

Testing for the genetic basis of cold tolerance in a wild mustard

Student: Will DeVries

Mentors: Thomas Mitchell-Olds, Lauren Carley, and Emily Chan

Part-time Independent Researcher
Summer 2016

Abstract:

Overwinter survival and cold tolerance are complex quantitative traits that are important in understanding adaptation to different environments. I evaluated 422 genotypes of *Boechnera stricta*, originating from different geographic locations and elevations, in a common garden genome wide association study (GWAS) experiment in the Rocky Mountains. I measured overwinter survival in 3895 GWAS plants that have been fully sequenced as a part of the original experiment. Using sequenced genetic information from the plants, I looked for correlations between survival and genetic variation in a quantitative trait locus previously implicated in cold tolerance adaptation in this species. I found variation in overwinter survival across broad population groups reflecting different genetic and evolutionary histories, as well as in relation to one aspect of genetic variation within the quantitative trait locus specifically. The population groups Utah and West had higher survival rates than those from Colorado in the Colorado common garden experiment. In a test for local adaptation, population group was a significant predictor of overwinter survival, while elevation, latitude and longitude were not. This indicates that the cold tolerance trait is likely not locally adapted.

Introduction:

How plants tolerate freezing is an important question in understanding adaptation to different environments, especially in northern latitudes and/or high elevations. Many plants exhibit adaptation to local environmental factors such as freezing conditions that vary based upon the origin of the plant (Gehan, Park, Gilmour, An, Lee, & Thomashow, 2015). Development of resistance to freezing temperatures in some plants is influenced by photoperiod, temperature and photosynthetic performance (Heo et al., 2014). All plants undergo biochemical changes when acclimating to the cold, but certain genes showing different patterns of regulation under low temperatures have been linked to some essential processes undertaken in response to cold conditions, such as accumulation of carbohydrates for energy storage and adjustment of photosynthesis to produce energy despite the freezing conditions. Despite this, the roles of many other genes influencing

cold tolerance remain unknown.

Understanding the different genes that influence the susceptibility of a plant to cold conditions may help increase crop yield by facilitating selective breeding of plants with increased tolerance of the cold (Heo et al., 2014) and will further our understanding of the genetic basis behind quantitative traits. Quantitative trait locus (QTL) mapping is an important tool for assessing the genetic basis of quantitative traits such as disease resistance, how well a plant is able to use nitrogen fertilizer (Heo et al., 2014), and cold tolerance. Quantitative traits cannot be measured using classic Mendeleian techniques because they display continuous variation as opposed to categorical variation, usually due to multiple genes and interactions controlling the traits (Seymour et al., 2012). Mapping markers linked to QTLs can identify regions of the genome that may contain genes involved in the expression of the quantitative trait such as cold tolerance. The analysis of QTLs can provide valuable insights about the evolution and adaption of plants to a specific environment by looking at the variation within the region under study and how that variation contributes to the expression of the trait of interest.

Boechnera stricta is perennial mustard found in many parts of the Rocky Mountains, where freezing temperatures are guaranteed in the winter. Variation in cold tolerance in *B. stricta* has been shown (Heo et al., 2014). In a previous study, cold tolerance in *Boechnera* was assessed in greenhouse settings as well as an outdoor research plot in Amsterdam, The Netherlands, where the plants were left over the winter of 2011-2012. Damage was assessed by measuring electrolyte leakage, survival, and maximum quantum yield of Photosystem II (Fv/Fm). The experiment looked specifically for QTLs correlated with cold tolerance. Under all field conditions, data from the Fv/Fm measurements identified one QTL on chromosome 7 that was correlated with freezing tolerance (Heo et al., 2014). There is a similar genomic region to this QTL in *Arabidopsis thaliana*, a close relative of *B. stricta*. This syntenic region in *A. thaliana* is thought to be involved in the processes involved in adjusting to the cold (Seymour et al., 2012). The role of fine-scale genetic variation within this region in contributing to survival and performance in natural habitats of *Boechnera* remains unknown.

The QTL region in *B. stricta* can be further studied by combining genomic data from known accessions with survival and performance of these accessions in a common

garden experiment. I hypothesized that genetic variation in the QTL region may be correlated with variation in cold tolerance across different genotypes of *B. stricta*. I assessed patterns in the geographic location, elevation of origin, and survival of 422 distinct genotypes of *B. stricta* after overwintering in Colorado in an experimental garden at 10,300' (3145 m) elevation, and compared these to patterns of genetic variation within the QTL region, which has already been fully sequenced.

Methods:

Study organism:

Boechnera stricta (Brassicaceae) is a short-lived perennial forb native to the Rocky Mountains and found over a large geographic and climatic range in North America (Rushworth, Song, Lee, & Mitchell-Olds, 2010). *B. stricta* is primarily self-pollinating but still maintains a great deal of genetic diversity, presumably caused by adaptation to local environmental differences (Gehan, Park, Gilmour, An, Lee, & Thomashow, 2015). *Boechnera* plants are adapted to cold weather; in fact, they must undergo vernalization, which is the induction of a plant's flowering caused by cold temperatures, to reproduce (Seymour, et al., 2012). However, local adaptations to environmental stress such cold conditions may vary depending on the specific circumstances that different populations of plants have adapted to; some *Boechnera* populations are found in dry, sagebrush-dense plains around 7000 feet above sea level, while others are found in alpine meadows and forest understories up to 11000 feet above sea level.

B. stricta is known to have a region on chromosome 7 that is correlated with cold tolerance in the plant (Seymour, et al., 2012). The presumed phenotypic variation maintained by local adaptations to varying environments may reflect underlying genetic variation in the QTL region on chromosome 7. Because *B.stricta* is used as a model organism to study ecology and evolutionary genetics (Rushworth, Song, Lee, & Mitchell-Olds, 2010), assessing the genetic basis of traits like cold tolerance is experimentally tractable in this species.

Currently, an experiment involving *Boechnera* is under way near the Rocky

Mountain Biological Laboratory in Gothic, CO. The plants in this experiment originated from a variety of geographic locations and elevations all over North America. The genomes of each genotype involved in this project have been fully sequenced, including the QTL region on chromosome 7. This ongoing study presents an opportunity to learn more about the genetic variation for cold tolerance in *Boechea*.

There are 5 population groups in *B. stricta*: Colorado, Utah, North, West, and Admix (Figure 1). These population groups show different genetic history in their genomes and originate from a variety of geographic locations and elevations. The difference in the population groups can be seen on a genetic level using a principal component analysis (Figure 2). The genome contains about 7 million SNPs or variations within a single nucleotide at a position in a genome. This variation can be condensed to about 1 million SNPs by narrowing them down to the common ones. The variation in the one million SNPs can be condensed and shown by genotype in a principal component analysis 3 dimensional scatterplot.

Plant materials:

An experimental common garden of *B. stricta* currently exists in a field site outside of Gothic, CO in a genome-wide association study (GWAS). Experimental plants are planted into natural *B. stricta* habitat in blocks of 200-250 individuals, with each plant having an individual plant ID number (Figure 3 and Figure 4). When the plants are censused, the survival information specific to each plant is paired with its ID number. The genotypes of the plants are paired with their respective ID number during data analysis. The genomes of these plants have already been fully sequenced for each of the 422 genotypes used in this study. These plants originate from a variety of geographic locations including Colorado, Montana, Idaho, and Wyoming as well as across a variety of elevations (Mitchell-Olds et al., unpub.). The plants were grown in a greenhouse at Duke University and transplanted to the common garden north of Scholfield Pass. They were planted in October of 2015 and overwintered through the spring of 2016. The site experienced winter conditions (837 cm of snow and a low temperature of -31° C; CITE) until snow melted on approximately June 10, 2016 (pers. obs.).

Assessment of Winter Survival:

The trait examined in this study is overwinter survival. Early in summer after snowmelt, the plants were censused for survival (alive or dead). Early-season survival data is assumed to specifically reflect overwinter mortality, since these measurements were taken before significant herbivory, drought, or other stressors had occurred. The plants were also censused later in the season to account for missed plants during the first census. If the plants were alive during the first census, but dead at the second, the plants were labeled as alive for the data analysis. If the plants were labeled as dead in the first census, but were found to be alive during the second, they were counted as alive. The combination of the early season census and the late season census gives the most accurate information on whether or not the plants' survival was dependent on the cold.

Data Analysis:

The survival data from both censuses during the summer was used to calculate the least squared means for survival of the different genotypes in the program JMP. We accounted for variation in survival due to location in the garden in the model when we calculated least square means, by including "block" (position) as a random effect in the model. Genotype was a significant factor in survival and accounted for 3.874 percent of the variation in survival (Figure 6). The LS means information was used to look for patterns in the survival of the different genotypes compared to population group, latitude and longitude, elevation, and genetic variation within the QTL on chromosome 7.

Population groups were previously identified using a principal component analysis reflecting broad patterns of variation in evolutionary history (Figure 2) using a principal components analysis of SNP variation across the whole genome. The population group assignments from this PCA were used in my analysis. Specifically within the QTL region on chromosome 7, SNP variation was described using a second principal component analysis (Figure 5). The first 5 axes of variation in this QTL were used to capture genetic variation in this region in further analyses. All genomic analyses were carried out in collaboration with Baosheng Wang, a bioinformatician at Duke.

I used four statistical models in order to interpret the survival data. The first model included all five QTL PCA axes, population group, and elevation. The second

model included latitude, longitude, elevation, and elevation crossed with latitude. During the data analysis, we noticed an outlier genotype from the Colorado population group that skewed the data. In order to see how this affected our results, we performed two more models that compared survival only to PCA 5, population group, and elevation. One of these final models included the outlier, and one model did not.

Results:

When measuring testing how overwinter survival varied in response to the first 5 PCA dimensions along with population group and elevation (Table 1), the fifth dimension (PCA 5) showed a significant effect ($P = 0.004$) as a predictor for overwinter survival while all other PCA axes did not. PCA 5 accounts for 12% of the variance in the QTL region of *B. stricta*'s genome. Population group was found to be significant as well. The 5 population groups differed from each other ($P = 0.0452$) in mean survival, with the Utah and West groups having significantly higher survival rates than the Colorado group (Figure 5). The Utah and west groups had significantly higher survival rates than the Colorado groups, but not the North or Admix groups.

In the second model, overwinter survival was compared to latitude, longitude, elevation, and elevation crossed with latitude. None of these predictors were significantly correlated with survival rate ($P_{\text{Longitude}}=0.639$; $P_{\text{latitude}}=0.8463$; $P_{\text{elevation}}=0.590$) (Table 2).

When examining Figure 6 an apparent outlier that originated from Colorado was discovered. The outlier genotype from Colorado shows a lot of variation on PCA 5 (Figure 5) but also experienced a low survival rate. Focusing in only on the two significant predictors and elevation, a third model was run with the outlier and a fourth was run without the outlier (Table 3 and Table 4). When the outlier was excluded, PCA 5 lost its significance in that particular model ($P = 0.2103$, Table 4), but population group was still a significant predictor of survival ($P = .0425$).

Discussion:

The evidence shown in the previous experiment identifying the location of a possible quantitative trait locus on chromosome 7 was supported by the significance of the QTL's PCA axis 5 on survival; however, this result is variable depending on which model was run. Examining specific nucleotide positions that drive variation on QTL axis 5 could help to determine whether a quantitative trait does exist in that location, and could provide the next step for researching this cold tolerance and overwinter survival in *B. stricta*. While the genetic basis of this trait may not solely exist on this chromosome, evidence from this project and the previous experiment (Heo, et al., 2014) suggest that it is a promising place to start.

Population group was consistently significant in determining cold tolerance. The population groups represent genotypes with the same broad genetic lineage. The significance of genetic lineage on overwinter survival indicates that cold tolerance variation may be affected by genetic history. In a previous experiment, flowering time in *B. stricta* exhibited localized adaptation, with elevation and latitude of origin found to be significant predictors of flowering time. Conversely, cold tolerance varied significantly across population groups and not with geographic origin or elevation. The significance in these population groups indicates that resistance to the cold may be more affected by a plant's ancestry than by any localized adaptation.

Interestingly, in the common garden experiment in western Colorado, the genotypes that originated from around the Colorado area showed a significantly lower survival rate than the populations from the West and Utah groups (Figure 2). Considering that the experiment was conducted in Colorado, the low survival rate of the local populations compared to the foreign ones is unusual. This provides further evidence that the quantitative trait of overwinter survival is not locally adapted, but perhaps instead an artifact of genetic history. It is possible that the genetic history of the Utah and West genotypes is what is causing their higher survival rates. Some physiological mechanism that differs among the population groups may contribute to the variation in survival for these genotypes. Genetic variation originating from the population group differences may

also make the Utah and West genotypes more likely to survive the cold than the Colorado genotypes, but the mechanisms allowing this remain unknown.

A 3-dimensional scatterplot showing the genetic variation for all genotypes on PCA axis 1, PCA axis 2, and PCA axis 5 for SNP variation within the QTL on chromosome 7 shows a group of genotypes that varies on PCA 5 (Figure 5); while most genotypes show little variation on PCA 5, this group clearly separates from the rest. These genotypes originate primarily from the Utah population group and all have similar elevations of origin (~9000 ft). This variance along PCA 5 could reflect genetic variation related to high-elevation adaptation, explaining the higher survival rate that those genotypes experienced. There were many Utah genotypes that did not show the same variation along PCA 5. Whatever underlying SNP variation is causing divergence of these accessions along PCA 5 may be related to overwinter survival, and thus cold tolerance. Individual SNP-by-SNP analysis was not conducted this summer, but may provide further insights into the genetic basis of overwinter survival in the future.

We explored ways to look for individual nucleotide variation in the QTL relating to survival by looking at specific, individual SNPs within the QTL, but the proper tools and time were lacking, however, this as an opportunity for further study regarding cold tolerance in *B. stricta*

In conclusion, *Boechea stricta* was found to exhibit variation in overwinter survival. While population and genetic variation within a putative cold-tolerance QTL displayed interesting results relating to overwinter survival, more precise and controlled work is still required to further our understanding of the specific genes relating to cold tolerance. The evidence that shows that cold tolerance may be affected by genetic history could provide the next step in understanding the genetics behind cold tolerance. The Utah genotypes that exhibited a higher survival rate as well as variation along PCA 5 could also hold answers to the role of variation within the QTL on chromosome 7 on cold tolerance.

Acknowledgements:

I would like to thank the Rocky Mountain Biological Laboratory for facilitating all of the work for this experiment. I would also like to thank Dr. Thomas Mitchell Olds for guiding me through the research process as well as providing me with all the resources I needed to complete the project. I would like to thank Baosheng Wang for preparing and analyzing the genetic data, Nadeesha Perera for the DNA preps used for sequencing, Julius Mojica for preparing the SNP data, the Joint Genome Institute for sequencing the genomes, Kathy Ghattas for preparing the plants for transplant, Baosheng Wang, Julius Mojica, Wes Mitchell, and Cori Reid for planting all the plants, Jennie Reithel, Ian Billick, and Jill Anderson for advice and preparations for the GWAS experiment, and NIH and Duke University for funding the GWAS project. I would like to especially thank Lauren Carley and Emily Chan for guiding through every step of the research process as well as making this project the best experience I could have ever asked for.

Literature Cited:

1. Anderson, J., Lee, C.-R., & Mitchell-Olds, T. (2014). Strong selection genome-wide enhances fitness tradeoffs across environments and episodes of selection . *Evolution* , 68 (1), 16-31.
2. Gehan, M., Park, S., Gilmour, S., An , C., Lee, C.-M., & Thomashow, M. (2015). Natural variation in the C-repeat binding factor cold response pathway correlates with local adaptation of Arabidopsis ecotypes. *The Plant Journal* , 84 (4), 682-693.
3. Heo, J.-Y., Feng, D., Niu, X., Mitchell-Olds, T., van Teinderen, P., Tomes, D., et al. (2014). Identification of Quantitative Trait Loci and a candidate locus for freezing tolerance in controlled and outdoor environments in the overwintering crucifer *Boechera stricta*. *Plant Cell Environ* , 37 (11), 2459-2469.
4. Koenig, D., & Weigel, D. (2015). Beyond the thale: comparative genomics and genetics of Arabidopsis relatives. *Nature Reviews: Genetics* , 16, 285-298.
5. Kuroki, M., Saito, K., Matsuba, S., Yokogami, N., Shimizu, H., Ando, I., et al. (2007). A quantitative trait locus for cold tolerance at the booting stage on rice chromosome 8 . *Thoe Appl Genet* , 115, 593-600.
6. Rushworth, C. A., Song, B.-H., Lee, C.-R., & Mitchell-Olds, T. (2010). *Boechera*, a model system for ecological genomics. *Mol Ecol.* , 20 (23), 4843-4857.
7. Sandve, S., Kosmala, A., Rudi, H., Fjellheim, S., Rapacz, M., Yamanda , T., et al. (2011). Molecular mechanisms underlying frost tolerance in perennial grasses adapted to cold climates. *Plant Science* , 180, 69-77.

8. Seymour, D., Filiault, D., Henry, I., Monson-Miller, J., Ravi, M., Pang, A., et al. (2012). Rapid creation of Arabidopsis doubled haploid lines for quantitative trait locus mapping. *PNAS*, 109 (11), 4227-4232.

Tables:

Source	Nparm	DF	Sum of Sq	F Ratio	Prob > F
pca1	1	1	0.000537	0.3099	0.5782
pca2	1	1	0.000307	0.1773	0.6741
pca3	1	1	0.002173	1.2543	0.2637
pca4	1	1	0.003228	1.8629	0.1734
pca5	1	1	0.022265	12.8494	0.0004
pop	4	4	0.017098	2.4669	0.0452
Elev_m2	1	1	0.000777	0.4483	0.5037

Table 1: Results from first mixed model test

Table shows results from the first mixed model run in the program JMP. In this model, survival rates for the different genotypes were compared to principal component axes as well as population group and elevation of origin. PCA 5 had a significant effect on survival (P value = 0.0004). Population group also had a marginally significant effect on survival (P value = 0.0452).

Source	Nparm	DF	Sum of Sq	F Ratio	Prob > F
Latitude	1	1	0.000132	0.0715	0.7894
Longitude	1	1	0.004131	2.2346	0.1361
Elev_m2	1	1	0.000736	0.3979	0.5287
Latitude*Elev_m2	1	1	0.001176	0.636	0.4259

Table 2: Results from second mixed model test

Table shows results from the second mixed model run in the program JMP. In this model, survival rates for the different genotypes were compared to latitude, longitude, elevation, and latitude crossed with elevation. In this model, no factors were significant (P > 0.05)

Source	Nparm	DF	Sum of Sq	F Ratio	Prob > F
pop	4	4	0.021623	3.1197	0.0156
pca5	1	1	0.02296	13.2509	0.0003
Elev_m2	1	1	0.000525	0.303	0.5824

Table 3: Final mixed model conducted including the outlier Colorado genotype. In the model, population group and PCA 5 remain significant.

Source	Nparm	DF	Sum of Sq	F Ratio	Prob > F
pop	4	4	0.017252	2.5057	0.0425
pca5	1	1	0.002714	1.5767	0.2103
Elev_m2	1	1	0.000725	0.421	0.517

Table 4: Final mixed model conducted excluding the outlier Colorado genotype. In the model, population group remain significant, but PCA 5 loses its significance.

Figures:

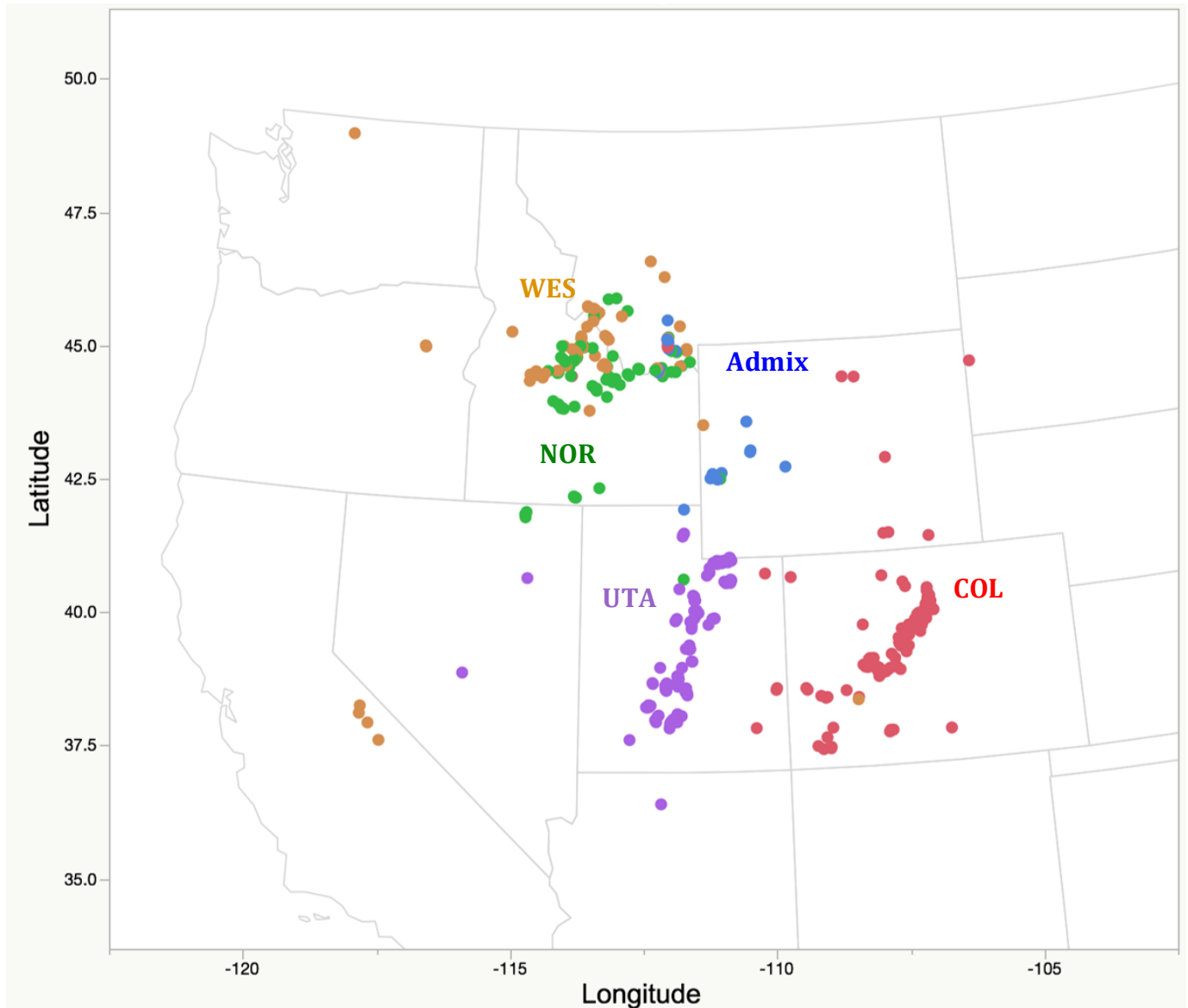


Figure 1: Geographic origins of the population groups. Western US Map of the origins of all population groups used in the Scholfield garden experiment. Colors represent the different population groups to which the genotypes belong. Locations are shown by latitude and longitude.

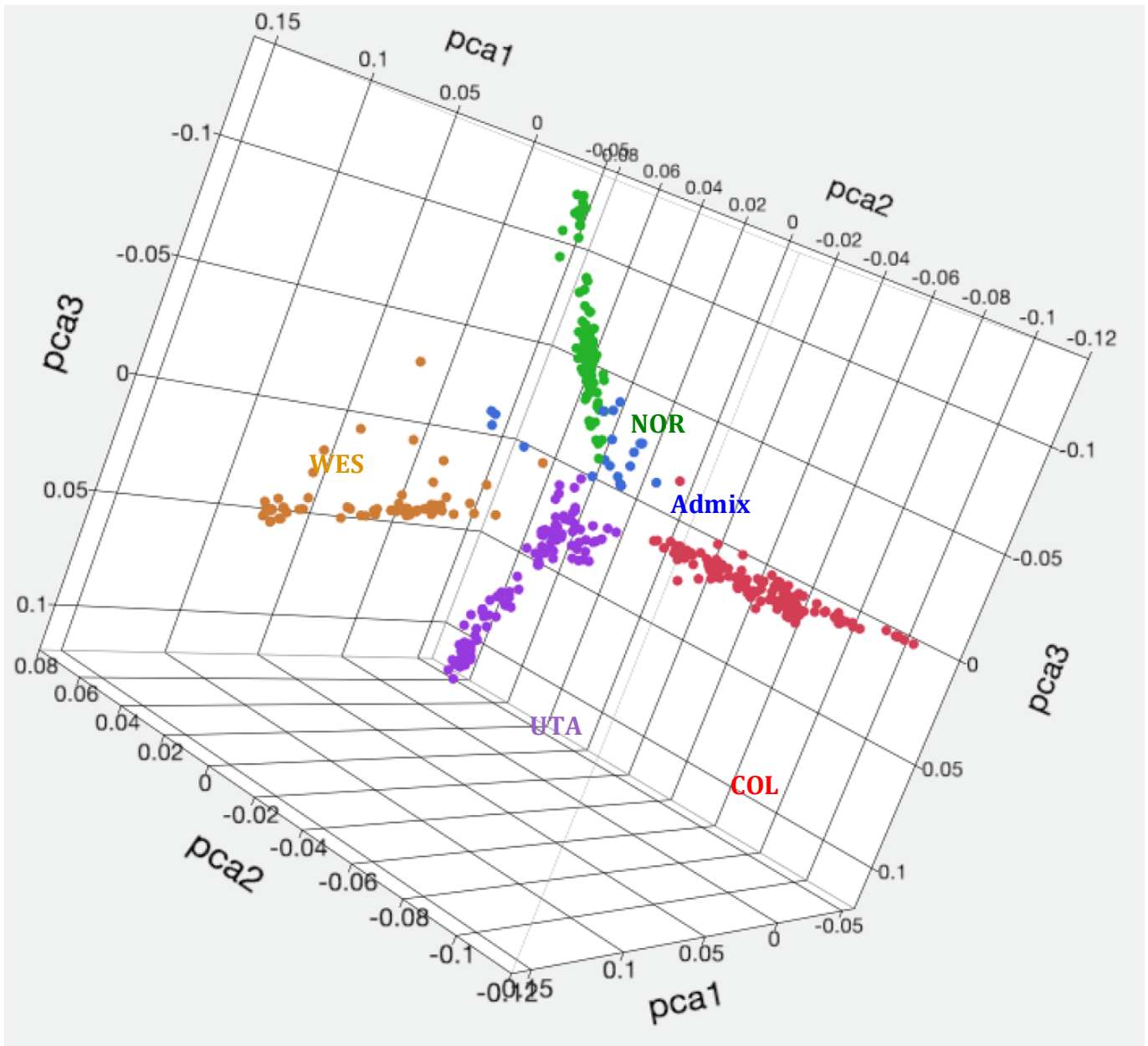


Figure 2: Principal Component Analysis for whole genome of *Boechera stricta*.

Figure is 5 dimensions in reality, but shown in three for simplicity. The axes represent variation among the various principal components. Population groups are generally grouped together. Variation between these groups is caused by genetic history and not necessarily adaptation on a local scale.



Figure 3: Scholfield Garden satellite map

This map shows the geographic location of the Scholfield site. The satellite image was taken on 10/12/2015. The coordinates of the garden are 39°02'07.80" N, 107°03'34.83" W. The elevation of the site is 10321 ft. The inset depicts the orientation of all the planting blocks, and is enlarged and shown in greater detail in Figure 4.

1-50	51-100	101-150	151-200			3705-3754	3755-3804	3805-3853	3864-3895
201-250	251-300	301-350	351-400			3905-3944	4390-4409	X	X
401-450	451-500	501-550	551-600			4190-4239	4240-4289	4290-4339	4340-4389
601-650	651-700	701-750	751-800						
801-850	851-900	901-950	951-1000						
1001-1050	1051-1100	1101-1150	1151-1200						
1201-1250	1251-1300	1301-1350	1351-1400						
1-50	51-100	101-150	151-200						
201-250	251-300	301-343	X						
677-726	727-776	777-826	827-876						
877-926	927-976	977-1005	1624-1643				CFR		
1324-1373	1374-1423	1424-1473	1474-1521				GWAS		
1524-1573	1574-1623	1953-2002	2003-2052						
2053-2102	2103-2152	2153-2202	2202-2252						
1401-1450	1451-1500	X	X						
2549-2598	2599-2648	2649-2698	2699-2748						
2749-2798	2799-2848	2849-2858	X						
3162-3211	3212-3261	3262-3311	3312-3361						
3362-3411	3412-3431	X	X						



Figure 4: Map of the Scholfield site garden

Map shows location of each block and the corresponding plant ID numbers for the GWAS and CFR experiment. The GWAS plants (shown in blue) are the ones used in this project.

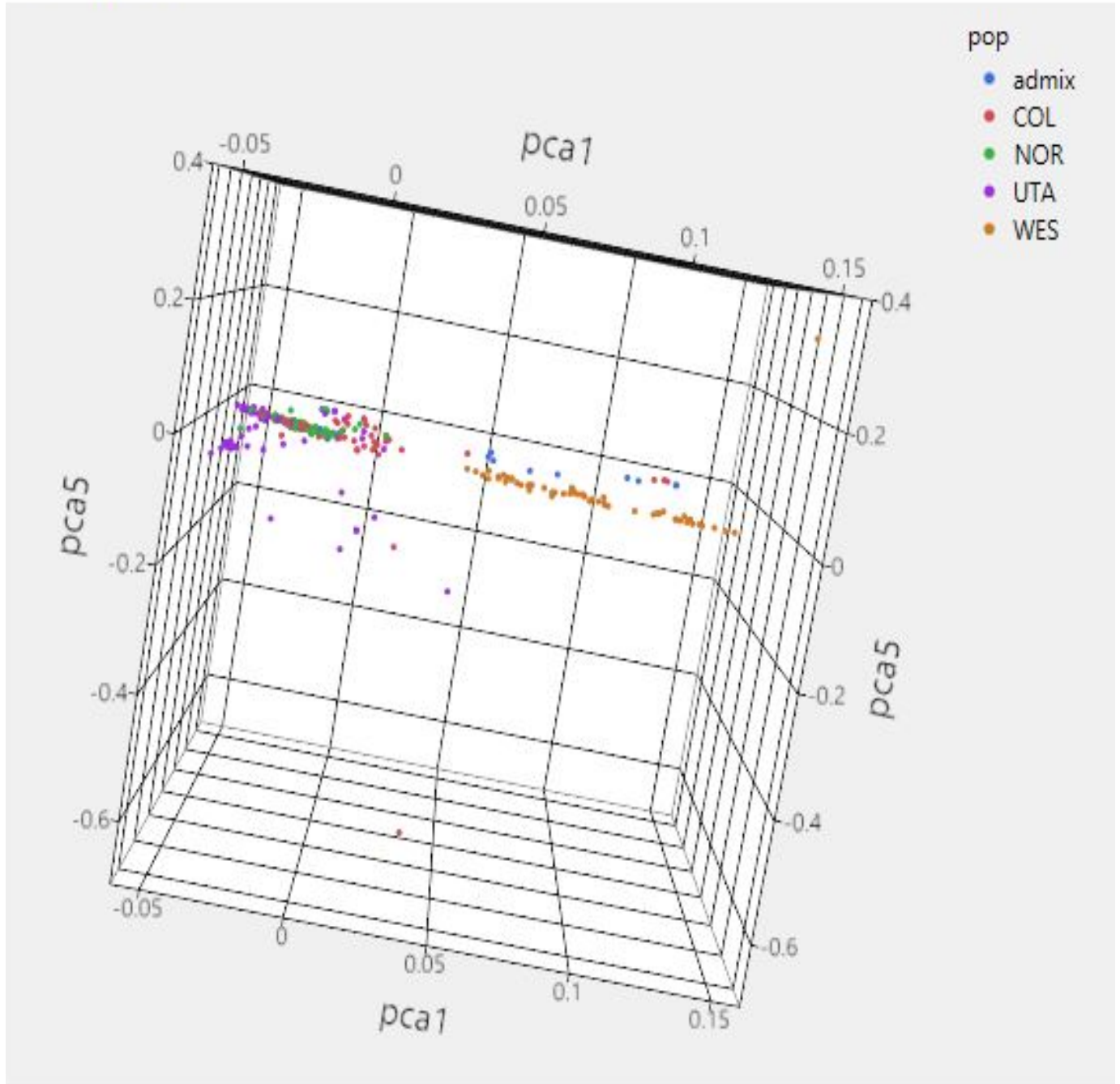


Figure 5: Principal component analysis for the linkage group 7 of *Boechnera stricta*. PCA 1, PCA 2, and PCA 5 are represented in this 3D scatterplot. Several outliers on the PCA 5 axis likely contribute to the significance of PCA 5 in overwinter survival. The outlier group is composed primarily of high-elevation genotypes from the Utah population group.

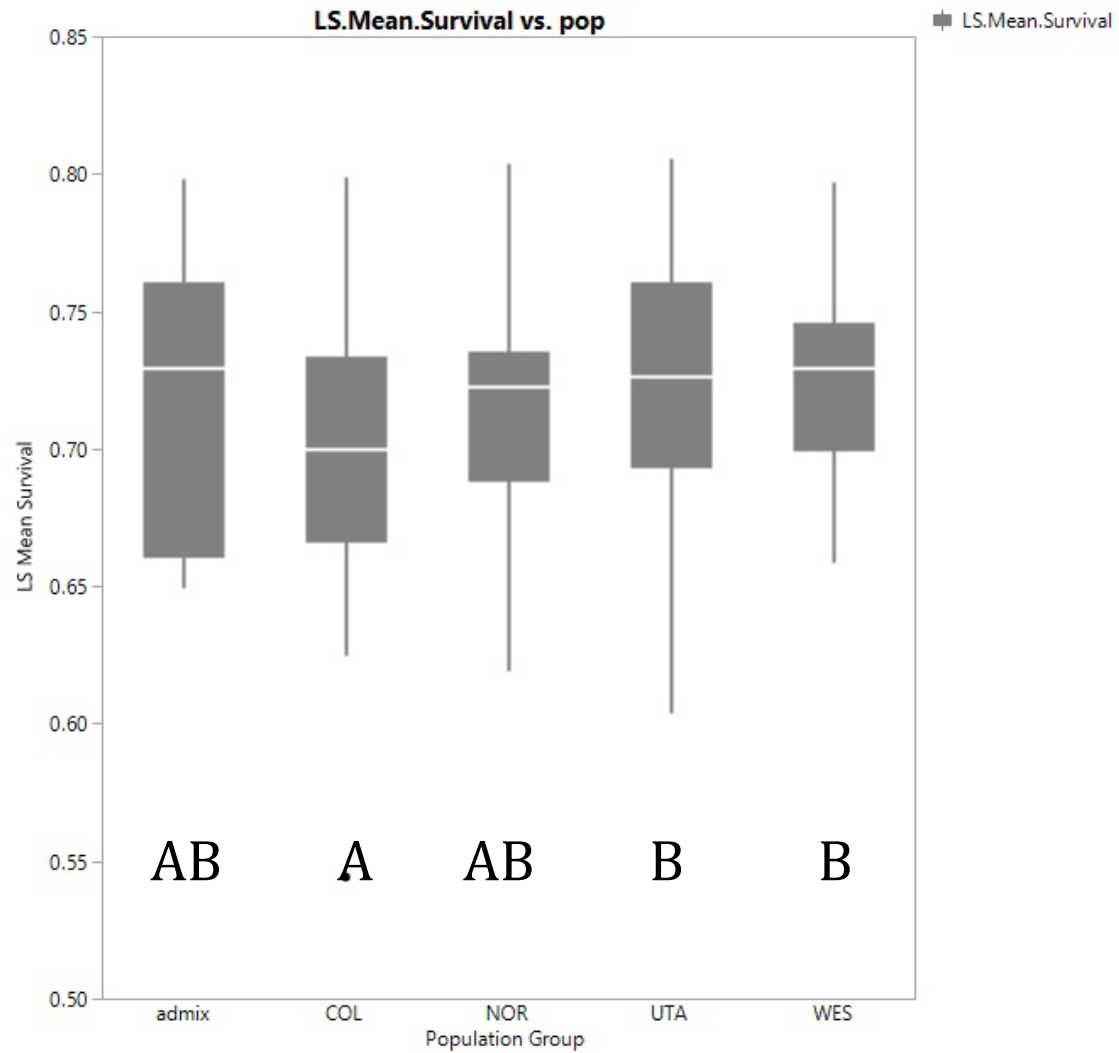


Figure 6: Survival least squared means for each population group. Least square means for each population group in the Scholfield Garden. The west and Utah population group showed significantly higher survival than the Colorado population group. Can you mark the significant differences on the figure?