

**Local to continental-scale variation in fitness and heritability in common bean
(*Phaseolus vulgaris*)**

Alice H. MacQueen¹, Colin K. Khoury^{2,3}, Phil Miklas⁴, Phillip E. McClean⁵, Juan M. Osorno⁵, Bryan C. Runck⁶, Jeffrey W. White⁷, Michael Kantar⁸, Patrick M. Ewing^{9*}

¹Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA

⁴International Center for Tropical Agriculture (CIAT), Cali, Colombia

⁵San Diego Botanic Garden, Encinitas, CA, USA.

⁴USDA-ARS, Grain Legume Genetics Physiology Research Unit, Prosser, WA, USA

⁵Department of Plant Sciences, North Dakota State University, PO Box 6050, Fargo, ND 58108, USA

⁶GEMS Informatics Center, University of Minnesota, Minneapolis, MN, USA

⁷Department of Agricultural and Biological Engineering, The University of Florida, Gainesville, FL, USA

⁸Department of Tropical Plant and Soil Science, University of Hawaii at Manoa, Honolulu, HI, United States of America

⁹USDA-ARS, North Central Agricultural Research Laboratory, Brookings, SD, USA

* Corresponding author

ABSTRACT

Selection during plant domestication and improvement often decreases genetic variation, including variation that confers adaptation to local conditions. We report spatial and temporal variation in fitness (seed yield), local adaptation, and segregating genetic variation within three races of common bean (*Phaseolus vulgaris* L.) with differing domestication histories and genetic diversities. Three-hundred and twenty-seven common bean genotypes had seed yield measured at subsets of 70 sites across North America between 1981–2015, as part of the Cooperative Dry Bean Nursery (CDBN). We estimated local adaptation using the metric home field advantage (HFA) and segregating genetic variation using heritability. The Durango and Mesoamerican races (Middle American genepool) had higher-than-expected ($p = 0.002$) HFA, equal to up to 34 years of average yield gains. Surprisingly, Nueva Granada (Andean genepool) and Durango yields became more heritable across the study period ($p < 0.001$), while Mesoamerican heritability decreased ($p < 0.001$). Both metrics detected diversity loss corresponding to the timings of major historical gene introgressions. Local adaptation remains an agronomically important phenomenon within some common bean races. Common bean adaptation to new conditions will be most rapid at locations with large local adaptation benefits and high heritability.

Key Words: Cooperative Dry Bean Nursery (CDBN), domestication, home field advantage, local adaptation, *Phaseolus vulgaris* (common bean)

1. INTRODUCTION

Genetic variation in phenotypes is the basis for effective natural and artificial selection. Through selection, crops become more productive and better adapted to new and changing environments. However, the domestication of crops consistently corresponds with a loss of genetic variation in “improved” or commercial genomes versus domesticated “landrace” or non-commercial genomes, and in domesticated genomes versus the genomes of wild relatives (Miller and Gross 2011; Moyer et al 2018). Domestication typically involves a genetic bottleneck followed by a long period of weak or variable selection. During the process of crop improvement, intense selection over short time frames is coupled with a reduction in effective population size and limited recombination, and often followed by long-distance migration events and rapid population expansion. To date, losses of genetic diversity between wild and improved populations range from 5 to 77% (Gray et al 2009, Lin et al 2014), and the only reported case of a gain in genetic diversity is in the Andean domestication event in common bean (*Phaseolus vulgaris* L.) (Schmutz et al 2014; Trucci et al, 2021).

Common bean is the most important food legume for direct human consumption. Common bean is currently grown on over 35 million ha and in nearly every country that records agricultural statistics (FAOSTAT 2015). This area encompasses across an enormous range of latitudes (from 52°N-32°S), altitudes (50-3000 m), growth temperatures (14-26 C), and average yearly precipitation (400-1600 mm) and precipitation regimes (Singh 1989; FAOSTAT 2015). In contrast, the habitat niche of wild common bean is narrow (Gepts 2009). The wild ancestors of common bean originated in central America (Bitocchi et al 2012) and were dispersed to the Andes ~165,000 years ago with a strong genetic bottleneck that lasted ~76,000 years (Schmutz et al 2014). From these two wild gene pools, common bean was domesticated at least twice, once in Mesoamerica (the ‘Middle American’ genepool) and once in the Andes (Gepts, et al. 1986; Schmutz, et al. 2014). The domestication

bottleneck was stronger in the Middle American than in the Andean genepool, probably because loss of diversity occurred in the Andes before domestication (Gepts et al., 1986; Bitocchi et al., 2012; Bellucci et al., 2014; Schmutz et al., 2014). Common bean was widely cultivated in Mexico and North America pre-European contact (Kaplan 1965), and New World settlers cultivated common bean from landraces maintained by Native Americans, primarily from the Middle American genepool, and later from European re-introductions, primarily from the Andean genepool (Vandemark et al 2014). These complex histories led to differences in genetic diversity, yield potential, and selection intensity among different races within these genepools. The first formal North American improvement efforts were initiated in the late 1800s and early 1900s by the United States state and federal governments (Vandemark et al 2014); since the 1950s, breeders have assessed comparative performance in part via the Cooperative Dry Bean Nursery (CDBN), the largest multi-environment trial of common bean in the United States and Canada (Myers 1988; Singh 2000). CDBN cooperators develop germplasm in their home regions, and are most interested in strong regional performance, typically releasing varieties with high seed yield that meet commercial expectations for seed size, shape, and color. However, they also test promising material for broad adaptation by measuring seed yield and other phenotypic traits across all participating CDBN field locations. The impacts of selection at both regional and continental scales on the genetic variation in common bean have not been measured; however, selection across heterogeneous environments is theorized to have major effects on population fitness and the suitability of different adaptive strategies (Levins, 1962; Botero et al 2015).

The CDBN dataset offers an exciting opportunity to examine the genetic component of fitness in common bean in three races within two gene pools which differ in their genetic diversity, and to determine the impact of this genetic variation on fitness and heritabilities at different spatial scales.

82 CDBN entries are advanced breeding materials without ‘locations of origin’ in the sense of their wild
83 progenitors. However, progenitors of these entries were grown in North America in pre-Columbian
84 times, and these specific entries were developed for regional production by breeders at many
85 locations that took part in the CDBN trials; thus, local adaptation of CDBN entries to specific
86 North American regions may have arisen from long-term selection of improved material within
87 those regions. If present, local adaptation might be leveraged in breeding programs that have
88 focused on broad-scale adaptation to achieve fitness gains (Ewing et al., 2019). For individual-based
89 selection, breeding programs often measure the heritability of traits of interest, the proportion of
90 phenotypic variation that is genetically determined. High heritability is essential for efficient
91 selection, but environmental conditions also influence the heritability on which any response to
92 selection depends (Hoffman, Hercus 2000; Hoffman, Merila 1999). Heritability may also vary
93 temporally as the array of genotypes measured typically changes during the breeding process
94 (Feldman, Lewontin 1975). Understanding spatial and temporal patterns of seed yield heritability
95 and the interaction between heritability and the portion of seed yield variation attributable to local
96 adaptation could provide fundamental insights into the adaptation and domestication process of *P.*
97 *vulgaris* (Table 1; Papa & Gepts 2003; Papa et al., 2007; Eckert et al 2010; Kraft et al 2014; Rodriguez
98 et al 2015). These patterns and interactions may differ among races due to their unique
99 domestication histories and genetic variabilities, with implications for the potential of bean races to
100 continue to adapt to changing conditions.

Table 1

Implications of home field advantage and heritability for breeding and adaptation. Combining agroecological ecoregion information and heritability of specific traits may help improve selection efficiency while providing insights into processes driving past selection.

		Heritability	
		Lower	Higher
Home Field Advantage	Larger	<i>Processes:</i> High environmental variation among locations, low genetic variation potentially enriched in locally important alleles. <i>Implications:</i> Inefficient selection on individuals for phenotypic improvement; testing at these sites may reveal conditionally beneficial alleles.	<i>Processes:</i> High environmental variation among locations, high genetic variation potentially enriched in locally important alleles. <i>Implications:</i> Efficient identification and selection of specialists that may contain large-effect candidate loci for introgression.
	Smaller	<i>Processes:</i> Low environmental variation among locations, low genetic variation potentially enriched in broadly important alleles. <i>Implications:</i> Inefficient selection on individual entries for phenotypic improvement, so family based methods are necessary.	<i>Processes:</i> Low environmental variation among locations, high genetic variation potentially enriched in broadly important alleles. <i>Implications:</i> Efficient selection on individual entries for phenotypic improvement.

102

103 We re-interpret the CDBN trials as a reciprocal transplant experiment to estimate local
 104 adaptation and yield heritability among common bean races, experimental locations, and across
 105 years. We first introduce a “home field advantage” (HFA) component for each CDBN entry into a
 106 traditional partitioning of seed yield variation and explore patterns of HFA across three races within
 107 the two genepools. Briefly, HFA is the fitness gain an entry realizes by growing in its home location,
 108 which is empirically defined as the location of highest fitness relative to other entries (Ewing et al,
 109 2019). We expected that HFA would explain a significant amount of yield variation within each bean
 110 race. We then examined overall variation in yield heritability across years and locations in the CDBN

and across races and genepools. We expected to observe substantial differences in heritability across space due to environmental differences (Wilson et al 2006; Finlay and Wilkinson, 1963). We also expected to observe decreases in heritability over time in the Middle American genepool as genetic sources of yield variation were eliminated, and potential increases in heritability over time in the Andean genepool, given its increase in genetic diversity from wild to domesticated gene pools. Finally, through identifying sites with high HFA and yield heritability, we discuss potential strategies for selection efforts to further incorporate local adaptation.

2. MATERIALS AND METHODS

2.1 Data Acquisition

Common bean seed yield (hereafter: ‘yield’) data were retrieved from MacQueen et al. 2020; briefly, yield data from digitized reports from 1981 to 2015 were obtained, which included data from 71 unique locations, and 327 entries (including released cultivars and unreleased advanced breeding lines) with 11 common check entries (Figure 1). CDBN entries include released cultivars and unreleased breeding lines at advanced breeding stages from at least 13 diverse North American market classes of common bean. The majority of these entries are members of one of three of the seven races of common bean; two of these races, Durango and Mesoamerican, arose from the Middle American domestication event and are more genetically diverse than Nueva Granada, which arose from the Andean domestication events (Mamidi, et al. 2011). 13,989 yield data points were used in this analysis.

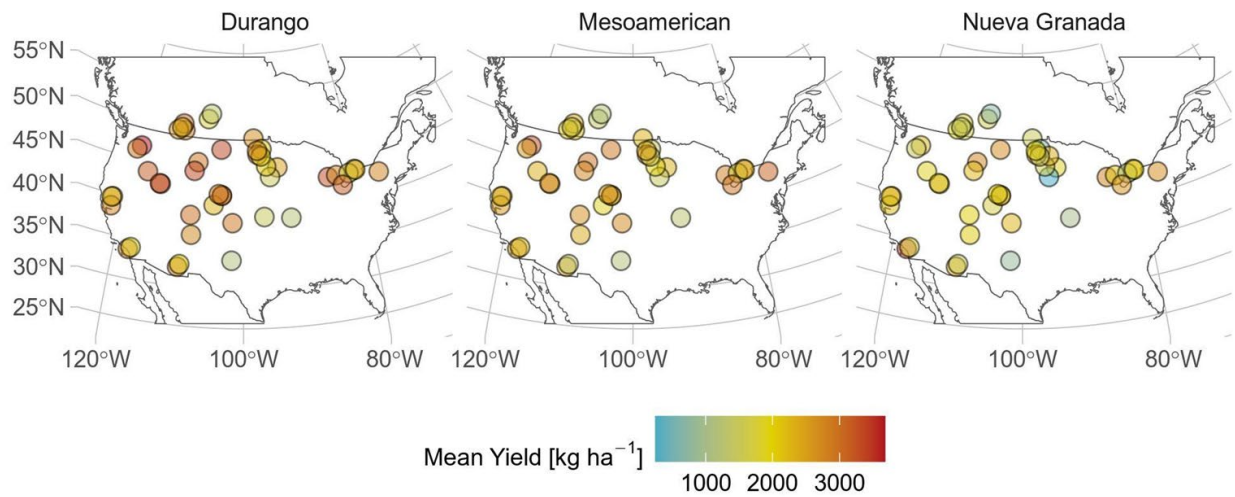


Figure 1: Locations and mean yields at sites within the CDBN variety trials. The analysis was restricted to continental North American sites used for at least three years.

The CDBN selection strategy for lines entered likely impacts partitioning of yield variance in several ways. This strategy was developed to test dry bean material for broad adaptation across CDBN locations. Within years, CDBN locations grew an identical complement of entries, with only rare entries missing yield phenotypes at single locations. The entries shift gradually through the years, with each entry typically present in the CDBN for three consecutive years. This shift is a direct result of the improvement strategy employed by network cooperators over the trial years and captures some of the regional breeding advances produced by additional non-CDBN field trials. This strategy confounds partitioning of yield variation into genetic and environmental components across years, but not across locations. Specifically, comparisons of heritability for different locations within years are not confounded by the presence of different genotypes. Heritabilities from different years within locations reflect both environmental differences across years and a change in the complement of genotypes used to compute the heritability.

2.2 Yield and home field advantage expectations

We first assessed overall fitness and fitness gains across the study period within three groups of CDBN entries: entries from the Nueva Granada race from the Andean genepool, and entries from two genetically distinct races within the Middle American genepool, the Durango and the Mesoamerican race using linear modeling.

We then quantified HFA. Home field advantage is the fitness benefit an entry receives by growing in its “home” location, after accounting for inherent site and genetic potential, and therefore is an entry-specific outcome of genotype-by-environment interactions (Blanquart et al, 2013). We estimated HFA following a modified version of Ewing et al. (2019). Most importantly, we included the additional step of accounting for entry genetic relationships to eliminate effects of double counting similar CDBN entries.

To estimate HFA, we first filtered out sites with fewer than three years of data. Next, we assigned each entry an empirical “home” site, the site where an entry performed best relative to other varieties. We calculated relative yields within each site-year by centering (mean = 0) and scaling (standard deviation = 1). Then, for each variety, we calculated the best linear unbiased predictor (BLUP) of relative yield at each site by using site as a random intercept in *lme4* 1.1 (Bates et al, 2015). The home site for each CDBN entry was the site with the largest BLUP.

With home site identified, we then tested whether it explained variation in CDBN yields using model comparison, including in conjunction with genetic relatedness, by comparing formulations of the ordinary least squares model:

$$y = X\beta + \epsilon$$

Where \mathbf{y} contains the n yields of each entry i at location j and year k , $\boldsymbol{\beta}$ contains coefficients of p predictors in the n by p model matrix \mathbf{X} , and $\boldsymbol{\varepsilon}$ is the associated error term. In the base model, \mathbf{X} included site, year, site-year, and entry identity. Model 2 added a predictor for HFA, a binary indicator of whether a site was that variety's home. Model 3 added terms for genetic relatedness to the base model. Genetic relatedness was determined using the *snp_autoSVD* function of the *bigsnpr* R package (Prive et al 2017), using 1,221,540 SNPs genotyped previously using genotyping by sequencing (MacQueen et al 2020); we used the first three principal components of the singular value decompositions of linkage-adjusted SNP matrices (Figure S1). Model 4 added to the base model both HFA and genetic relatedness terms, including an interaction. In models with an HFA term, the coefficient β_{HFA} is the average home field advantage across entries.

We compared these models using Akaike's information criterion (AIC); Model 2 was always the most parsimonious. We then calculated variance explained by each predictor in Model 2 using the *Anova* function in *car* 3.0 (Fox and Weisberg, 2019). We then used 999 permutations to test the significance of home field advantage (Ewing et al, 2019). After permuting yields within site-year, we re-assigned home site and re-calculated home field advantage as described above. These permuted HFAs were as expected based on variation in the dataset; we compared this with the observed HFA using a two-tailed test.

To observe trends in HFA across the study period, we modified Model 1 to add a year by home field interaction. The coefficients of these terms were the home field advantages within each year. Likewise, we estimate home field advantage conferred by each site, by adding a site by home field interaction to Model 1.

2.3 Spatial and temporal variation in heritability

We calculated narrow sense heritability for yield per-site on an entry-mean basis, by combining data from all breeding trials, then decomposing the variance components within each trial. Briefly, we built linear models using the R package *lme4* version 1.1 to calculate heritability within each site-year using the variance in components at each site with effects of entry, site, and year effects, as well as entry-by-site and entry-by-year interactions. Heritability was calculated as

$$h^2 = V_{\text{entries}} \left(\frac{1}{\text{rep} * \text{environment}} * MS_{\text{entries}} \right)^{-1}$$

Where V_{entries} the variance among the entries, environment is the number of locations in the study and MS_{entries} is the squared value of expected yields for the entries (Bernardo, 2020, page 163). To quantify the change in heritability over time, we used average heritability across a five-year sliding window. For example, heritability for 1983 was the average heritability of 1981-1985. Finally, to model the spatial variation of heritability across North America, we used ordinary kriging implemented in *fields* version 11.6 (Furrer et al., 2009). For spatial variation, heritability was the average heritability over the 35-year study period and across races.

2.4 Relationship between home field advantage, heritability and kinship

Finally, we identified locations suitable for breeding locally adapted crops. We defined such sites as both having a higher-than-median heritability and conferring higher-than-median home field advantages. We also quantified whether home site location had a genetic basis by correlating the spatial distance among each variety's home site with the genetic distance among varieties. Home distance was great circle distance using *sp* 1.3. Genetic distance was reconstructed from the first three principal components of the linkage-corrected kinship matrices. We tested the correlation using a Mantel test with 999 permutations in *vegan* 2.5 (Oksanen et al, 2019).

2.5 Statistical analysis

All analyses were performed independently for each race. Unless otherwise noted: statistics were performed in R 3.6.3 *base* and *stats* packages; plots were made with *ggplot2* 3.3.0 and annotated with regression formulae using *ggpmisc* 0.3.3 (R Core Team, 2020; Wickham, 2016; Aphalo, 2019); centering and scaling refer to a group mean of zero and standard deviation of one; errors are standard.

3. RESULTS

3.1 Home Field Advantage

Each dry bean race studied had a different yield potential ($F_{2,13170} = 608$; $p < 0.001$, Figure 2a; Table S1): By 2015, Nueva Granada entries were expected to yield $2370 \pm 40 \text{ kg ha}^{-1}$, Durango entries, $3060 \pm 30 \text{ kg ha}^{-1}$, and Mesoamerican entries, $2850 \pm 40 \text{ kg ha}^{-1}$. In agreement with previous studies (Vandemark et al 2014), yields of all races improved during the study period, albeit at different rates ($F_{2,13170} = 4.00$, $p = 0.02$). Nueva Granada yields grew by $16 \pm 2 \text{ kg ha}^{-1} \text{ yr}^{-1}$ ($p < 0.001$), and Durango yields grew by $14 \pm 1 \text{ kg ha}^{-1} \text{ yr}^{-1}$ ($p < 0.001$). Mesoamerican yields grew the fastest, by $20 \pm 2 \text{ kg ha}^{-1} \text{ yr}^{-1}$ ($p < 0.001$).

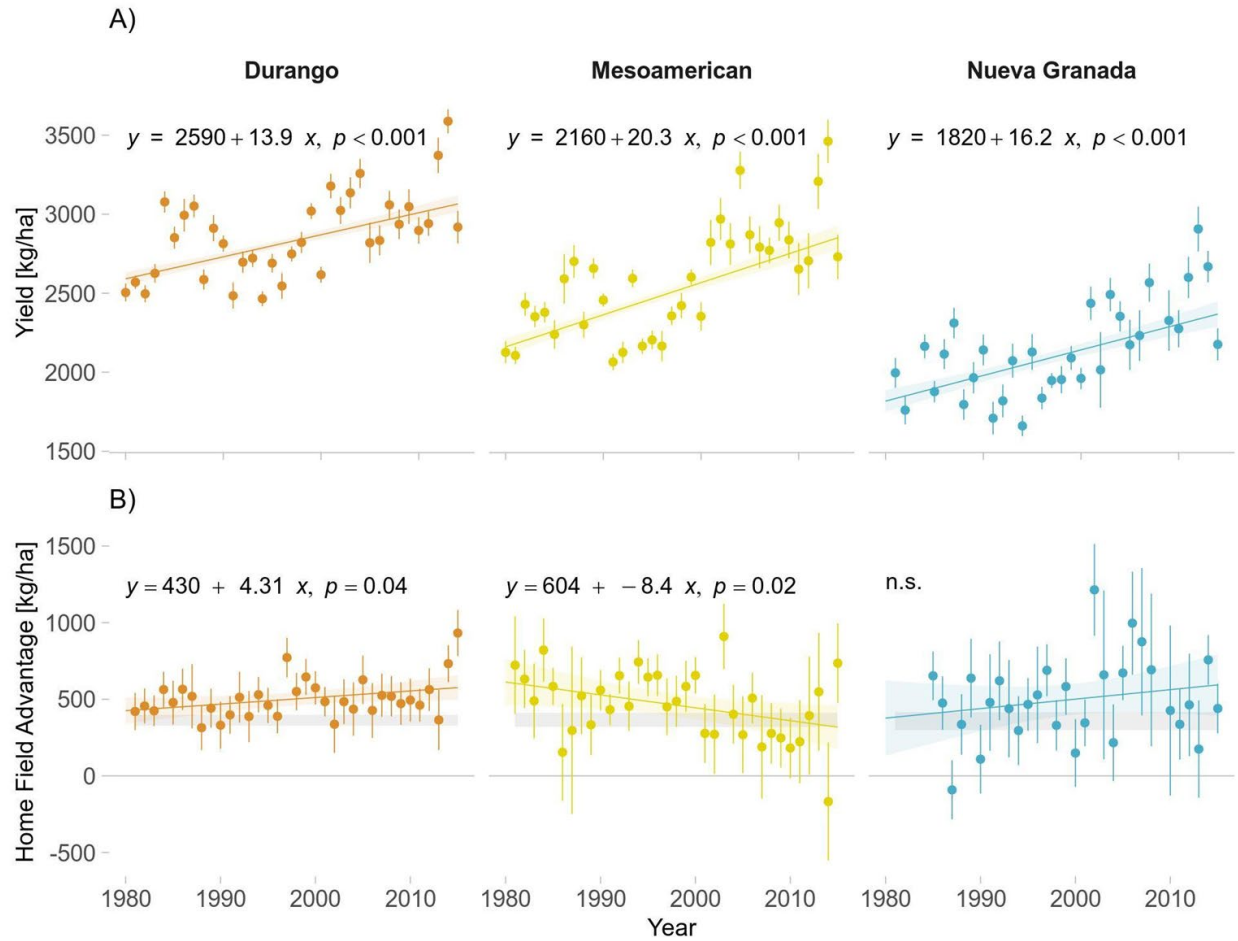


Figure 2: Trends in A) yield and B) home field advantage across the study period by race. In B), the gray bars indicate the 90% confidence interval of the expected home field advantage based on permutation within site-year. Error bars are standard errors. Shading around regression lines indicates 95% confidence intervals. n.s.: not significant at $p < 0.05$.

We partitioned yield variation into genetic and environmental components within the two gene pools and three races (Table 2). Emphasizing the importance of local conditions, the largest source of variability in yield across all races was a location-by-year interaction (44.9-48.7%), followed by an effect of location alone (29-33.1%). In contrast, year-to-year variation and genotype variation were small (2.4-3% and 3.2-4.1%). To test for a contribution of local adaptation to these fitness gains, we also assessed the variation in yield explained by each CDBN entry's home field advantage (HFA). In Nueva Granada entries, HFA accounted for 1.1% of yield variation, and reduced unexplained yield variation by 6.4% (Table 2). In Durango entries, HFA accounted for 1.3% of yield

variation and reduced residual variation by 8.3%. Similarly, HFA accounted for 1.3% of Mesoamerican yield variation, and reduced residual variation by 7.7%. Including principal components (PCs) as covariates for genetic relatedness did not reduce HFA; in fact, including genetic relatedness as a predictor of yield reduced model parsimony ($\Delta AIC = 339$ (Nueva Granada), 890 (Durango), and 608 (Mesoamerican); Table S2). These three PCs accounted for 39%, 44%, and 42% of genetic variance in the Nueva Granada, Durango, and Mesoamerican entries, respectively (Figure S1).

Table 2
Partitioning of CDBN Yield Variances

	Race		
	Durango	Mesoamerican	Nueva Granada
Home Site	1.3%	1.3%	1.1%
Residuals	14.1%	15.8%	15.6%
Genotype	3.3%	4.1%	3.2%
Site	33.1%	30.9%	29.0%
Year	2.4%	3.0%	2.5%
Site-Year	45.7%	44.9%	48.7%
Residual Reduction[†]	8.3%	7.7%	6.3%

[†] Proportion of residual variance explained by home site

We next determined if race-average HFA were larger than expected and whether these average HFA were agronomically significant. Nueva Granada entries had yield gains that were not significantly different from the expectation of 357 kg ha⁻¹ based on permutations of variation in Nueva Granada yields ($p = 0.21$; Figure 2b). In Durango entries, HFA conferred an average yield gain of 472 kg ha⁻¹, 109 kg ha⁻¹ larger than expected (90% confidence: 72-146 kg ha⁻¹; $p = 0.002$). The

Durango HFA was 15% of expected 2015 yields, equivalent to 34 years of yield gains for this race. In Mesoamerican entries, HFA conferred a yield gain of 453 kg ha⁻¹, 89 kg ha⁻¹ larger than expected (90% confidence: 44 -134 kg ha⁻¹; $p = 0.002$). The average Mesoamerican HFA was 16% of expected 2015 yields and equivalent to 22 years of yield gains for this race. Thus, some environmental specialization had been preserved in modern North American varieties deriving from the Mesoamerican race, but not the Nueva Granada race.

Though environmental specialization was present within two dry bean races, we found inconsistent evidence that dry bean breeders tried to leverage HFA, in that there were inconsistent, and only moderately significant changes in HFA among races and across time ($F_{2,94} = 4.82$; $p = 0.01$; Table S3). HFA increased slightly across the study period in Durango entries, by 4.3 ± 2.0 kg ha⁻¹ yr⁻¹ ($p = 0.04$; Figure 2b). In Mesoamerican entries, however, HFA decreased by 8.4 ± 3.5 kg ha⁻¹ yr⁻¹ ($p = 0.02$) and by the end of the study period, was no larger than expected based on permutation.

3.2 Spatial and Temporal Decay in Heritability

We next determined narrow-sense heritability in yield across 2170 location-year combinations in the CDBN, without accounting for HFA. Yield heritabilities were highest in the northwestern United States and southwestern Canada and decreased along a south-eastern gradient (Figure 3). Heritabilities were stable over large geographic ranges (~1013 km), although year-to-year fluctuations in the direction of the gradient in yield heritability were evident (Figure S2). More commonly used trial locations did not have more consistent heritabilities: in fact, there was a positive relationship between the number of trial years at a location and the standard deviation in heritabilities (yield stability) across trial years ($p = 0.014$, Figure S3A). Location-specific heritability was also unrelated to yield stability at the location ($p > 0.05$; Figure S3B). Heritability was higher in higher quality site-years ($p < 0.05$), with environmental quality measured as the average yield of

273 entries of a race in that site-year (Fig S3C; Finlay and Wilkinson, 1963). However, this trend was
274 confounded with values across time, which reflected the direct act of selection for improvement on
275 CDBN entries. Specifically, there were trends in yield heritability across years which varied by bean
276 race. As hypothesized, heritability decreased in Mesoamerican entries by 0.005 yr^{-1} ($p < 0.001$). In
277 contrast, Nueva Granada and Durango entries showed an increase in yield heritability over time,
278 with yield heritability increasing by 0.007 yr^{-1} and 0.004 yr^{-1} , respectively ($p < 0.001$). To test for a
279 genetic basis for environmental specialization, we determined whether more closely related CDBN
280 entries had spatially closer home sites. Entries in the Durango (Mantel's $r = 0.11$; $p = 0.001$) and
281 Mesoamerican (Mantel's $r = 0.068$; $p = 0.01$) races showed this relationship, while entries in the
282 Nueva Granada race did not ($p = 0.6$). Finally, we found no tradeoff between heritability and home
283 field advantage within races ($F_{2,97} = 1.27$; $p = 0.3$; Figure 4).

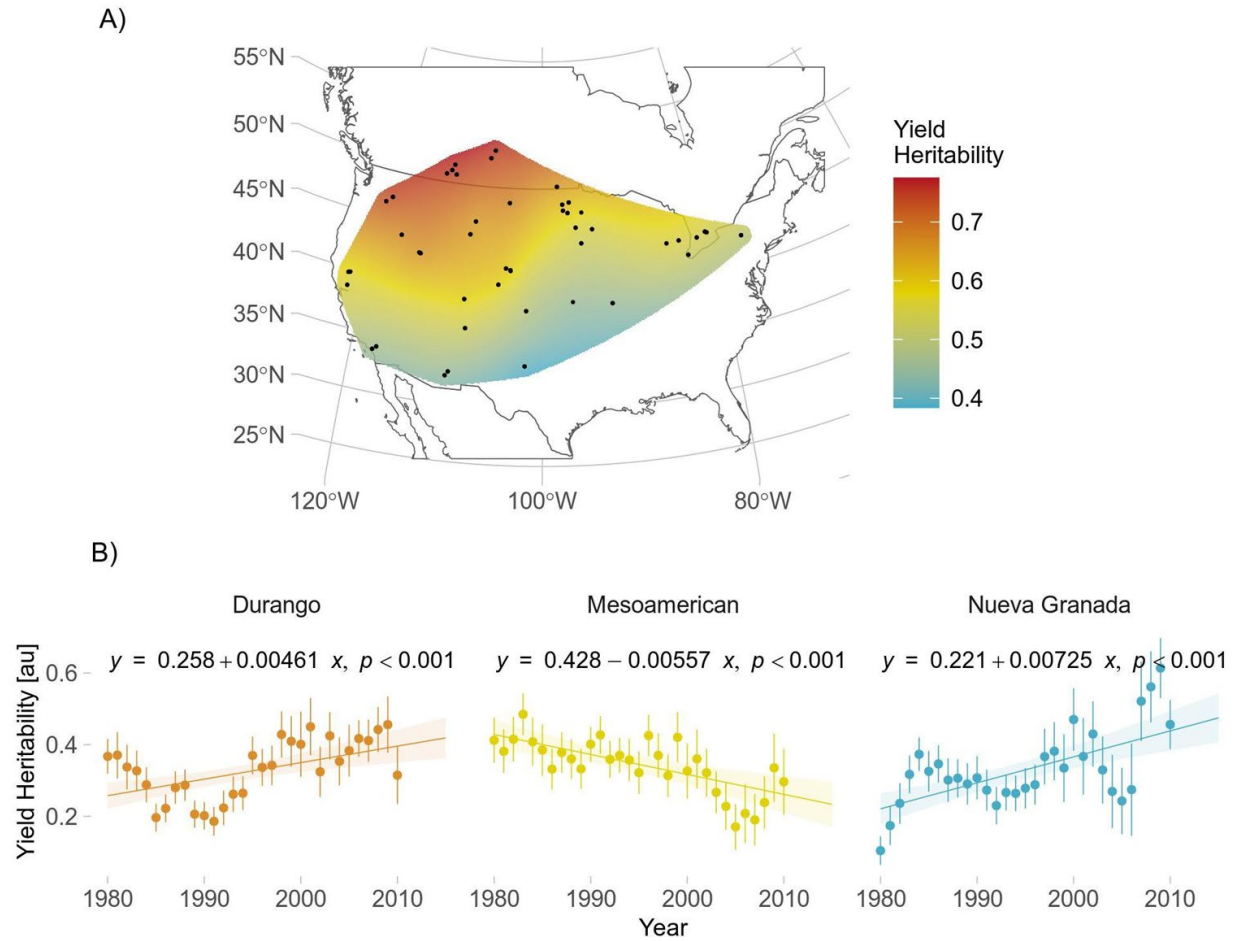


Figure 3: A) Ordinary kriging of yield heritability based on the average heritability over 30-years of trials at 70 field sites. B) Mean and standard deviation of site heritability for 70 field sites. Symbology is as in Figure 2.

4. DISCUSSION

4.1 Orthogonal Metrics to Interpret Past Selection and Guide Future Improvement

Modern breeding techniques have achieved remarkable success in increasing the fitness and phenotypic stability of cultivated plants through breeding programs that focus on broad adaptation at large spatial scales. Efforts such as the CDBN have selected for both broad and regional environmental adaptation; however, ecological theory suggests that across environments that vary spatially or temporally, the most fit population will be a mix of narrowly adapted specialists (Levins,

1962). Ascertaining the presence and amount of environmental specialization remaining in broadly adapted crop species may provide an opportunity to further improve crop fitness by leveraging local adaptation. In other words, to continue improving common bean, we may need to select for locally important alleles, (i.e., conditionally neutral alleles that affect performance in individual environments without affecting performance in alternate environments). Yet unknown is whether these locally adapted alleles or gene complexes are still segregating within advanced common bean breeding material, or whether they have been removed as part of the genetic bottleneck during the improvement of these races. Here, we measure HFA as a proxy for yield variation that may be attributed to local adaptation. In Durango and Mesoamerican entries, local adaptation conferred agronomically significant yield gains that were equivalent to decades worth of breeders' yield improvements.

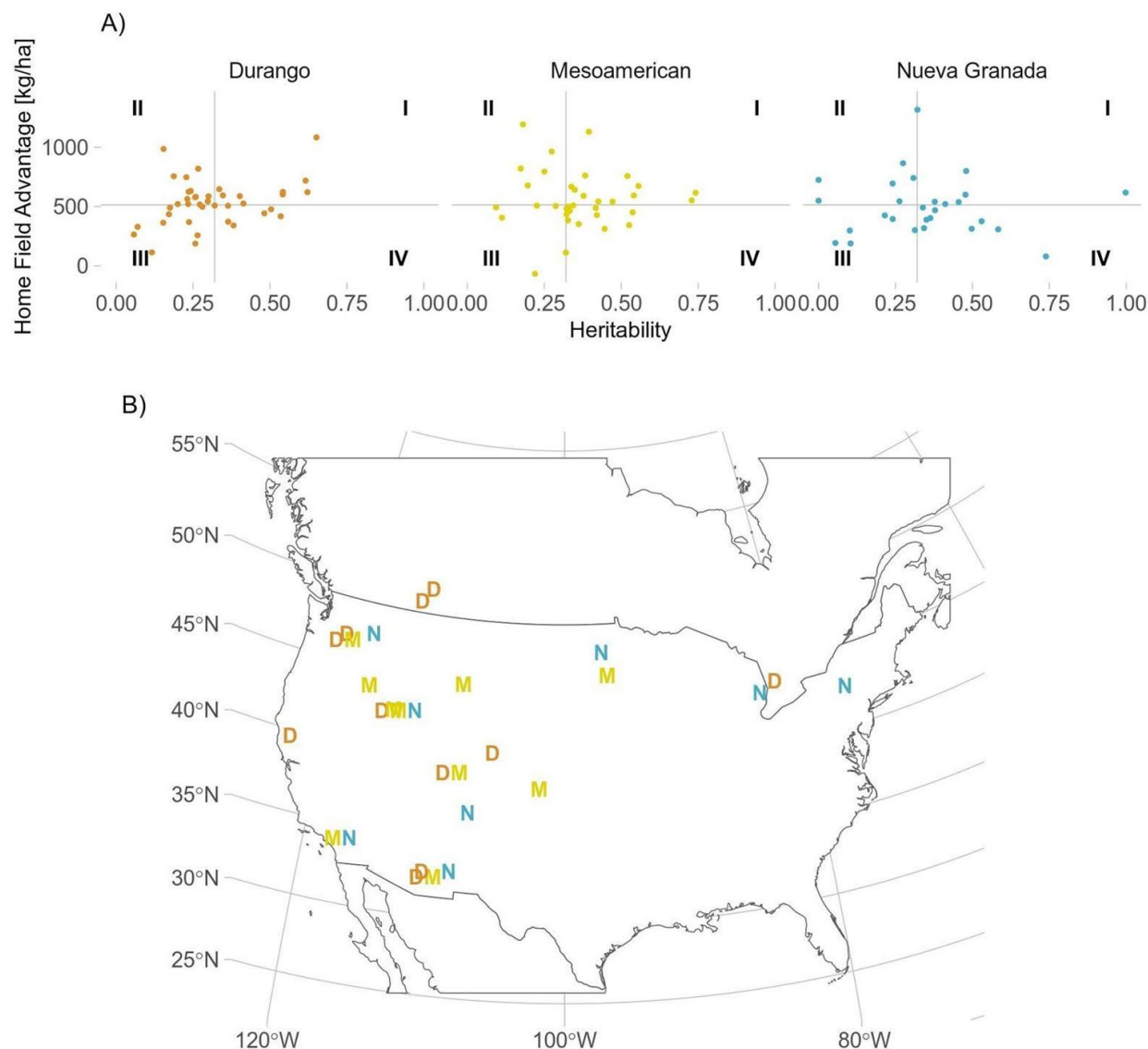


Figure 4: A) Relationship between yield heritability and home field advantage at each site, by race. Lines indicate the median value for heritability and home field advantage across all sites. B) Top locations (quadrant I) for selecting for home field advantage. D – Durango; M – Mesoamerican; N – Nueva Granada races. Figure S3 identifies locations with CDBN abbreviations.

We correlate these HFAs with the amount of variance segregating for entry-mean heritability (Figure 4). We envision that the interplay between HFA, an entry-specific quantity, and heritability, a function of all genotypes grown at a location, could impact the kinds of selection practiced in the future at specific locations of the CDBN (Table 1). When trait heritability is high and HFA are not present for any CDBN entries (Table 1; Figure 4, quadrant IV), the current breeding approaches that

select upon individual entries and families are appropriate. When trait heritability is high and HFA are present for CDBN entries (Table 1; Figure 4, quadrant I), breeders have an additional opportunity to screen and select for locally important alleles which may confer benefits at that location without accruing trait costs at other locations. These alleles could be introgressed more broadly into other material to improve performance in these environments. When trait heritability is low and HFA is present for CDBN entries (Table 1; Figure 4, quadrant II), individual-based selection is inefficient; however, entries at these locations could still be screened for locally important alleles which may improve performance in these environments. When neither trait heritability nor HFA are present (Table 1; Figure 4, quadrant III), neither traditional breeding approaches nor screening for locally important alleles are likely to aid further common bean improvement.

Given the different evolutionary histories of both gene pools pre- and post-domestication, we expected distinct responses to selection in the Andean and Middle American entries over the course of the CDBN. The low-yielding, low genetic diversity Nueva Granada entries from the Andean gene pool experienced modest improvements in yield across the study period and showed little evidence of local adaptation to specific regions within North America. Nonetheless, there was strong evidence for increasing adaptation at a continental scale, in that yield heritability increased in this race over the 35-year study period. These increases in yield heritability over time occurred without an increase in HFA. We thus suggest that Andean CDBN entries still lack genetic variation conferring adaptation to North American growing conditions. If this is true, then further introgression from the Middle American gene pool or from other species may offer opportunities for continued yield improvement and local adaptation in this gene pool. Alternatively, Andean entries also require distinct management strategies and more inputs to maximize their yields; many CDBN

locations may have managed for Middle American entries rather than Andean entries, which would limit our ability to detect an HFA in Andean material.

The high-yielding, high genetic diversity Durango race also showed modest improvements in yield across the study period, which coincided with a modest increase in local adaptation and increases in adaptation at a continental scale over time. This increase in HFA and in yield heritability suggests increasing adaptation to local and continental conditions in this race over the course of improvement in the CDBN. We observe a drop in Durango yields and yield heritability from the line-of-best-fit (Figure 2a, Figure 3b) beginning in the mid-1990s. This immediately follows the introgression of rust resistance gene *Ur-3* into the Durango race from a Mesoamerican source, which began in 1988 and was widely deployed in the mid-1990s (Hurtado-Gonzales et al 2017; MacQueen et al 2020). We speculate that the broad introduction of rust resistance loci into the Durango race in this time period resulted in a temporary (~6 year) reduction in segregating variation for yield within this race.

Finally, the moderately yielding Mesoamerican race showed the largest improvements in yield across the study period, at the expense of local adaptation and heritability of yield in more modern entries. Both the decrease in heritability and HFA over time suggest a reduction in genetic diversity in this race during the course of improvement. We observed a sustained drop in yields relative to the line-of-best-fit (Figure 2a) in the Mesoamerican race in the 1990s. This observation was also made in previous work which saw particularly low biomass and seed yields for Mesoamerican varieties introduced in the 1990s that had upright determinate architecture (MacQueen et al 2020). After the 1990s, very few Mesoamerican CDBN entries have this architecture, as CDBN breeders had adopted upright indeterminate architecture for newer CDBN entries in the Mesoamerican race (Kelly 2001; Soltani et al 2016). This adoption highlights the complexity of defining fitness in

agricultural systems: ‘fitness’ includes both overall seed yield, and numerous anthropocentric values such as harvestability and commercial value that may override yield considerations. In the mid-2000s, we also observed a sustained drop in yield heritability relative to the line-of-best-fit. Major alleles for both lodging and plant height stopped segregating within the Mesoamerican race in this time period; we speculate that loss of variation linked to these alleles could have caused this temporary drop in heritability of yield. These major changes to entries in the Mesoamerican race may also have caused the negative trend in HFA magnitude present in Mesoamerican entries, as locally adaptive alleles may have been lost in genomic regions adjacent to architecture, height, and lodging genes. In general, the introgression of major resistance or architecture alleles into populations is equivalent to an artificially induced selective sweep. When this occurs, we expect to see considerable linkage drag and the loss of locally adapted genetic variants. This signal was observed in the CDBN, where following introgression, many years were required to regain local adaptation (HFA) after using a common parent to introduce novel, important traits.

4.2 HFA across species: comparing apples and oranges

Previous work quantifying HFA in maize found that a large proportion of yield variation was attributable to local adaptation (Ewing et al 2019) but was unable to determine if accounting for genetic relatedness was an effective substitute for HFA. We found that HFA was robust to genetic and familial relationships within CDBN material, supporting the utility of the HFA metric. We were also able to compare HFA across these species. When compared to the annual breeding gains observed in each trial, common bean HFA was four-fold larger than maize HFA. In absolute terms, maize HFA was $1.0 \text{ Mg ha}^{-1}\text{year}^{-1}$, equivalent to 4.5 years of observed yield gains; in addition, the HFA observed was smaller than expected (Ewing et al., 2019). In contrast, in this study, HFA was $450\text{-}470 \text{ kg ha}^{-1} \text{ yr}^{-1}$, equivalent to 22 and 34 years of annual breeding gains for Mesoamerican and

Durango races, respectively; these HFA were larger than expected. There are several possible explanations for the difference in HFA between species. The current study encompasses yield trials and improvement efforts over a much broader spatial scale and range of environments: thousands of kilometers, versus a few hundred in Ewing et al (2019). Selection in common bean has occurred in many regional breeding programs with disparate goals, which commonly develop specific sub-races (market classes) (Vandermark et al, 2014); this may have enhanced regional adaptation in common bean. Yet at the same time, several of these bean programs also employ shuttle breeding, which commonly leads to the development of more broadly adapted germplasm (Ortiz et al, 2007). The stable or decreasing HFA in Nueva Granada and Mesoamerican CDBN entries suggests that the presence of regional adaptation is weakening, in line with CDBN goals. Irrespective of the underlying cause of the size difference in HFA between dry bean and maize, in both species and breeding systems (i.e. self-pollinated dicot, hybrid monocot), local adaptation is both valuable to future breeding efforts and decreasing due to improvement efforts.

5. CONCLUSION

The interplay between local adaptation and total genetic variation may impact plant adaptation to changing conditions. Local adaptation was worth decades of yield gains in gene pools from both domestication events over the 35 years of common bean improvement represented by the CDBN but was decreasing or absent in two of three gene pools. Surprisingly, we found that yield heritability increased over time in races from both domestication events, indicating that genetic diversity in races of both gene pools may have increased, perhaps due to introgression from the other gene pool or from other species. The presence of yield variation attributable to local adaptation underscores the possibility of selecting at particular CDBN locations for alleles that confer adaptation to those environments. Resilient food systems of the future will take advantage of all potential avenues of

improvement, including local adaptation, to sustainably produce food adapted to novel and changing production environments.

6. ACKNOWLEDGEMENTS

The authors thank the many individuals who helped compile this dataset and provide feedback on drafts, as well as their funding sources. This research was partially funded by support from the National Science Foundation, Plant Genomes Research Program, Grant IOS-1612262 to AHM.

7. DATA AVAILABILITY

Data and accompanying analyses are available at <https://github.com/alice-MacQueen/cdbn-home-away>.

8. AUTHOR CONTRIBUTIONS

All authors contributed conceptualization and manuscript review. AHM, MK, and PME designed and performed analyses, and wrote the manuscript with contributions from all authors. PM, PEM, and JMO executed field trials. JWW and JMO compiled the phenotypic data from the annual CDBN reports.

9. REFERENCES

1. Aphalo, P. J. (2018). Ggpmisc: Miscellaneous Extensions to ‘ggplot2’. R package version 0.3, 3.
2. Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4*: Linear Mixed-Effects Models Using Eigen and S4.
3. Beebe, S. (2012). Common bean breeding in the tropics. *Plant breeding reviews* (pp. 357–426). Hoboken, NJ, USA: John Wiley & Sons Inc.
4. Bernardo R (2020) Breeding for quantitative traits in plants, 3rd edn. Stemma Press, Woodbury, Minnesota

- 431 5. Bellucci, Elisa, Elena Bitocchi, Alberto Ferrarini, Andrea Benazzo, Eleonora Biagetti, Sebastian
432 Klie, Andrea Minio, et al. 2014. Decreased Nucleotide and Expression Diversity and Modified
433 Coexpression Patterns Characterize Domestication in the Common Bean. *The Plant Cell* 26 (5):
434 1901–12. <https://doi.org/10.1105/tpc.114.124040>.
- 435 6. Bitocchi, Elena, Elisa Bellucci, Alessandro Giardini, Domenico Rau, Monica Rodriguez,
436 Eleonora Biagetti, Rodolfo Santilocchi, et al. 2013. Molecular Analysis of the Parallel
437 Domestication of the Common Bean (*Phaseolus Vulgaris*) in Mesoamerica and the Andes. *New*
438 *Phytologist* 197 (1): 300–313.
- 439 7. Bitocchi, Elena, Laura Nanni, Elisa Bellucci, Monica Rossi, Alessandro Giardini, Pierluigi
440 Spagnoletti Zeuli, Giuseppina Logozzo, et al. 2012. Mesoamerican Origin of the Common Bean
441 (*Phaseolus Vulgaris* L.) Is Revealed by Sequence Data. *Proceedings of the National Academy of Sciences*
442 109 (14): E788–96. <https://doi.org/10.1073/pnas.1108973109>.
- 443 8. Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., & Pebesma, E. J. 2013. Applied spatial data
444 analysis with R (Vol. 2). New York: Spring
- 445 9. Botero, C. A., Weissing, F. J., Wright, J., & Rubenstein, D. R. 2015. Evolutionary tipping points
446 in the capacity to adapt to environmental change. *Proceedings of the National Academy of Sciences*,
447 112(1), 184-189.
- 448 10. Eckert, C. G., Kalisz, S., Geber, M. A., Sargent, R., Elle, E., Cheptou, P. O., ... & Winn, A. A.
449 (2010). Plant mating systems in a changing world. *Trends in Ecology & Evolution*, 25(1), 35-43.
- 450 11. Ewing, P. M., Runck, B. C., Kono, T. Y., Kantar, M. B. (2019). The home field advantage of
451 modern plant breeding. *PloS one*, 14(12), e0227079.
- 452 12. FAOSTAT, 2015 Food and Agriculture Organization of the United Nations, pp.,
453 <http://faostat3.fao.org/>
- 454 13. Farid, M., Earl, H. J., Pauls, K. P., & Navabi, A. (2017). Response to selection for improved
455 nitrogen fixation in common bean (*Phaseolus vulgaris* L.). *Euphytica*, 213(4), 99.
- 456 14. Feldman, M. W., and R. C. Lewontin. 1975. The Heritability Hang-Up. *Science* 190 (4220): 1163–
457 68.
- 458 15. Finlay KW, Wilkinson GN. The analysis of adaptation in a plant-breeding programme.
459 *Australian journal of agricultural research*. 1963;14(6):742–54.
- 460 16. Fischer, R. A. (2015). Definitions and determination of crop yield, yield gaps, and of rates of
461 change. *Field Crops Research*, 182, 9-18.
- 462 17. Fox, John, & Weisberg, Sanford. (2019). An {R} Companion to Applied Regression, Third
463 Edition. Thousand Oaks CA: Sage. URL:
464 <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>

18. Furrer, R., Nychka, D., Sain, S., & Nychka, M. D. (2009). Package ‘fields’. R Foundation for Statistical Computing, Vienna, Austria. <http://www.idg.pl/mirrors/CRAN/web/packages/fields/fields.pdf> (last accessed 22 December 2012).
19. Gepts, P., T. C. Osborn, K. Rashka, and F. A. Bliss. 1986. Phaseolin-Protein Variability in Wild Forms and Landraces of the Common Bean (*Phaseolus Vulgaris*): Evidence for Multiple Centers of Domestication. *Economic Botany* 40 (4): 451–68. <https://doi.org/10.1007/BF02859659>.
20. Gray, Melissa M., Julie M. Granka, Carlos D. Bustamante, Nathan B. Sutter, Adam R. Boyko, Lan Zhu, Elaine A. Ostrander, and Robert K. Wayne. 2009. Linkage Disequilibrium and Demographic History of Wild and Domestic Canids. *Genetics* 181 (4): 1493–1505. <https://doi.org/10.1534/genetics.108.098830>.
21. Hoffmann, Ary A., and Miriam J. Hercus. 2000. Environmental Stress as an Evolutionary Force. *BioScience* 50 (3): 217–26. [https://doi.org/10.1641/0006-3568\(2000\)050\[0217:ESAAEF\]2.3.CO;2](https://doi.org/10.1641/0006-3568(2000)050[0217:ESAAEF]2.3.CO;2).
22. Hoffmann, Ary A., and Juha Merilä. 1999. Heritable Variation and Evolution under Favourable and Unfavourable Conditions. *Trends in Ecology & Evolution* 14 (3): 96–101. [https://doi.org/10.1016/S0169-5347\(99\)01595-5](https://doi.org/10.1016/S0169-5347(99)01595-5).
23. Hurtado-Gonzales, O. P., Valentini, G., Gilio, T. A., Martins, A. M., Song, Q., & Pastor-Corrales, M. A. (2017). Fine mapping of Ur-3, a historically important rust resistance locus in common bean. *G3: Genes, Genomes, Genetics*, 7(2), 557-569.
24. Kaplan, L. 1965. Archeology and domestication in American *Phaseolus* (beans). *Econ. Bot* 19(4):358-368. doi:10.1007/BF02904806
25. Kelly, J. D., 2001 Remaking bean plant architecture for efficient production, pp. 109-143 in *Advances in Agronomy*. Academic Press.
26. Kraft, K. H., Brown, C. H., Nabhan, G. P., Luedeling, E., Ruiz, J. D. J. L., d’Eeckenbrugge, G. C., et al.. (2014). Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annum*, in Mexico. *Proceedings of the National Academy of Sciences*, 111(17), 6165-6170.
27. Levins R. 1962. Theory of fitness in a heterogeneous environment. I. The fitness set and adaptive function. *The American Naturalist* 96 (891): 361–73.
28. Lin, Tao, Guangtao Zhu, Junhong Zhang, Xiangyang Xu, Qinghui Yu, Zheng Zheng, Zhonghua Zhang, et al. 2014. Genomic Analyses Provide Insights into the History of Tomato Breeding. *Nature Genetics* 46 (11): 1220–26. <https://doi.org/10.1038/ng.3117>.
29. MacQueen, A. H., White, J. W., Lee, R., Osorno, J. M., Schmutz, J., Miklas, P. N., et al. (2020). Genetic Associations in Four Decades of Multienvironment Trials Reveal Agronomic Trait Evolution in Common Bean. *Genetics*, 215(1), 267-284.

30. Mamidi, S., M. Rossi, D. Annam, S. Moghaddam, R. Lee et al., 2011 Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant Biol.* 38: 953. <https://doi.org/10.1071/FP11124>
31. Miller, Allison J., and Briana L. Gross. 2011. From Forest to Field: Perennial Fruit Crop Domestication. *American Journal of Botany* 98 (9): 1389–1414.
32. Moyers, Brook T, Peter L Morrell, and John K McKay. 2018. Genetic Costs of Domestication and Improvement. *Journal of Heredity* 109 (2): 103–16. <https://doi.org/10.1093/jhered/esx069>.
33. Myers, J., 1988 The Cooperative Dry Bean Nursery. Reports of the Bean Improvement Cooperative: 209–210.
34. Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., et al. (2013). Community ecology package. R package version, 2(0).
35. Ortiz, R., Trethowan, R., Ferrara, G. O., Iwanaga, M., Dodds, J. H., Crouch, J. H., et al. (2007). High yield potential, shuttle breeding, genetic diversity, and a new international wheat improvement strategy. *Euphytica*, 157(3), 365-384.
36. Papa, R., & Gepts, P. (2003). Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical and Applied Genetics*, 106(2), 239-250.
37. Papa, R., Bellucci, E., Rossi, M., Leonardi, S., Rau, D., Gepts, P., et al. (2007). Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Annals of Botany*, 100(5), 1039-1051.
38. Parker, T. A., Berny Mier y Teran, J. C., Palkovic, A., Jernstedt, J., Gepts, P. (2020). Pod indehiscence is a domestication and aridity resilience trait in common bean. *New Phytologist*, 225(1), 558-570.
39. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
40. Ramirez-Villegas, J., Khoury, C. K., Achicanoy, H. A., Mendez, A. C., Diaz, M. V., Sosa, C. C., et al. (2020). A gap analysis modelling framework to prioritize collecting for ex situ conservation of crop landraces. *Diversity and Distributions*, 26(6), 730-742.
41. Resende, R. T., Piepho, H. P., Rosa, G. J., Silva-Junior, O. B., e Silva, F. F., de Resende, M. D. V., & Grattapaglia, D. (2020). Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. *Theoretical and Applied Genetics*, 1-18.
42. Rodriguez M, Rau D, Angioi SA, Bellucci E, Bitocchi E, Nanni L, Knupffer H, Negri V, Papa R, Attene G. 2013. European *Phaseolus coccineus* L. Landraces: population structure and adaptation, as revealed by cpSSRs and phenotypic analyses. *PLoS ONE* 8: e57337.

- 533 43. Roger S. Bivand, Edzer Pebesma, Virgilio Gomez-Rubio, 2013. Applied spatial data analysis with
534 R, Second edition. Springer, NY.
- 535 44. Schmutz, Jeremy, Phillip E. McClean, Sujana Mamidi, G. Albert Wu, Steven B. Cannon, Jane
536 Grimwood, Jerry Jenkins, et al. 2014. A Reference Genome for Common Bean and Genome-
537 Wide Analysis of Dual Domestications. *Nature Genetics* 46 (7): 707–13.
538 <https://doi.org/10.1038/ng.3008>.
- 539 45. Singh, S. 1989 Patterns of Variation in Cultivated Common Bean (*Phaseolus vulgaris*, Fabaceae).
540 *Economic Botany* 43(1):39-57.
- 541 46. Singh, S., Gepts, P., & Debouck, D. (1991). Races of common bean (*Phaseolus vulgaris*,
542 Fabaceae). *Economic Botany*, 45, 379–396. <https://doi.org/10.1007/BF02887079>
- 543 47. Singh, S., 2000 50 years of the Cooperative Dry Bean Nursery. Reports of the Bean
544 Improvement Cooperative: 110–111.
545 <https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.13713>
- 546 48. Soltani, A., M. Bello, E. Mndolwa, S. Schroder, S. M. Moghaddam *et al.*, 2016 Targeted Analysis
547 of Dry Bean Growth Habit: Interrelationship among Architectural, Phenological, and Yield
548 Components. *Crop Science* 56: 3005.
- 549 49. Trucchi E, Benazzo A, Lari M, Iob A, Vai S, Nanni L, Belucci E, Bitocchi E, Raffini F, Xu C et
550 al. 2021. Ancient genomes reveal early Andean farmers selected common beans while preserving
551 diversity. *Nature Plants* 7: 123–128.
- 552 50. Vandemark, G. J., Brick, M. A., Osorno, J. M., Kelly, J. D., & Urrea, C. A. (2014). Edible grain
553 legumes. *Yield gains in major US field crops*, 33, 87-123.
- 554 51. Voss-Fels, K. P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., et al. (2019).
555 Breeding improves wheat productivity under contrasting agrochemical input levels. *Nature plants*,
556 5(7), 706-714.
- 557 52. Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., et al. (2016).
558 Springer-Verlag. New York.
- 559 53. Wilson, A. J., J. M. Pemberton, J. G. Pilkington, D. W. Coltman, D. V. Mifsud, T. H. Clutton-
560 Brock, and L. E. B. Kruuk. 2006. Environmental Coupling of Selection and Heritability Limits
561 Evolution. *PLOS Biology* 4 (7): e216. <https://doi.org/10.1371/journal.pbio.0040216>.

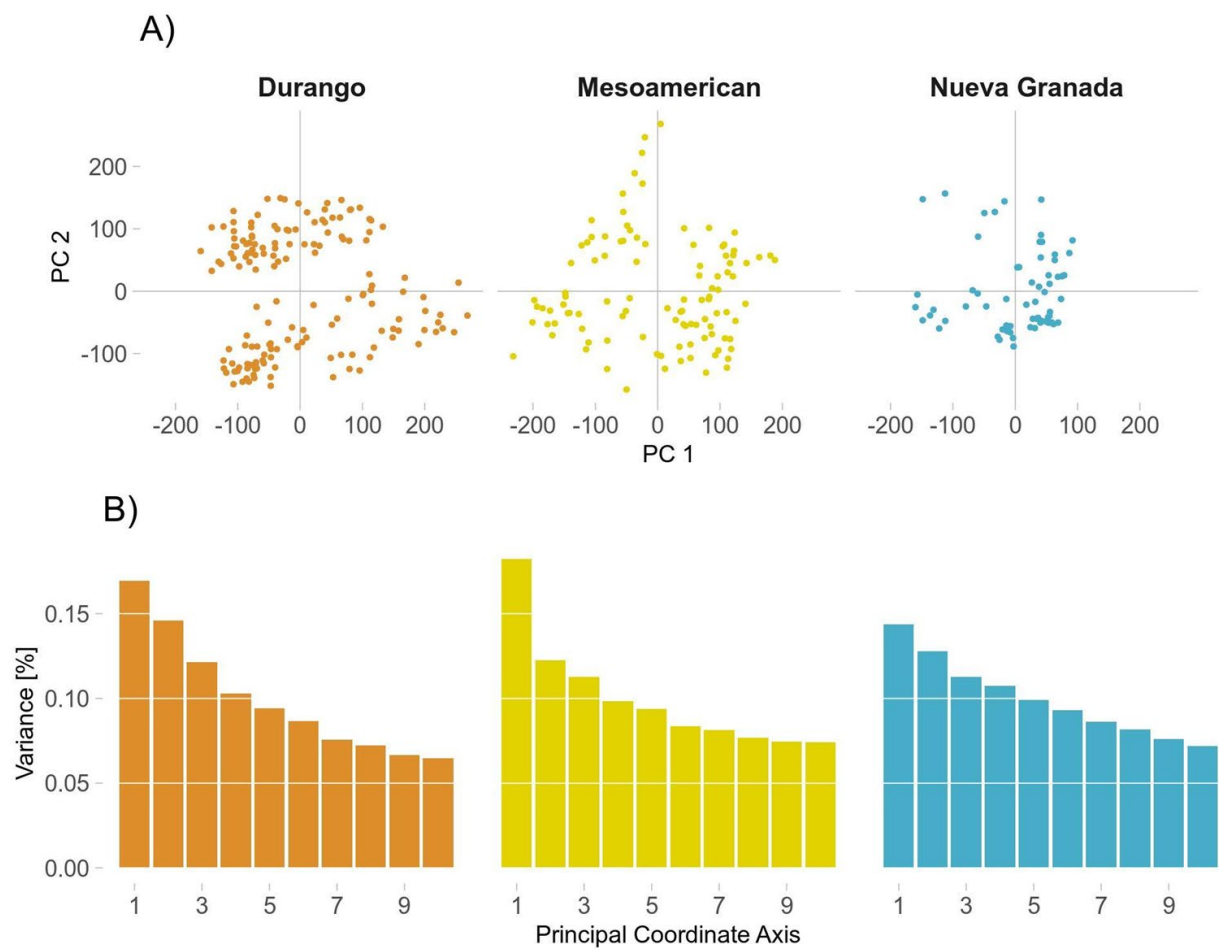


Figure S1: Summary of kinship among varieties within race. A) ordination of varieties along principal component axes 1 and 2. B) Scree plot of variance explained by successive principal axes.

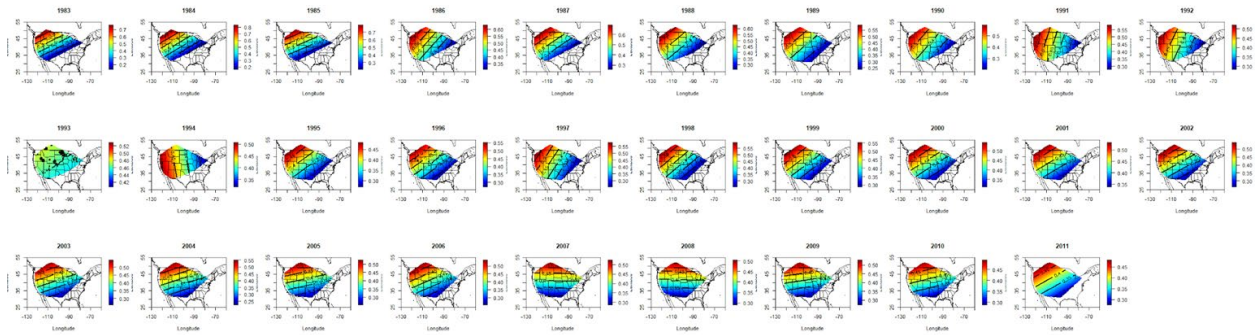


Figure S2. Ordinary kriging of yield heritability for each 5-year sliding window (1981-2013) centered on the third year of the window.

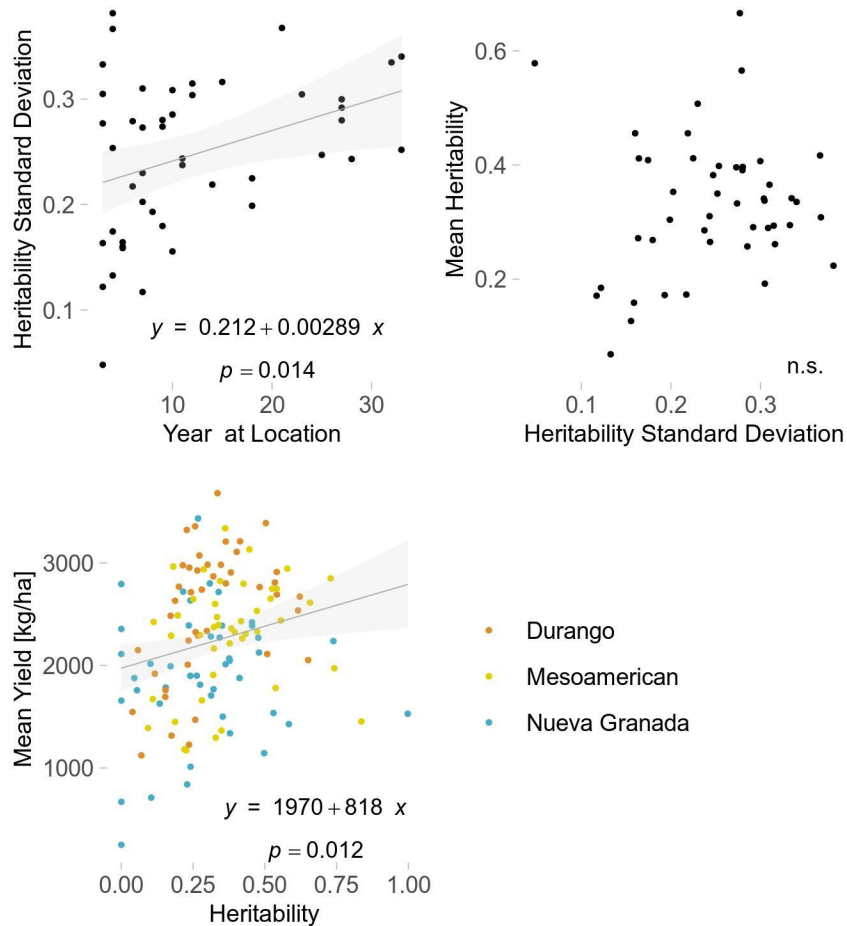


Figure S3: a) relationship between standard deviation in heritability at each site, and the number of trials at that site; b) relationship between mean and standard deviation in heritability at each site; c) relationship between mean yield and mean heritability at each site. n.s. – not significant at $p < 0.10$.

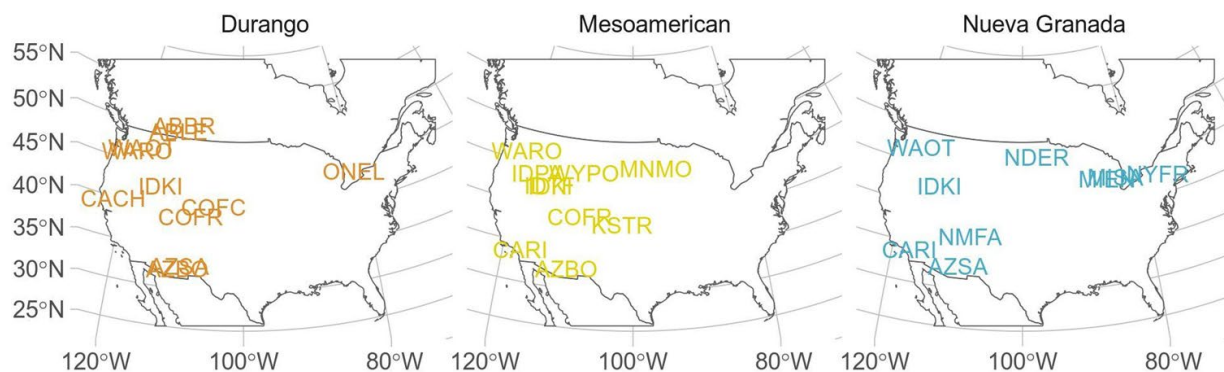


Figure S4. High heritability and high HFA location combinations by bean race. Abbreviations correspond to CDBN standard abbreviations.

Table S1

ANOVA of Yield Across Time

	Sum Sq	Df	F value	Pr(>F)
Year	252,500,000	1	292.70	0.000
Race	1,050,000,000	2	608.26	0.000
Year*Race	6,895,000	2	4.00	0.018
Residuals	11,370,000,000	13174	-	-

Table S2

AIC¹ of Competing Home Field Advantage Models

	Durango	Mesoamerica n	Nueva Granada
With Home Field	6772	5095	3434
Base (Variety, Site, Year)	7323	5421	3599
With Home Field, Kinship	7662	5703	3773
With Kinship	8143	5997	3926

¹

Akaike Information Criterion

580

Table S3
ANOVA of Home Field Advantage Across Time

	Sum Sq	Df	F value	Pr(>F)
Year	4.755	1	0.00	0.992
Race	36,290	2	0.43	0.654
Year * Race	410,000	2	4.82	0.010
Residuals	3,994,000	94	-	-

581