Response to the reviewers for “Genotype-by-Environment interactions and local adaptation shape selection in the United States National Chip Processing Trial”

We would like to thank the reviewers for taking the time to make thoughtful comments on this manuscript that helped us to improve the paper. We have addressed the comments the reviewers made as follows.

To begin, we would like to point out that, when revising the manuscript, we found some small errors in the data introduced when putting these data onto the database. While this had no effect on many of the models used in the study, it did slightly change the results in the regression GWAS models (**Lines 344:360**). We have updated both these and their associated figures to reflect the changes made.

**Reviewer #1:**

*The argument for using envGWAS on elite lines is not strong, the major assumption of envGWAS analysis is to leverage thousands of generations of natural selection in wilds, or hundreds of generations of selection in landraces that has occurred in specific geographies. Here you are making individual crosses and using three cycles of selection in a local environment. There is a large chance any relationships are spurious.*

We have added a stronger argument for the use of envGWAS with these data (**Lines 141:153**). Specifically, we would like to point out that these individuals are clonally propagated and undergo an extremely high selection intensity (~3.4) when summed over the course of those three years in a single geographic location. While it is true that local adaptation may take hundreds of generations of selection to be established in a landrace, here we are not making a single year of selection, crossing, and selecting from the progeny of the previous generation. We are selecting the same clonal individuals for three years. We would also note that, because we are selecting clones, any genetic variance for local adaptation due to non-additive genetic effects will be preserved across field years. In the standard equation to calculate response to selection, the narrow-sense heritability is used. Here, to appropriately capture response to selection, the broad-sense heritability should be used, which, for even a complex trait like yield, can be 0.6 or higher in potato. Together, this intense selection in a small geographic range combined with the clonal selection scheme should result in a similar response to selection for local adaptation compared to much longer timescales in wild or landrace populations. *Shouldn't latitude and longitude be covariates, not responses? It seems likely that you will be only getting signals of flowering time for latitude. For longitude what is the interpretation? What is the east-west gradient that you are mapping?*

We included latitude and longitude as responses in the model to test for associations with a more general measurement of the environment than the other variables we tested here. We have added an interpretation of the longitude model (**Lines 271:274**), but to quickly summarize longitude is correlated with other aspects of the environment, e.g., precipitation (precipitation: *r2* = 0.78), elevation (*r2* = -0.51), and soil pH (*r2* = -0.84).

*The markers you have associated with breeding program may not be signals of local adaptation, but signals of traits correlated with the different breeding program targets that you described in the introduction. What genes are in LD with the significant markers? Do the gene annotations have anything to do with the climate or geographic variables? Are the annotations associated with production characteristics?*

We have addressed this in the manuscript (which was mistakenly removed from an earlier version!, **Lines 429:438**). Our main takeaway from the program-specific models was that, as you said, we do not know whether the significant markers were due to breeders’ preferences unrelated to the environment, population structure, or true genotype-by-environment interactions. We included this analysis to say that, because we do not see any overlap between the regions of the genome identified in these case-control models and our envGWAS or regression GWAS models, we don’t expect breeder choice or population structure within a single program to be causing spurious associations for the latter. This has now been incorporated into the discussion.

We chose not to look at the gene annotations as we only have course-scale mapping resolution using the array. We would not have appropriate evidence to point to specific genes, and it would be far too easy to build a story by picking the genes in LD with the significant SNPs that fit our narrative. *For the environmental association what is the distribution of the significant SNPs, are the alleles that reduce precipitation found in lines bred in droughty environments? Are these differences of in temp and precip meaningful based on the growing regions where the lines occur?*

I am not sure exactly what you mean by the distribution of significant SNPs, but I believe you are asking about whether allele frequency for the significant SNPs is correlated with the environmental variables we used. If so, we would point out that, by definition, that must be true as GWAS evaluates the significance of a SNP by regressing the phenotype (in this case, environmental variables) onto the marker genotype (while accounting for population structure and any other covariates you supply). If a SNP was not found in drier environments more often than in rainy environments, it would not be found to be significant.

As for whether the effect sizes are meaningful, this is of course up for debate and interpretation. We believe that finding a single SNP that accounts for more than 3 cm of precipitation or explains as much as 4.4% of the variance in precipitation during selection is meaningful, but others may disagree with that assessment. It is important to note that these SNPs aren’t acting to change precipitation or temperature, they would indicate a SNP that is more likely to be selected when precipitation or temperature is high and/or low. In that sense, we may expect the estimated effect sizes to be lower as the environments are not 100% consistent over years, which lowers the “heritability” of these “traits.”

*Exploring GxE QTL is quite interesting, this is the most novel part of the paper.*

Thank you for these kind words! *Why are 1 MB, 2MB and 5 MB important what is LD in the population?*

We have included estimates of LD in US domesticated potato, which generally range from less than 1 Mb to ~5 Mb to reach an *r*2 of < 0.1 (**Lines 311:314**). We chose a conservative estimate considering that this population spans a very diverse set of genotypes over many years. We highlighted those distances to demonstrate that these SNPs are likely unique within each test.  *The logic of your model is a bit difficult to follow, you select for few generations and then don't find local adaptation, it seems that this is actually your expectation not support for an alternative hypothesis. The main finding of the envGWAS seems to be to confirm that the assumption that it does not work in elite material is correct.*

Our interpretation of the results is that we *do* find evidence of local adaptation from the significantly associated markers in the envGWAS models. While we do not find evidence of directional selection acting to increase the frequency of locally adapted alleles over time in our GPSM models, our results suggest that breeders are repeatedly making selections for alleles that are locally adapted even as these alleles are reduced in frequency in the population in subsequent multi-environment trialing. *The discussion seems to contradict the intro, in the discussion you talk about potato being bred for broad adaptation but in the intro you talk about how potato breeding is regional and therefore you are likely to see local adaptation, which is it?*

It is both. Potato is bred with the *intention* of growing it across diverse sets of environments (generally, a variety is only considered worthy of commercial production if it performs well across either the entire northern and/or southern US), but early selection is structured in such a way that breeders are likely *unintentionally* selecting for local adaptation, even if the alleles underlying it are purged during subsequent trialing and selection. Our suggestion is that breeders make selections which produce genotypes that perform quite well where they were selected and then throw away any locally high-performing lines that do not perform well nationally, even if those lines fit all the necessary criteria to be implemented as ecotypes. While this type of selection scheme may make sense in other crops, we suggest that potato might benefit from moving to ecotype development where breeders can make more rapid progress within their specific environments for two main reasons: breeding progress is extremely slow in potato and throwing away any genetic gain that is being made is likely not worth potential savings in the long run and the breeding and seed pipelines are already regionally implemented, which would make a transition to ecotypes relatively easy compared to other crops. We think this may especially be the case considering the relative importance of GxE for agronomic traits like yield and specific gravity and especially for quality traits in fresh market varieties.**Reviewer #2:**

*The authors utilize a rich data set of phenotypic and genotypic data to investigate local adaptation, environmental variables, and selection. The paper is very well written and does a good job of applying ecological / population genetics principles to a standard plant breeding trial with entries from multiple programs. This trial, and indeed potato breeding in the US, is unique and offers an excellent opportunity to study a fundamental plant breeding question: local versus broad adaptation. The authors use novel methods to generate phenotypes for GWAS and discuss the benefits and limitations of these approaches.*

Thank you for these kind words!

*To provide a better context to the study, it would be good to know the relative importance of selecting for yield (especially total yield) in the early stages of a potato breeding program. In general, selection in the first field years is against negative traits and for high heritable traits in mostly unreplicated trials. Do the breeders submit clones to the NCPT based on a standard criteria or are these decisions made differently by each breeder?*

We have made this section of the analysis clearer in the manuscript. To briefly summarize, the only analysis that included yield data was the regression GWAS analyses, which is a separate analysis from the local adaptation. In the regression GWAS, the goal is to find marker alleles associated with response to environment across the trial location, but doesn’t take into account which breeding program those varieties originated from. In short, the relative importance of yield in those early selection years shouldn’t affect whether we can find evidence of alleles underlying GxE for yield. Generally though, the selection criteria for material submitted to the NCPT is supposed to meet a standard minimum criterion for several different traits, including yield, specific gravity, skin and flesh color, and shape and size. This is not strictly enforced, and breeders may opt to submit material they believe to be promising even if it doesn’t meet one of those standards.

*Is it possible that some breeders put more emphasis on traits other than yield than other breeders? I think the results from the selection site associated markers should address other traits that could explain the differences in material from different programs in addition selection for local adaptation.*

This is quite true! We have included a section in the conclusion on breeder preference and population structure being confounded with the selection site associated markers (**Lines 429:438**). I am hesitant to outright suggest potential traits that might be underlying these significant associations as we have no evidence to back those suggestions up.

*I think that this could be approached by looking at differences in performance for other traits by program and by comparing these regions with QTL regions for traits that are important for chip varieties in addition to total yield.*

The main problem with implementing this approach is that yield is one of the only high quality, quantitative traits that is taken during these trials (many other traits are 1-5 scores). While we would love to do this analysis (I actually think it would be a beautiful way to test what effect breeder preference has) it isn’t possible to do with these data. We do know that many traits *are* being selected in a program-specific fashion, particularly disease/pest resistance. However, in many instances we would argue that those traits are also indicative of selection on GxE, as disease/pest pressure varies across environments (and some pests/diseases are unique to certain programs e.g., golden nematode is only a real pressure in New York).

*Similarly, for the year the clones entered the NCPT, have the selection criteria changed over time?*

The selection criteria have remained relatively stable over time, though some small changes to the relative importance of disease resistance have changed over the years.

*The seemingly large QTL peaks on chromosomes 2 and 12 are worthy of more discussion. This Manhattan plot is noisy, possibly indicating a problem with this "trait". As programs differ in the turnover of parents, perhaps generation would be better measured by the cross or release year of the parents.*

Unfortunately, potato is unique in how parents are chosen compared to many other crops. In potato, the parents are often not from the same “generation” or cross/release year. While we could potentially use the average release year of the parents, this could potentially introduce another level of complexity as this average would ignore the year the grandparents were released. Here, all the potato included in a generation were initially crossed and selected in the same years, which is the simplest solution to that complex problem. We have clarified this issue in the text.

*One minor comment is about the use of yield (kg per plot) in Figure 1 instead of standardized yield per area to account for the differences in plot sizes between sites.*

We have added a second panel to figure 1 that shows yield per area and also made a specific note in the caption that the plot sizes and number of plants are not standard across trial locations. We decided to emphasize that neither total yield nor yield per area are perfect metrics considering the planting density is different between programs but also the number of plants is different in CA.

*Overall, this is an interesting study that will stimulate potato breeders to think differently about their breeding targets.*

Thank you!