boxplots indicate nucleotides.
1148
1149
1150


a
Great Basin
genetic distance (offset)
350
550
b
frequency
WNA
ENA
mean genetic distance (offset)
c
Great Basin genetic distance (offset)
t-test p = 0.006, permutation p = 0.01
low
high
Great Basin cheatgrass abundance

1165 **Fig. 5: Genomic predictions of strong local adaptation occur in regions where cheatgrass is**
1166 **most dominant.** (**a**) Geographic distribution of the genomic offset estimated for each invasive
1167 genotype (n=194). The genomic offset or maladaptation is the genetic distance between observed
1168 invasive genotypes and the genotype-environment predictions in the invaded range based on the
1169 native range genotype-environment association. (**b**) Histograms of the mean genetic distance
1170 (offset) of 1000 null permutations in western North America (WNA, n=127) and eastern North
1171 America (ENA, n=67), relative to their estimated mean genetic distance (red lines). (**c**) Within
1172 the Great Basin (polygon in a, n=55), the mean genetic distance (offset) is significantly lower in
1173 areas where cheatgrass occurs in high (*i.e.*, representing >15% vegetation cover) vs. low
1174 abundance. Boxplots indicate median (middle line), 25th, 75th percentile (box), and whiskers
1175 cover the data extent.