CROP-2021-07-0504-ORA

Response to Reviewer comments

Editor

Thank you for the opportunity to revise the manuscript and especially thank you to the reviewers for their deep reading and thoughtful comments. The suggestions from you and the reviewers have made the critical points clearer and raised worthwhile questions for further study. We hope that our revisions meet your expectations.

**Comment:** Please consider the comments by reviewers especially those concerning presentation and analysis of genotypic data, how absolute yield can be interpreted in relation to conclusions about HFA, and the division of HFA v. Heritability into quadrant

**Response:** Thank you. We have responded to all comments and carefully clarified the presentation of genotypic data, the relationship between yield and HFA as presented and their meanings for adaptation; and the definition and interpretation of quadrats in Figure 4.

Reviewer 1

**Comment 1:** This is a fascinating paper that reviews results in the medium-to-long term of 35 years of a regional common bean yield nursery in North America.

**Response:** Thank you

**Comment 2:** Line 177 presents the standard statistical model that is expanded in 4 models. Might the authors present statistical models for the 4 models used, in continuation of line 177, so that the reader can visualize the contrasts and progressive evolution of the models?

**Response:** Thank you. We added an additional table, now Table S1, which summarizes this paragraph.

**Comment 3:** The distribution of sites with useful HFA: Referring to Figure 4B, the frequency of useful HFA is much greater in the western USA that it is to the east of the Rocky Mountains, although many sites were represented in the upper Midwest (Figure 1). One notes that three of the most most active breeding programs are found to the east of the Rockies in Michigan, Nebraska, and N Dakota. One suspects that in this period of 1981-2015, a great number of the entries emerged from these programs. Did this lead to wider adaptation and lower HFA across the American Midwest, as opposed to the dryer western USA where local adaptation appears to play a greater role? Or is there some other explanation for the distribution of HFA?

**Response:** This is an interesting observation and a feasible interpretation. We had noticed that the distribution of good breeding sites loosely follows the specialization of breeding programs detailed in Vandermark et al (2014): Durango lines traditionally have been the focus of the arid northwest, kidney (Nueva Granada) of the northeast. Additionally, good breeding sites as defined here also need high heritability (notwithstanding your comments) - which is higher in general in the northwest. This may be due to greater variation in growing conditions, for example, disease pressure, in the Midwest, combined with irrigated and highly uniform field conditions (laser leveled, etc) in the Pacific Northwest. A follow-up analysis comparing HFA and heritability across breeding programs and/or combining this with environmental data would be a worthwhile subsequent paper. We included this in the discussion (ca. Line 355).

**Comment 4:** In the use of maps in the manuscript, would it be worthwhile to include a map showing the sites of programs that proposed lines, and the respective numbers of lines in the CBDN? In line with the above, is it possible to separate analysis by sites that figured in the selection process of lines, from those sites that played no role in the development of lines? Sites that were employed in the development of lines would be a different perspective on what a “home site” represents. Would this reveal a contrast in HFA? Would sites that participated in the line development process show greater HFA?

**Response:** Thank you for these observations and worthwhile questions. The metadata for the lines in the CDBN, including the source breeder’s institution for each line, is in Supplementary Table 2 of the MacQueen et al 2020 paper, at: <https://doi.org/10.18738/T8/KZFZ6K>.

The question of HFA-home vs breeding home is interesting. We have this information for ~80% of the entries in the panel, and indeed, we would expect source breeders to focus on developing lines suited to their regional environments. We considered including this, but ultimately determined it was best suited for a follow-up paper. All sites were included in the selection of lines once they entered the CDBN, but to varying degrees that would be difficult to quantify - especially for earlier entries, where in many cases we do not have a source breeder or institution.

**Comment 5:**  Race Nueva Granada. I think that it is an overstatement to say that race Nueva Granada showed “little evidence of local adaptation” (lines 343-4). Per cent contribution to yield variance was less but similar to other races (Table 2). Figure 4 shows 8 sites with useful HFA for Nueva Granada, 10 for Mesoamerica and 11 for Durango. Nueva Granada may present less instance of HFA but it is only a little less frequent that the other races.

**Response:** Thank you for this astute observation. The HFA metric applied at the population level is (almost) always positive and “statistically significant” using traditional metrics (eg ANOVA), including for completely random data. We clarified this in the text (Line 197). The stated lack of evidence for local adaptation is based on a permutation test, which was not significant for Nueva Granada lines - i.e. while we saw high HFA with Nueva Granada, it wasn’t higher than expected by chance. The sites in figure 4 both confer high HFA and have high heritability, thus counting sites isn’t necessarily evidence for HFA’s importance. See also Reviewer 2’s second comment.

**Comment 6:**  “Yield heritabilities were highest in the northwestern United States and southwestern Canada and decreased along a south-eastern gradient” (line 277). Could northern latitudes and high near-desert temperatures induce a photoperiod response in some materials that would widen phenotypic response and thereby increase heritability?

**Response:** We also are wondering about the cause of this heritability gradient - and whether it is due to greater phenotypic variation or lower environmental variation. Earlier CDBN entries were sensitive to photoperiod, though not to the extent of, for example, soybean (White et al, 1989). Photoperiod sensitivity is believed to have been bred out of modern entries, but not formally studied. Additional, highly likely possibilities are uniform environmental conditions of low disease pressure, irrigation, and laser-leveled fields in the semi-arid northwest, and greater disease pressure in more humid regions. Please also see comment 3 above.

* White, J.W., Laing, D.R., 1989. Photoperiod response of flowering in diverse genotypes of common bean (Phaseolus vulgaris). Field Crops Research 22, 113–128.

**Comment 7:** There is an inherent conundrum in employing the concept of heritability. We almost universally assume that high heritability is good, and low heritability is bad. This is the case when a breeder is applying selection. But when heritability is used as a measure of genetic

variability (which seems to be its primary role in this manuscript), one must first have confidence

that the genetic variability is useful and is not driven by inclusion of inferior materials;

heritability can be a two-edged sword. Since the article deals with a nursery of elite lines and

varieties, we assume that heritability is not the result of the presence of very bad materials in

terms of yield. On the other hand, to what extent is low heritability necessarily bad? ...and bad for what?

**Response:** Thank you for this important comment. In this case the low heritability is an indicator of a particular site not being useful for selection, so low heritability in this case is more of a statement on the usefulness of the site for inclusion in a broad set of trials. Low heritability sites tended to be lower-yielding in general (Fig S3b), suggesting lower phenotypic variation in CDBN entries at those sites. It also is related to the spatial extent of the potential breeding program: Multiple, close sites with high heritability might enable efficient selection, but will not necessarily be high yielding sites or produce varieties that yield well at other sites.

**Comment 8:** The manuscript cites a case of the Mesoamerican race that “showed the largest improvements in yield across the study period, at the expense of local adaptation and heritability.” (line 366) This race presented the highest gain per year (20 kg/ha; line 235) while HFA decreased (line 272) but was still important (line 262). The text explains well what could have happened, and how selection for specific traits may have resulted in gain but narrowed the genetic base and reduced heritability. It is a question of perspective whether this is a problem or a sign of success in attaining broad adaptation in new materials. The authors seem to conclude that it is both: “...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” (line 407), these goals apparently being to identify stable genotypes. This is short term payoff but may be long term risk. It is an open question whether the genes that conferred stability are accompanied by genes that imply a genetic risk, such as susceptibility to a disease or to changing climate. The more important implication may be that genetic variability for yield per se is being reduced, limiting future gain. Given that the entries in the CDBN are drawn from multiple breeding programs, the fact that materials do not show more local adaptation and variability could indicate that multiple programs are using similar genetic variability. In this regard, understanding the source of materials (as in point 2 above) could be informative. The analysis presented here can serve as a wake-up call for breeders, but what is critical is what is occurring in breeding programs where crossing and introgression are creating lines for future trials. Can anything be said about recent developments in introgression, based on formal or informal (BIC reports) literature?

**Response:** Thank you for this comment, which strikes at the heart of the question we are addressing. Most CDBN breeders have emphasized that their top priority is breeding for their region, with broad adaptation being an added benefit - a couple do try to generally improve bean production and/or pre-breed. Per loss of potential for long-term yield gain, this is a common concern that our results do suggest - as do a number of other papers especially related to the inclusion of wild relatives and landraces making it imperative to both improve populations but maintain genetic variance (Gepts, 2004; Smith et al.. 2015; Cobb et al., 2019). At the same time, yields do continue to increase, there still appears to be genetic gains to be made, and genetic variance to select from.

Combining ability certainly may explain some of the results. Crosses between North Dakota and Michigan State material usually are successful due to similar environments, but crossing from those programs with, for example, Nebraska tends to produce worse results. The program source of entries might be a good way to further interrogate these results combined with knowledge of combining ability among programs; unfortunately, that data is not always available, particularly for the earlier years. See Supplementary Table 2 of MacQueen et al. (2020). We also want to emphasize that the focus on population-level local adaptation masks large variation in home field advantage among entries (new Figure S1). Further, we emphasize recent work suggesting that continued introgression has helped maintain genetic diversity in common bean, especially between genepool introgression. This may be helpful in maintaining genetic variance and continued yield gains (Chacón-Sánchez et al., 2021; Lobaton et al., 2018).

We have enriched various places in the discussion with these responses with a particular eye towards operationalizing the interpretation of the analysis.

* Smith, S., Bubeck, D., Nelson, B., Stanek, J., & Gerke, J. (2015). Genetic diversity and modern plant breeding. In Genetic diversity and erosion in plants (pp. 55-88). Springer, Cham.
* Gepts, P. (2004). Crop domestication as a long-term selection experiment. Plant breeding reviews, 24(2), 1-44.
* Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., et al. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder’s equation. Theoretical and applied genetics, 132(3), 627-645
* Chacón-Sánchez, M. I., Martínez-Castillo, J., Duitama, J., & Debouck, D. G. (2021). Gene flow in Phaseolus beans and its role as a plausible driver of ecological fitness and expansion of cultigens. Frontiers in Ecology and Evolution, 9, 312.
* David Lobaton, J., Miller, T., Gil, J., Ariza, D., Fernanado de la Hoz, J., Soler, A., .et al.  (2018). Resequencing of common bean identifies regions of inter-gene pool introgression and provides comprehensive resources for molecular breeding. The plant genome, 11(2), 1-2.

**Comment 9:** These considerations about heritability may raise questions about statements such as, race NuevaGranada shows “increasing adaptation at a continental scale, in that yield heritability increased in this race over the 35-year study period.” (line 346) It could be better to say that “both yield and yield heritability increased...” as increasing yield would be better indicator of continental adaptation. A combined increase of yield and yield heritability would seem to be the optimal situation.

**Response:** Thank you for this revision.

**Comment:**   Line 63, remove “across”

**Response:** Thank you.

**Comment:**   Line 67, capitalize “central”

**Response:** Thank you

**Comment:**   Line 113 and elsewhere. I am not familiar with the use the term “specialists” to describe locally adapted materials. Is this standard vocabulary?

**Response:** Thank you for this point. We agree that specialization and local adaptation are closely related but non-identical concepts. We use specialization as a short-hand for locally adapted materials assuming that areas closer together have similar habitat characteristics. Clarified at line 113.

**Comment:**  Line 134 cites 71 environments, lines 34 and 298 cite 70

**Response:** Corrected to 71, thank you.

**Comment:**   Line 267. Should this not read “the Mesoamerican and Durango races”?

**Response:** Thank you. Corrected as: Preserved in Durango, lost, in Mesoamerican, and absent in Nueva Granada per Figure 2b.

Reviewer 2

**Comment 1**: The manuscript is well written and the results are clearly presented.

**Response**: Thank you.

**Comment 2**:My main concern regarding the work is that most of the claims in the results and the discussion related to genetic diversity could be much better supported if the genotypic data from McQueen et al. 2020 were reanalized to obtain actual estimates of genetic diversity within the three races. The authors only use these data to calculate kinship values to fit the model. However, the genetic data could be better exploited to actually estimate and compare genetic diversity by year (or time period) for each race. If the number of data points per year / race is low, at least it would be informative to assess the global genetic diversity within races. For example, the authors assume that the durango race has high genetic diversity, based on a citation to Mamidi et al. 2011. However, this may not be true for the specific panel of CDBN.  Based on the results, I would speculate that the durango race has much lower genetic diversity within the panel than the mesomerican race.

**Response:** We were specifically interested in the proportion of genetic variation that controlled yield/fitness variation in the CDBN. However, we agree that population-level genetic diversity statistics such as nucleotide diversity could have changed over the course of the CDBN, and this might affect our results. Since this is elite breeding material and is typically inbred for eight or more generations prior to testing in the CDBN, we do not expect heterozygosity measures to vary significantly in this material. We add information about nucleotide diversity to the manuscript at lines 277-284 and line 386-392.

For all comparisons, we used a consistent set of SNPs present in the entire CDBN in order to compare nucleotide diversity (pi) between the bean races. We analyzed nucleotide diversity for 1.22M SNPs that were present in the CDBN individuals at a MAF > 5% and sequenced in at least 150 CDBN genotypes. Overall, the Durango and Mesoamerican races had very similar nucleotide diversities across these SNPs: median pi for Durango was 0.231 (IQR 0.939 - 0.405; SE 0.013) and 0.233 for Mesoamerican (IQR 0.088 - 0.407; SE 0.016). In contrast, Nueva Granada had a median pi of 0.042 (IQR 0 - 0.239; SE 0.022). We would argue that, in the CDBN, Durango and Mesoamerican nucleotide diversity is functionally equivalent, whereas Nueva Granada nucleotide diversity is much lower.

To obtain a reasonable (>10) sample size within races for pairwise nucleotide diversity estimates, we estimated nucleotide diversity within races for five year windows within the CDBN timeframe. In general, we did not have enough Nueva Granada genotypes within most 5 year intervals to compute nucleotide diversity with any precision. Median nucleotide diversity in the Durango and Mesoamerican races did not differ significantly across windows from 1985 and 2015. In Durango, median pi was lowest in 1985 and 1990 and increased to a ~0.23 by 1995, though this difference was not significant. Mesoamerican nucleotide diversity was ~0.23 from 1985 to 2000 and decreased thereafter, though this difference was also not significant.

**Comment 3:** In particular, the claim on lines 383-385 indicating that introgression of resistance or architecture alleles  is equivalent to a selective sweep needs to be supported by genetic data. Unless this can actually be shown, please indicate that this is a possible explanation of the results but it is not a proven hypothesis.

**Response:** Thank you. We have revised the text to clarify that this is a hypothesis.

**Comment 4:** Regarding the results on HFA for 2015 (lines 257 - 267), the authors highlighted that in 2015 the HFA was higher than expected and claim that this represents 34 and 22 years of (conventional breeding?) yield gains. However, the absolute yield looks lower than the expectation by the linear regression. Hence, my conclusion would be that HFA could be compensating for some other factor that reduced yield in 2015. If that is the case, it would be good to identify factors explaining this behavior. If my conclusion is not right, please consider the trend of absolute yield to improve the interpretation of  the outcomes for this year. I think this is important to clarify because the claim at lines 316-318 of the discussion looks as completely based on this result.

**Response:** We agree. The 34 and 22 years are just the average home field advantage divided by the average annual yield gains, and HFA percentages are based on expected yields due to genetic gain, as you suggest. We have revised this section for clarity.

**Comment 5:** Finally, it is not clear how the x and y cuts of figure 4a are chosen to build the quadrants. It is very important to clearly define and justify the quadrant boundaries if this chart is proposed as a tool for guidance of breeding decisions.

**Response:** We agree. The x and y cuts in Fig 4a are median HFA and heritability values of all sites, thus, top breeding sites have higher heritability than most and confer higher HFA than most. In general, we view this chart as an illustration of Table 1 rather than a decision support tool. Clear definitions of what is “high” HFA or heritability are best determined within each breeding project based on project goals, germplasm, and budget. A number of formal frameworks can help (ex. Hastie and Dawes, 2009; Malczewski and Rinner, 2015). We clarified these points in the discussion and figure caption.

* Hastie, R., & Dawes, R. M. (2009). *Rational choice in an uncertain world: The psychology of judgment and decision making*. Sage Publications.
* Malczewski, J., & Rinner, C. (2015). *Multicriteria decision analysis in geographic information science* (p. 331). New York: Springer.

**Comment:** Since the calculation of HFA was modified from Ewing et al. 2019, I think it would be great to see in a supplementary figure if these is any difference between the distribution of HFA calculated by the Ewing method and the method used in this manuscript, as presented in figure 2B.

**Response:** Thanks for this request. The modified method is preferred when there is rapid turnover in sites that preclude robust estimates of site effects across years, as in this manuscript (they were relatively stable in Ewing et al., 2019). Conceptually, a moderately high average yield for variety at one site used for three years is more believable than a very high yield at another site used for one year year; the modified method ensures the three year site is chosen as home. At the entry level, differences were minor. At the population level BLUP increases precision. The inclusion of kinship was the second modification, and was rejected for reasons of parsimony.

This is now clarified in the methods section (lines 174-179) and in a new Figure S1.

**Comment:** The variable rep is not explained in the equation for heritability. For non expert readers it may be great to provide a brief rationale of the formula. For example, it could be explained why the number of environments increases heritability, or why the expected yield value reduces heritability.

**Response:** We calculated the heritability based on an entry-mean basis from “Breeding for Quantitative Traits”, Bernardo, 2020 page 163. Specifically, we need to calculate the variance of the progeny mean which is equal to the variance of an individual plot divided by the number of observations for a group of progeny (rep\*environments). This is why we included rep in the calculation, it is necessary to estimate the entry variance which is needed for the heritability calculation.  By increasing the number of environments, you decrease the variance so it should improve the estimate of heritability. The expected yield should not decrease heritability as heritability is a function of the variance of the trials.

**Comment:** In Figure 2B please show for Nueva Granada the regression and p-value, even if the regression is not significant.

**Response:** Thank you. We have updated the figure.

**Comment:** Line 283-284. The linear relationship between yield and heritability looks more clear for the durango race, compared to the other races. It may be worth to make the analysis separated by race.

**Response:** We originally separated these by race, but found that slopes were not significantly different among races, so reverted to the combined model.