**From local to continent-scale variation in fitness and heritability in common bean (*Phaseolus vulgaris* L.)**

**Potential Authors**

Alice H. MacQueenꟸ1,

Jeffrey W. White, Phil Miklas, Phillip E. McClean, Juan M. Osorno, Bryan Runck, Thomas Kono, Colin K. Khoury,

Michael Kantar3, Patrick M. Ewing2\*

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

2North Central Agricultural Research Laboratory, Agricultural Research Service, USDA, 2923 Medary Ave, Brookings, SD, 57006, USA

3Department of Tropical Plant and Soil Science, University of Hawaii at Manoa, Honolulu, HI, United States of America

4GEMS Agroinformatics Initiative, University of Minnesota, Minneapolis, MN, United States of America

5Minnesota Supercomputing Institute, Minneapolis, MN, United States of America

6International Center for Tropical Agriculture (CIAT), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, Cali, 763537, Colombia

ꟸFirst author

\* Corresponding author

**Target Audience:** Those interested in long term effects of selection and domestication and the food security implications thereof

**Target journal:** New Phytologist

**ABSTRACT**

* In domesticated species under continuous selection for broad environmental adaptation, it is unknown whether locally adapted alleles still segregate within highly improved, contemporary genetic material. In common bean (*Phaseolus vulgaris* L.), we quantify long-term improvement impacts on fitness heritability and fitness variation attributable to local adaptation.
* We quantify patterns of adaptation breadth in 327 commercial common bean entries planted across 70 sites from 1981–2015, for the Cooperative Dry Bean Nursery (CDBN) trials. We also quantified patterns of heritability across sites and time to estimate the efficiency of selection for seed yield. Most common bean varieties cultivated in the US belong to one of three races with differing domestication and improvement histories, which resulted in contrasting patterns of local adaptation and heritability.
* The Durango and Mesoamerican races both enjoyed higher-than-expected (*p* = 0.002) home field advantage, equal to 34 years of selection gains. The Andean races did not demonstrate local adaptation. Andean and Durango yields became more heritable across the study period (*p* < 0.001), but Mesoamerican fitness heritability decreased (*p* < 0.001).
* During improvement in the CDBN, we find agronomically significant increases in fitness attributable to local adaptation in both domestication clades. Surprisingly, we find that heritability has been increasing over time in both domestication clades.

**Key Words:** Domestication, time-series, long term selection, food security, home field advantage

# INTRODUCTION

Genetic variation in phenotypes is the fundamental property that makes both natural and artificial selection effective. However, there is a consistent loss of genetic variation when “improved” or commercial genomes are compared to domesticated “landrace” or non-commercial genomes, and again when domesticated genomes are compared to 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 in common bean? range from 5 to 77% between wild and improved populations (Gray et al 2009, Lin et al 2014), and the only case with a gain in genetic diversity is in the Andean domestication event in common bean (Schmutz et al 2014).

Common bean (*Phaseolus vulgaris*) is the most important food legume for direct human consumption. Common bean is currently grown across a large range of latitudes (from 52 N to 32 S), altitudes (50 – 3000m), growth temperatures (14-26 C), and average yearly precipitation (400-1600mm) and precipitation regimes (Singh 1989; FAOSTAT 2015), on over 35 million ha globally with production in nearly every country that keeps statistics (FAOSTAT 2015). In contrast, the habitat niche of wild common bean is extremely narrow (Gepts 2009). Common bean originated in Mesoamerica (Bitocchi et al 2012) and was 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 and once in the Andes (Gepts, et al. 1986; Schmutz, et al. 2014). The domestication bottleneck was stronger in the Mesoamerican than in the Andean gene pool, 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). The gain in genetic diversity during the Andean domestication event was likely a result of gene flow from the more genetically diverse Mesoamerican common bean (Schmutz et al 2014). Common bean was widely cultivated in Mexico and North America during pre-Columbian times (Kaplan 1965), and New World settlers cultivated common bean from landraces cultivated by Native Americans and from European introductions (Vandemark et al 2014). The first 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, these improvements have typically taken advantage of a multi-environment trial known as 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 at their home locations, then test that material for broad adaptation across all CDBN field locations, typically releasing varieties with high fitness (largely seed yield) across CDBN locations. This approach maximizes the expected performance of released varieties across a large geographical area, but may ignore benefits of environmental specialization. Recently, hundreds of CDBN entries that group into three major races from both independent domestication events have been genotyped at 1.2 million SNPs and used for genetic work (MacQueen et al 2020). The CDBN dataset offers an exciting opportunity to examine the genetic component of fitness in two domestication clades which differ in their genetic diversity, and to examine the impact of common bean improvement across time on genetic variation in fitness at local and continental scales.

CDBN entries are advanced breeding material without ‘locations of origin’ in the sense of their wild progenitors. However, most? progenitors of these entries were grown in North America in pre-Columbian times, and these specific entries were developed by breeders at many locations that took part in the CDBN trials; thus, local adaptation of CDBN entries to specific North American regions may have arisen from long-term selection of improved material within those regions. If present, local adaptation could also be leveraged in breeding programs which have focused on broad-scale adaptation for large fitness gains (Ewing et al., 2019). For individual-based selection, breeding programs often measure heritability - the proportion of phenotypic variance explained by additive genetic variance relative to the total phenotypic variance. High heritability is essential for efficient selection, but environmental conditions also influence the heritability on which any response to selection depends (Hoffman, Hercus 2000; Hoffman, Merila 1999), and heritability may also vary temporally as the array of genotypes measured changes during the breeding process (Feldman, Lewontin 1975). Understanding spatial and temporal patterns of seed yield heritability and the interaction between heritability and the portion of seed yield variation attributable to local adaptation could provide fundamental insights into the adaptation and domestication process of *P. vulgaris* L., which may differ by domestication clade (Papa & Gepts 2003; Papa et al., 2007; Eckert et al 2010;Kraft et al 2014; Rodriguez et al 2015).

Here, we introduce a home field advantage (HFA) component for each CDBN entry into a traditional partitioning of seed yield variation and explore patterns of HFA across these races and gene pools. Our assumption was that higher performance at an entry’s home location would explain a significant amount of seed yield variation within each race. We then examined overall variation in seed yield heritability across years and locations in the CDBN and across races and gene pools. 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 gene pool as genetic sources of yield variation were eliminated, and potential increases in heritability over time in the Andean gene pool, given its increase in genetic diversity from wild to domesticated types. We conclude by investigating the potential for local adaptation to be incorporated into breeding programs, by characterizing the relationship between yield heritability and home field advantage across CDBN locations.

# MATERIALS AND METHODS

### 2.1 Data Acquisition

Common bean seed 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. CDBN entries include released cultivars and unreleased breeding lines at advanced breeding stages from most bean types grown in North America and represent at least 13 diverse market classes of common bean that group into three major races from both independent domestication events (Mamidi, et al. 2011). In addition to the 13989 yield datapoints used in this analysis, each entry was genotyped using genotyping by sequencing resulting in 1,221,540 SNPs.

The selection strategy in the CDBN likely impacts partitioning of yield variance in this material in a number of ways. This strategy in the CDBN was developed to test dry bean material for broad adaptation across CDBN locations. Within years, CDBN locations grew an identical complement of CDBN entries, with only rare entries missing yield phenotypes at single locations. The entries in the CDBN shift gradually through the years, with each entry typically present in the CDBN for three consecutive years. This shift in CDBN genotypes is a direct result of the improvement strategy employed by CDBN cooperators over the trial years. This strategy confounds partitioning of yield variation into genetic and environmental components across years, but not across locations. Specifically, comparisons of heritabilities for different locations within years are not confounded by the presence of different genotypes. Heritabilities of 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 Expectation of home field advantage

HFA 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.

We began estimating HFA by filtering out sites with fewer than three years of data. Next, we assigned each entry a “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 highest BLUP score.

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

***y*** = **X*β* + *ε***

Where ***y*** contains the *n* yields of each entry *i* at location *j* and year *k*, ***β*** contains coefficients of *p* predictors in the *n* by *p* model matrix **X**, and ***ε*** is the associated error term. In the base model, **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 to the base model terms for genetic relatedness. Genetic relatedness was determined using the snp\_autoSVD() function of the bigsnpr R package (Prive et al 2017); 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 in 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; 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 what would be 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

Where Ventries the variance among the entries, environment is the number of locations in the study and MSentries 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 considered to be 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 having a higher-than-median heritability, and also 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

### Unless otherwise noted: statistics were performed in R 3.6.3 base and stats packages, and 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); entering and scaling refer to a group mean of zero and standard deviation of one; errors are standard errors; and all analyses were performed independently for each race.

# RESULTS

*3.1 Home Field Advantage*

We first assessed overall fitness and fitness gains across the study period. We assessed this for three groups of CDBN entries: entries from the Andean domestication clade, and entries from two genetically distinct races within the Middle American domestication clade, the Durango and the Mesoamerican race. Each of the major dry bean races had a different fitnesses by 2015 (*F*2,13170 = 608; *p* < 0.001, Figure 2a; Table S1): Andean entries were expected to yield 2370 ± 40 kg ha-1, Durango entries, 3060 ± 30 kg ha-1, and Mesoamerican entries, 2850 ± 40 kg ha-1. In agreement with previous studies (Vandemark et al 2014), all races’ seed yields improved during the study period, albeit at different rates (*F*2,13170 = 4.00, *p* = 0.02). Andean seed yields grew by 16 ± 2 kg ha-1 yr-1 (*p* < 0.001), and Durango seed yields grew by 14 ± 1 kg ha-1 yr-1 (*p* < 0.001). Mesoamerican seed yields grew the fastest, by 20 ± 2 kg ha-1 yr-1 (*p* < 0.001).

We partitioned yield variation into genetic and environmental components within the two domestication clades and three races (Table 2). 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 seed yield explained by each CDBN entry’s home field advantage (HFA). In Andean entries, HFA accounted for 1.1% of yield variation, and reduced unexplained seed yield variation by 6.4% (Table 2). In Durango entries, HFA accounted for 1.3% of seed yield variation and reduced residual variation by 8.3%. Similarly, HFA accounted for 1.3% of Mesoamerican seed 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 seed yield reduced model parsimony (Δ*AIC* = 339 (Andean), 890 (Durango), and 608 (Mesoamerican); Table S2). These three PCs accounted for 39%, 44%, and 42% of genetic variance in the Andean, Durango, and Mesoamerican lines, respectively (Figure S1).

We next determined if race-average HFA were larger than expected and whether these average HFAs were agronomically significant. Andean entries had seed yield gains that were not significantly different from the expectation of 357 kg ha-1 based on permutations of variation in Andean seed yields (*p* = 0.21; Figure 2b). In Durango entries, HFA conferred an average seed 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 lines, 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 has been preserved for the Mesoamerican, but not the Andean domestication clade.

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), though 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). There was a significant positive relationship between heritability and environmental quality of each location-year, with environmental quality measured as the average yield of entries of a race in that site-year (Fig S3C; Finlay and Wilkinson, 1963). However, this trend was confounded with trends across time, which reflected the direct act of selection for improvement on CDBN entries. Specifically, there were trends in yield heritability across years which varied by bean race. As hypothesized, heritability decreased in Mesoamerican entries by 0.005 yr-1 (*p* < 0.001). In contrast, Andean and Durango entries showed an increase in yield heritability over time, with yield heritability increasing by 0.007 yr-1 and 0.004 yr-1, respectively (*p* < 0.001). To test for a genetic basis for environmental specialization we determined if more closely related CDBN entries had spatially closer home sites. Entries in the Durango (Mantel’s *r* = 0.11; *p* = 0.001) and Mesoamerican (Mantel’s *r* = 0.068; *p* = 0.01) races showed this relationship, while entries in the Andean race did not (*p* = 0.6). Finally, we found no tradeoff between heritability and home field advantage within races (*F*2,97 = 1.27; *p* = 0.3; Figure 4).

# DISCUSSION

*4.1 Past Selection Indicates Orthogonal Axes of Future Improvement*

Modern breeding techniques have achieved remarkable success in increasing the fitness and phenotypic stability of cultivated plants through breeding programs that do not optimize for specific regions. Efforts such as the CDBN have indirectly? selected for broad environmental adaptation; however, ecological theory suggests that across environments that vary spatially or temporally, the most productive 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. Thus, to continue to improve 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 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 this species. Here, we measure HFA as a proxy for the amount of 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 background yield improvements. We correlate these HFA with the amount of variance segregating for entry-mean heritability (Figure 4). We envision that the interplay between HFA, an entry-specific quantity, and h2, 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 involving individual-based selection 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 particular environments. When trait heritability is low and HFA are 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 both trait heritability and HFA are not present for CDBN entries (Table 1; Figure 4, quadrant III), neither traditional breeding approaches nor screening for locally important alleles will prove valuable for further common bean improvement.

Given the different evolutionary histories of both domestication clades pre- and post-domestication, we expected distinct responses to selection in the Andean and Mesoamerican entries over the course of the CDBN. The low-yielding, low genetic diversity Andean domestication clade 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 heritabilities 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 are still maladapted to North American growing conditions. Further introgression from the Middle-American domestication clade or from other species may offer opportunities for continued yield improvement and local adaptation in this domestication clade.

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 heritabilities 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, 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. 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). We also observed a sustained drop in yield heritabilities relative to the line-of-best-fit in the mid-2000s. 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 heritabilities 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 and/or architecture alleles into populations is equivalent to an artificially induced selective sweep. In this case, we expect to see considerable linkage drag and the loss of locally adapted genetic variants. This signal was observed in the CDBN where after introgression, many years were required to regain local adaptation (HFA) after using a common parent for important traits.

*4.2 HFA across species: comparing apples and oranges*

Previous work quantifying HFA in maize found large benefits of local adaptation (Ewing et al 2019), but was unable to determine if accounting for genetic relatedness was an effective substitute for HFA. Here, we found that HFA was robust to familial relationships within breeding programs across sites, supporting the utility of this metric, and additionally compared HFA across domestication clades within common bean. Extending this comparison across species, HFA relative to annual breeding gains was four-fold larger in common bean than in maize (*Zea mays* L.). Maize HFA was 1.0 Mg ha−1year−1, which, while larger than observed in common bean in this study, was equivalent to just 4.5 years of observed yield gains and was smaller than expected (Ewing et al., 2019). In contrast, in this study, HFA was 450-470 kg ha−1 yr−1, equivalent to 22 and 34 years of annual breeding gains for Mesoamerican and Durango races, respectively, and was larger than expected. A number of processes may underlie this difference in HFA between species. First, the geographic extent of the current study spans thousands of kilometers, versus a few hundred in Ewing et al (2019), and encompasses a much broader range of environments. Additionally, selection of common bean has historically happened at regional scales for specific sub-races (market classes) (Vandermark et al, 2014), although stable or decreasing HFA in Andean and Mesoamerican varieties suggests this historic pattern 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 vs. hybrid monocot), local adaptation is both valuable to future breeding efforts and weakening as a result of improvement efforts.

Reversing this trend would necessitate a return to selection in more dispersed breeding programs and at narrower environmental scales. Such selection might be more efficient in certain locations where both HFA and heritability are high; these race-location matches we observed for dry bean (Supplemental Figure 4) loosely follow the pattern described in Vandermark et al (2014) which suggested regional breeders focused on specific market classes which descend from different races. Challenges to adopting a more localized breeding strategy include identifying the extent of “local”, ensuring the introgression of globally important agronomic traits (e.g. plant architecture and disease resistance), and maintaining diverse gene pools of locally adapted parents. Still, the benefits of leveraging favorable genotype-by-environment interactions - local adaptation - may have profound implications for food security.

# Conclusion

Plant populations in natural systems are expected to be adapted to local conditions; in contrast, crop breeding programs typically select for high fitness across a breadth of conditions (Finlay and Wilkenson, 1963). During improvement in the CDBN, we find agronomically significant yield gains attributable to environmental specialization in both domestication clades. Surprisingly, we find that yield heritability has been increasing over time in races from both domestication clades, indicating that genetic diversity in both clades may have increased from introgression events. In the CDBN we see the expected evidence for broad adaptation contingent on population (race), influenced by independent domestication events that have experienced different selection histories. We show that focusing on narrower geographies could provide independent increases in fitness relative to the common selection strategies. Resilient food systems of the future are going to have to take advantage of all the potential avenues of improvement in order to sustainably produce food adapted to future unknown environments.

# 

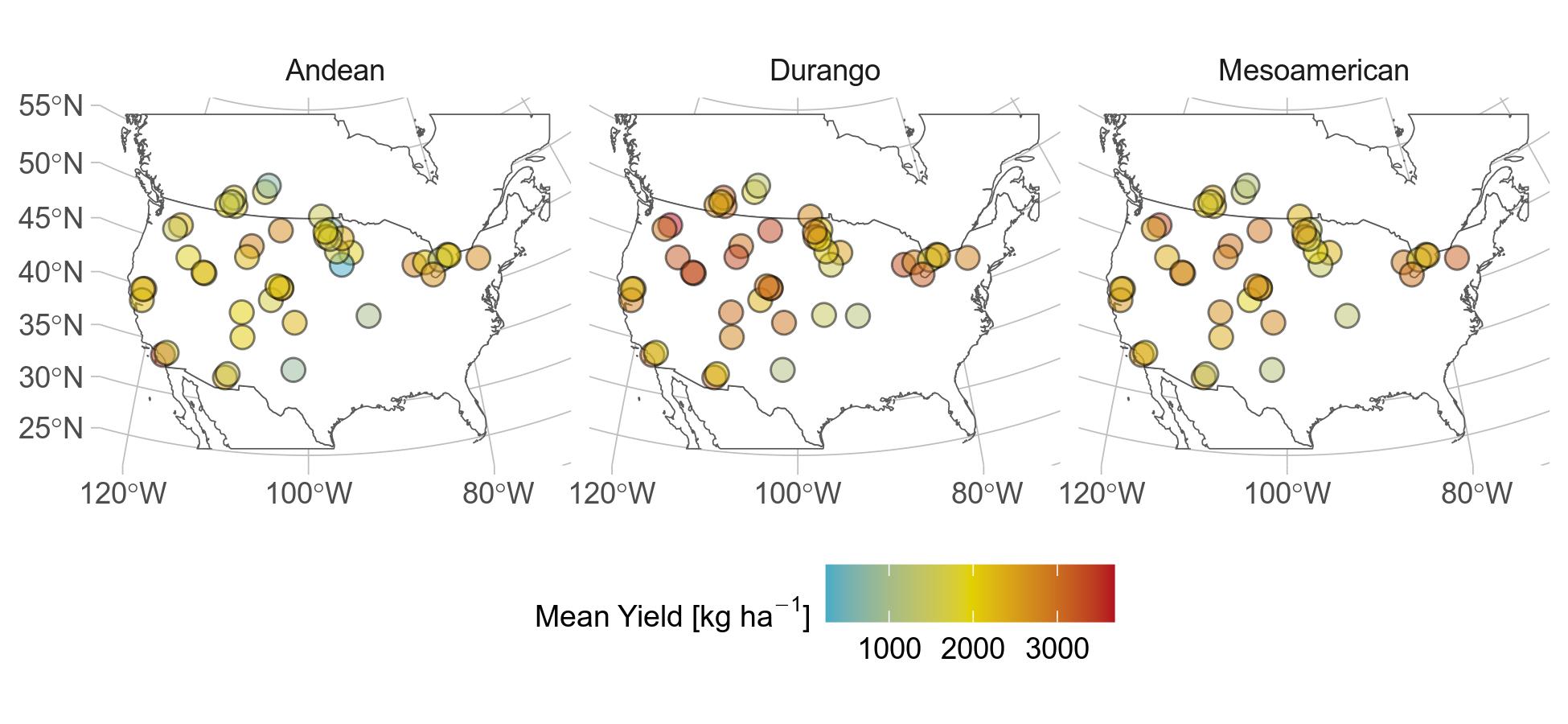
# References

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). Ime4: Linear Mixed-EffectsMoels 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
5. Bellucci, Elisa, Elena Bitocchi, Alberto Ferrarini, Andrea Benazzo, Eleonora Biagetti, Sebastian Klie, Andrea Minio, et al. 2014. Decreased Nucleotide and Expression Diversity and Modified Coexpression Patterns Characterize Domestication in the Common Bean. *The Plant Cell* 26 (5): 1901–12. https://doi.org/10.1105/tpc.114.124040.
6. Bitocchi, Elena, Elisa Bellucci, Alessandro Giardini, Domenico Rau, Monica Rodriguez, Eleonora Biagetti, Rodolfo Santilocchi, et al. 2013. Molecular Analysis of the Parallel Domestication of the Common Bean (Phaseolus Vulgaris) in Mesoamerica and the Andes. *New Phytologist* 197 (1): 300–313.
7. Bitocchi, Elena, Laura Nanni, Elisa Bellucci, Monica Rossi, Alessandro Giardini, Pierluigi Spagnoletti Zeuli, Giuseppina Logozzo, et al. 2012. Mesoamerican Origin of the Common Bean (Phaseolus Vulgaris L.) Is Revealed by Sequence Data. *Proceedings of the National Academy of Sciences* 109 (14): E788–96. https://doi.org/10.1073/pnas.1108973109.
8. Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., & Pebesma, E. J. (2013). Applied spatial data analysis with R (Vol. 2). New York: Spring
9. Eckert, C. G., Kalisz, S., Geber, M. A., Sargent, R., Elle, E., Cheptou, P. O., ... & Winn, A. A. (2010). Plant mating systems in a changing world. Trends in Ecology & Evolution, 25(1), 35-43.
10. Ewing, P. M., Runck, B. C., Kono, T. Y., Kantar, M. B. (2019). The home field advantage of modern plant breeding. PloS one, 14(12), e0227079.
11. FAOSTAT, 2015 Food and Agriculture Organization of the United Nations, pp.,<http://faostat3.fao.org/>
12. Farid, M., Earl, H. J., Pauls, K. P., & Navabi, A. (2017). Response to selection for improved nitrogen fixation in common bean (Phaseolus vulgaris L.). Euphytica, 213(4), 99.
13. Feldman, M. W., and R. C. Lewontin. 1975. The Heritability Hang-Up. *Science* 190 (4220): 1163–68.
14. Finlay KW, Wilkinson GN. The analysis of adaptation in a plant-breeding programme. Australian journal of agricultural research. 1963;14(6):742–54.
15. Fischer, R. A. (2015). Definitions and determination of crop yield, yield gaps, and of rates of change. *Field Crops Research*, *182*, 9-18.
16. Fox, John, & Weisberg, Sanford. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/
17. 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).
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. Kaplan, L. 1965. Archeology and domestication in American Phaseolus (beans). Econ. Bot 19(4):358-368. doi:10.1007/BF02904806
24. Kelly, J. D., 2001 Remaking bean plant architecture for efficient production, pp. 109-143 in *Advances in Agronomy*. Academic Press.
25. 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 annuum, in Mexico. Proceedings of the National Academy of Sciences, 111(17), 6165-6170.
26. Levins R. Theory of fitness in a heterogeneous environment. I. The fitness set and adaptive function. The American Naturalist. 1962 Nov 1;96(891):361–73.
27. 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.
28. 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.
29. 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>
30. Miller, Allison J., and Briana L. Gross. 2011. From Forest to Field: Perennial Fruit Crop Domestication. American Journal of Botany 98 (9): 1389–1414.
31. 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.
32. Myers, J., 1988 The Cooperative Dry Bean Nursery. Reports of the Bean Improvement Cooperative: 209–210.
33. 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).
34. 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.
35. 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.
36. 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.
37. Vandemark, G. J., Brick, M. A., Osorno, J. M., Kelly, J. D., & Urrea, C. A. (2014). Edible grain legumes. Yield gains in major US field crops, 33, 87-123.
38. Voss-Fels, K. P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., et el. (2019). Breeding improves wheat productivity under contrasting agrochemical input levels. *Nature plants*, *5*(7), 706-714.
39. 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.
40. 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.
41. Roger S. Bivand, Edzer Pebesma, Virgilio Gomez-Rubio, 2013. Applied spatial data analysis with R, Second edition. Springer, NY.
42. Schmutz, Jeremy, Phillip E. McClean, Sujan Mamidi, G. Albert Wu, Steven B. Cannon, Jane Grimwood, Jerry Jenkins, et al. 2014. A Reference Genome for Common Bean and Genome-Wide Analysis of Dual Domestications. *Nature Genetics* 46 (7): 707–13. https://doi.org/10.1038/ng.3008.
43. Singh, S. 1989 Patterns of Variation in Cultivated Common Bean (*Phaseolus vulgaris*, Fabaceae). Economic Botany 43(1):39-57.
44. Singh, S., Gepts, P., & Debouck, D. (1991). Races of common bean (Phaseolus vulgaris, Fabaceae). Economic Botany, 45, 379–396. <https://doi.org/10.1007/BF02887079>
45. Singh, S., 2000 50 years of the Cooperative Dry Bean Nursery. Reports of the Bean Improvement Cooperative: 110–111. <https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.13713>
46. Soltani, A., M. Bello, E. Mndolwa, S. Schroder, S. M. Moghaddam *et al.*, 2016 Targeted Analysis of Dry Bean Growth Habit: Interrelationship among Architectural, Phenological, and Yield Components. Crop Science 56**:** 3005.
47. 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/.
48. 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.
49. Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., et al. (2016). Springer-Verlag. New York.
50. Wilson, A. J., J. M. Pemberton, J. G. Pilkington, D. W. Coltman, D. V. Mifsud, T. H. Clutton-Brock, and L. E. B. Kruuk. 2006. Environmental Coupling of Selection and Heritability Limits Evolution. PLOS Biology 4 (7): e216. https://doi.org/10.1371/journal.pbio.0040216.

# Figures

## Figure 1

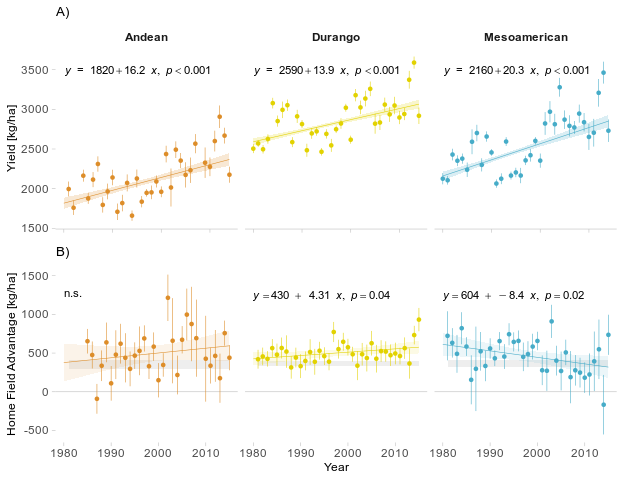
**Figure 1:** Locations and mean yields at sites within the CDBN variety trials. Only continental North American sites used for at least three years are shown.



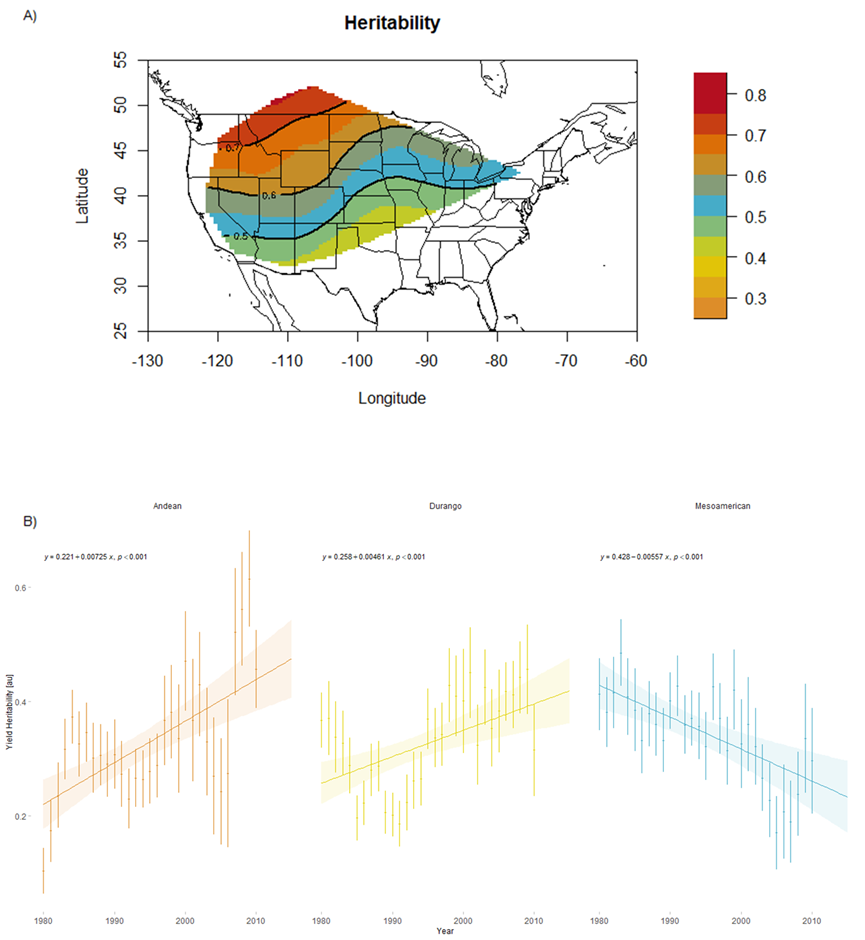
## 

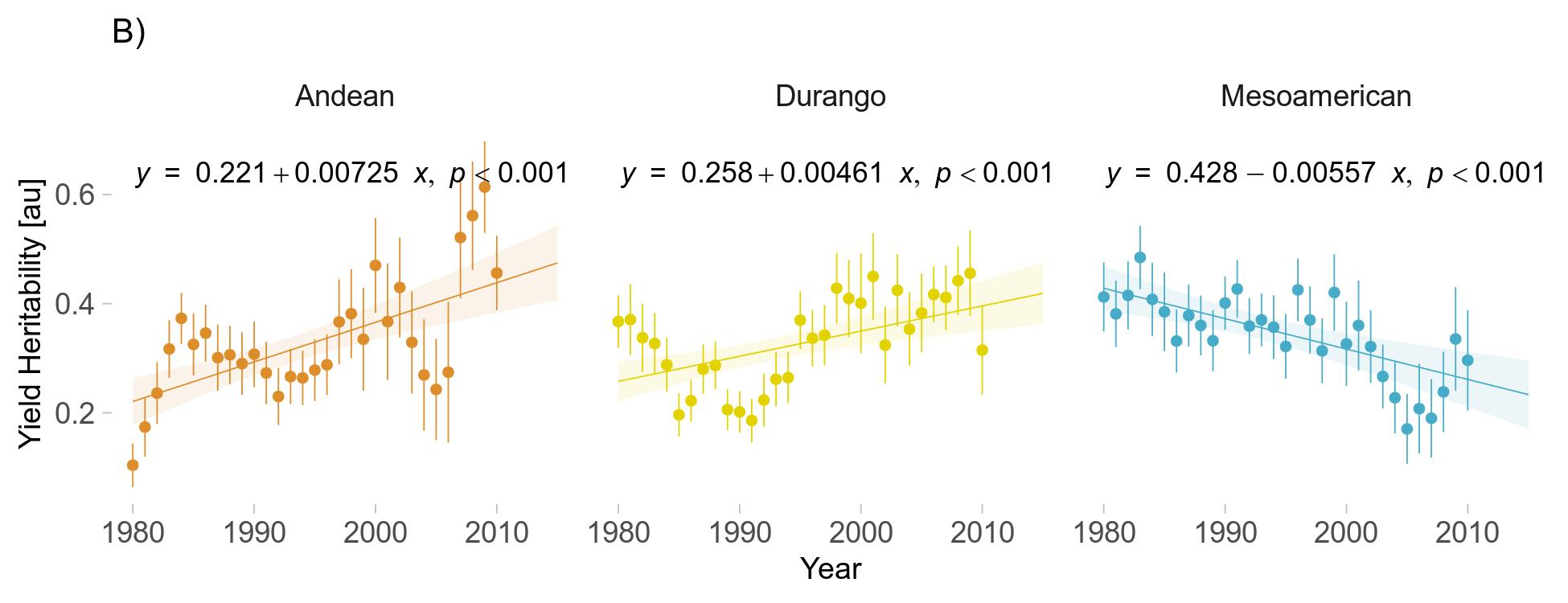
## Figure 2

**Figure 2**: Trends in A) yield and B) home field advantage across the study period by race. In B), the gray bar indicates 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.



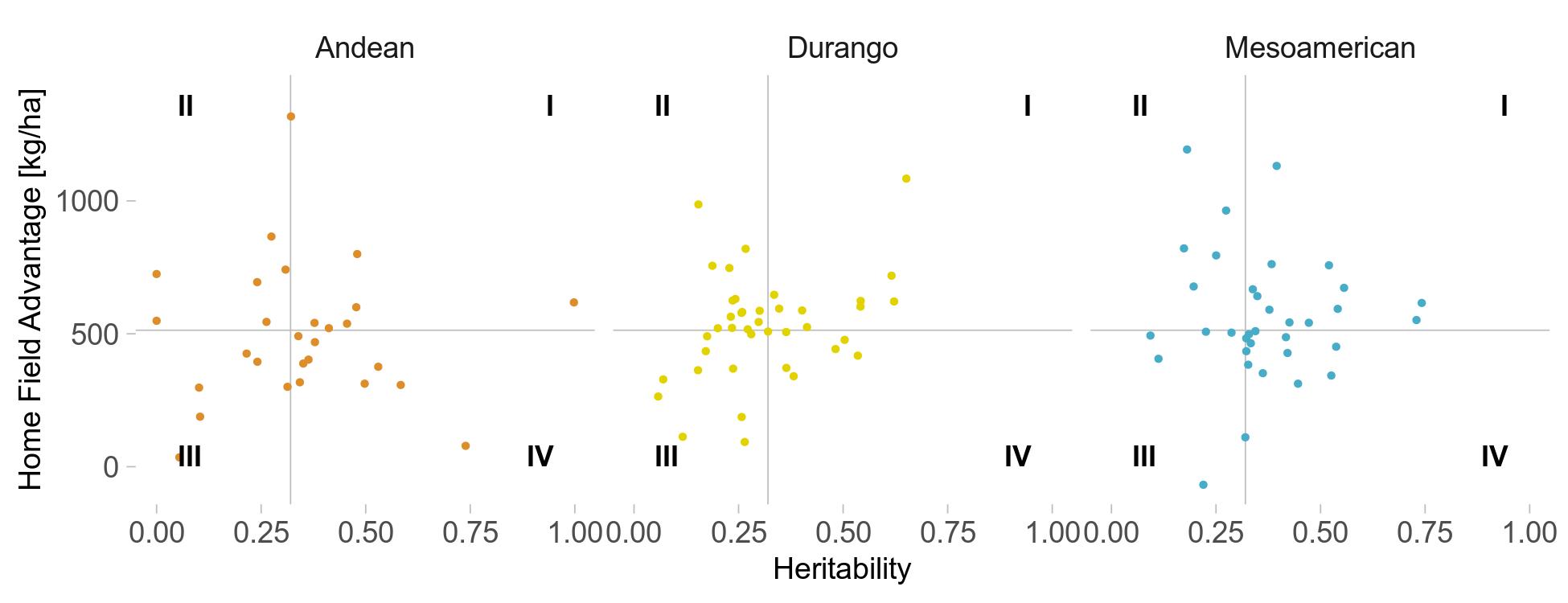
## Figure 3

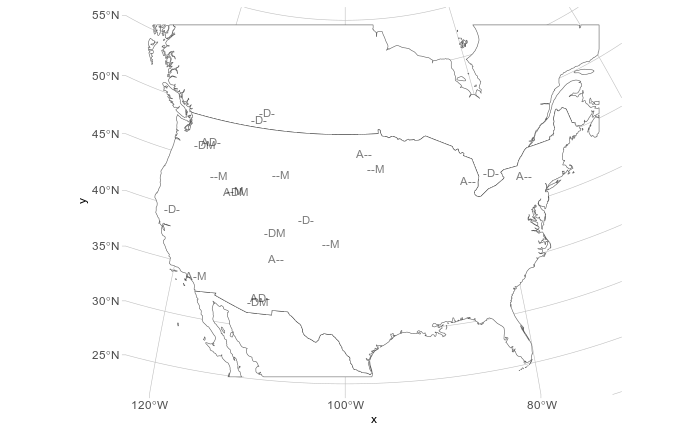
**Figure 3:** *A) Ordinary kriging of yield heritability based on the average hertailblity over 30-years of trials at 70 field sites,. B) Mean and standard deviation of site heritability for 70 field sites; sites with a high heritability and low variance provide better opportunity for selection. *

**

## Figure 4

**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.





Heritability for specialists would either be lower (everywhere) or more variable (between locations)

# Tables

## 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. In this study heritability spatially and temporally variable, high heritability traits tend to be <0.8 tend to have mendelian inheritance, making selection efficient in any environment. Traits with low heritabilities (>0.4) often differ between environments due to the genotype x environment interaction causing larger phenotypic variance and making selection inefficient on individuals. In this study the observed decay in heritability over time implies that major genes were fixed a long time ago and that selection for narrow locally adapted types is likely a place where improvement can still be made.

|  |  |  |  |
| --- | --- | --- | --- |
| Table 1: Implications of home field advantage and heritability for breeding and adaptation. | | | |
|  | | **Heritability** | |
| **Low** | **High** |
| **Home Field Advantage** | **Present** | *Processes*: High environmental variation among locations, low genetic variation potentially enriched in locally important alleles.  *Implications*: Inefficient selection on individuals for phenotypic improvement; testing at these sites may reveal conditionally beneficial alleles. | *Processes*: High environmental variation among locations, high genetic variation potentially enriched in locally important alleles.  *Implications*: Efficient identification and selection on specialists that may contain large effect candidate loci for introgression. |
| **Absent** | *Processes*: Low environmental variation among sites, low genetic variation potentially enriched in broadly important alleles.  *Implications*: Inefficient selection on individuals for phenotypic improvement, so family based methods are necessary. | *Processes*: Low environmental variation among locations, high genetic variation potentially enriched in broadly important alleles.  *Implications*: Efficient selection on individuals for phenotypic improvement (e.g. entries in the data set). |

## Table 2

|  |  |  |  |
| --- | --- | --- | --- |
| Table 2. Partitioning of CDBN Yield Variances | | | |
| Partitioning of CDBN Yield Variances | | | |
|  | Race | | |
| Andean | Durango | Mesoamerican |
| Home Site | 1.1% | 1.3% | 1.3% |
| Residuals | 15.6% | 14.1% | 15.8% |
| Genotype | 3.2% | 3.3% | 4.1% |
| Site | 29.0% | 33.1% | 30.9% |
| Year | 2.5% | 2.4% | 3.0% |
| Site-Year | 48.7% | 45.7% | 44.9% |
| Residual Reduction*1* | 6.3% | 8.3% | 7.7% |
| *1*Proportion of residual variance explained by home site | | | |

# 

# Supplement

## Figure S1

**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

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

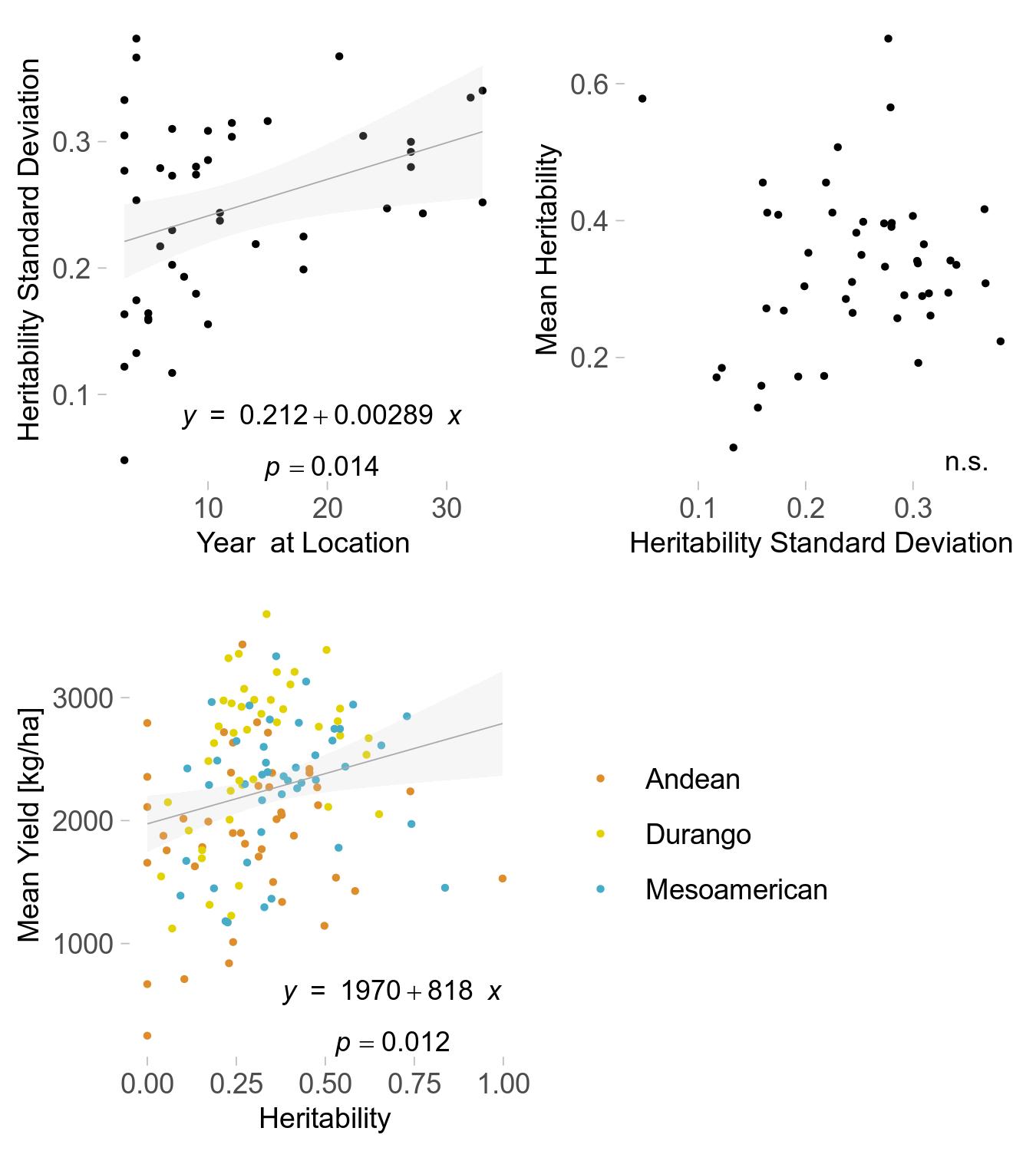
## 

## 

## 

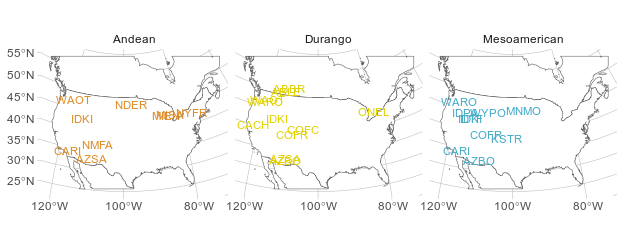
## Figure S3

**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.



## Figure S4

Figure S4. High heritability and high HFA location combinations by bean race.



## 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 | | | |
|  | Andean | Durango | Mesoamerican |
| With Home Field | 3434 | 6772 | 5095 |
| Variety, Site, Year | 3599 | 7323 | 5421 |
| With Home Field, Kinship | 3773 | 7662 | 5703 |
| With Kinship | 3926 | 8143 | 5997 |
| 1Akaike Information Criterion | | | |

## Table S3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 | - | - |