INVESTIGATING THE GENETIC DIVERSITY OF THE BARLEY COMPOSITE CROSS PARENTS

By

Yongyi Wen

A capstone project submitted for Graduation with University Honors

May 10, 2024

University Honors

University of California, Riverside

APPROVED

Dr. Daniel Koenig

Departmental Affiliation: Botany and Plant Sciences

Dr. Richard Cardullo, Howard H Hays Jr. Chair

University Honors

ABSTRACT

Plants are remarkably able to colonize and succeed in a wide variety of environments. The number and types of genetic changes that contribute to adaptation to different areas of the world remains mostly unclear. In this experiment, I investigated the genetic diversity of the barley composite cross parents that are geographically and genetically diverse. To explore the parents' genetic diversity, I performed genome alignment for each of the parents using long read sequences produced using the Pacific Biosciences HiFi Technology and use the most suitable bioinformatic tools to identify and analyze the structural variations. I identified 366,149 structural variations, and over 60% of them are deletions. We also observe that structural variants are most abundant near the centromere and at the ends of chromosomes. GO enrichment analysis found deleted genes associated with DNA transposition, integration, DNA repair enzymatic activity, and tissue development, which might contribute to the role in adaptation. The deleted plant developmental gene shows high similarity to MAIN and MAIN LIKE protein in Arabidopsis thaliana database, which may have a role in TEs silencing, genome stability, gene expression and correct meristem cell division in the barley genome.

ACKNOWLEDGMENTS

First and foremost, I want to express my appreciation to Dr. Daniel Koenig for invaluable guidance and mentorship throughout this study. I want to thank Dr Keely Brown for her mentorship when I first joined the lab. I want to thank Ruth Sarahi Perez Alfaro, our lab assistant specialist, for providing the high-quality DNA extraction. I would also like to thank members of Koenig and Seymour labs for the inspiration and suggestions they provided during undergrad lab meeting presentations. I appreciate the internship and fellowship opportunities provided by Dynamic Genome program at UCR, and the fellowship opportunity by Undergraduates for careers in agricultural science and biotechnology that has encouraged me to move forward with the research. At the end, I want to acknowledge the researchers and breeders, Harry Halan, Mary Martini, Coit Suneson, and G. A. Wiebe, for their hard work to maintain composite cross experiments over the last century.

TABLE OF CONTENTS

INTRODUCTION

Barley (*Hordeum vulgare L.*) is an important agricultural crop that serves as food, animal feed, and malt for the brewing industry (Visioni 2023). Barley is a nutritious rich food source for humans because of the high dietary level of fiber, minerals, higher protein content, β-glucan, and antioxidant polyphenols. Barley is the only grain crop to contain all eight vitamin E isomers. Vitamin E content and antioxidant capacity in barley are reported to be genotype-dependent (Sakellariou 2023). Barley is phenotypic and genetic diverse, allowing it to adapt to a wide variety of environments. The ability of barley to survive in a large global distribution has made it a model system for understanding crop adaptation in an effort to combat the impacts of climate change, increase agricultural yield and generate nutritious varieties.

Here, I explore genetic variation amongst a collection of barley varieties that founded the composite cross experiments at the beginning of the last century. The composite crosses are long term competition experiments maintained over the last one hundred years (Harlan 1929). CCII experiment began in the 1920s with the intercrossing of 28 selected barley varieties around the world in a half diallel design. Researchers cross the parents with every other parent except at least once in this design. CCV is a composite cross created by funnel design with 30 selected barley varieties as parents. The progeny of these crosses were then used to found a long term competition experiment in Davis, California and Bozeman, Montana. The population was planted in the fall, allowed to compete and harvested at the end of the year. The harvest was then used to continue the experiment the following year. As a result, natural selection could alter the genetic composition of the experiment over evolutionary time.

Previous research has shown that the parents of composite cross are highly diverse. An Illumina resequencing approach identified over 12 million variants segregate between the 28

parents of the CCII population (Landis 2023). Analysis of population structure showed that the parents represent most of the common variation found in barley. This work focused on single nucleotide polymorphisms, but the existence of other types of changes has not been investigated.

Structural variation is a change in the length, copy number, orientation, or location of a DNA sequence longer than 50 bp (Yuan 2021). Because they are bigger in size compared to single nucleotide polymorphism, they have greater influence on proteins and gene expression. There are five different types of structural variation. Deletion, insertion, inversion duplication, and translocation. Figure 1 shows the diagram of each structural variant type. Deletion is the fragment that was removed. Insertion is the sequence of nucleotides added between two adjacent nucleotides in the sequence. Inversion is a continuous nucleotide sequence that is inverted in the same position. Duplication is alteration where the copy number of a given region is greater than the reference sequence. Translocation is chromosome breaks and the fragmented pieces re-attach to the same or different chromosomes.

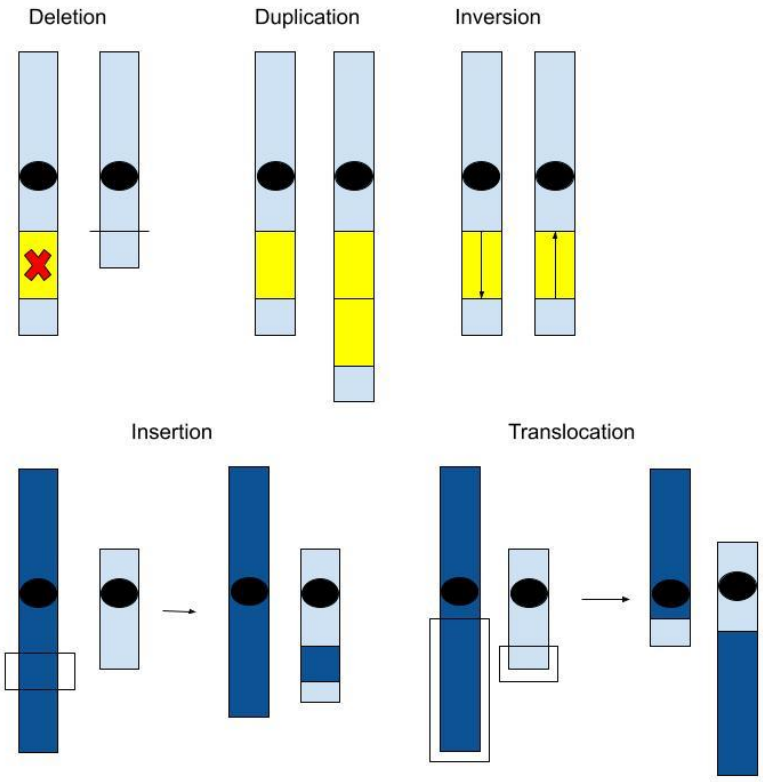


Figure 1

A recent study with barley stimulation data suggests that structural variants can provide valuable insights into phenotype, gene expression and function (Weisweiler 2022). Another study revealed structural variants are related to the degree of boron toxicity resistance in barley. The tolerant individuals are associated with a high transcription of the *Bot1* gene (Sutton 2007). This suggests that by studying structural variation we can better understand changes in transcriptional regulation and their connection to phenotypic variation. Therefore, my research will focus on distribution of structural variation in the composite cross parents.

Accurate genome-wide identification of structural variation has been difficult in the past because of the limited sequencing fragment lengths. Here, we identify structural variants using

medium coverage long read sequencing of barley genomes. My research will be fundamental to understanding how structural variants drive adaptation in the barley composite cross long-term experiment.

The purpose of this project is to identify structural variation genome-wide among the CCII, to explore the genomic distribution of this variation, and to identify the number and types of genes affected by SVs. Structural variants that overlap genes can have influence on gene expression and that may drive changes in phenotype related to adaptation.

MATERIAL AND METHODS

Study Material

The forty-seven barley varieties that founded the CCs were ordered from the USDA national germplasm repository. A single plant was grown from each accession and tissue was harvested for long fragment DNA extraction with lab assistant specialist Ruth Perez Alfaro.

DNA Extraction and DNA Sequencing

To preserve sequence continuity and minimize alignment errors caused by sequence repetition, we choose technology that can provide us with a bigger output fragment, Pacific Biosciences. The lab performs big fragment DNA extraction to obtain the DNA information.

To get the high-quality DNA extraction, the Nanobind plant Nuclei Big DNA Kits from Circulomics are used. The youngest leaves tissue from a plant between 1 and 2 months are collected. During nuclei isolation, plant tissues are grinded with mortar and pestle, or Tissue Lyser with grinding jars for more than 20 minutes, followed by lysing the ground tissue and centrifuge at maximum speed 15 rpm for 15 minutes at room temperature. The lysate is filtered using a funnel and four layers of miracloth that were soaked in buffer NIB beforehand. After centrifuge and discard supernatant, a paintbrush is used to resuspend the nuclei pellet with ice cold buffer NIB and the wash is performed 3 times. During DNA extraction, the resuspend plant nuclei pellet in proteinase K is lysed with buffer Carlson Lysis buffer to form a pellet debris, and the supernatant is transferred to a new tube. Nanobind and isopropanol is added and gently mixing is required. Two washes are required with buffer PW1. Lastly, buffer EB is added to release the DNA. Quality control is performed to ensure the quality of DNA in the extract: DNA quantity>7ug, concentration > 100 ng/ul, 260/280 ratio >1.8-2, 260/230 ratio >2. DNA fragments on the gel should be bigger than 20kb for the 15kb-20kb target DNA shear size. The DNA

samples are sent to Davis California for library construction and sequencing on a Pacific Biosciences Sequel2 instrument.

Genome Alignment and Structural Variation Calling

Each of the 47 sequencing samples was aligned to the Barley Morex 3 reference genome (Mascher 2021) using the software minimap2 (Li 2018). Sniffle was used to identify structural variation, providing a VCF file, which contains genetic variation information with header section, variant data section. Sample quality metrics, per site depth, per site quality, per site missingness, individual missingness, and heterozygosity statistics were extracted using vcftools --site-depth --site-quality--missing-site --missing-indv --het (Danecek 2011). Coverage per site statistic was extracted by bcftools query -f (Danecek 2021). Next, we used Rstudio 4.3.0 (RStudio 2023) and the package, ggplot 2 geom_histogram (Wickham 2016) to visualize the statistic. We filter out the structural variation based on quality score, coverage, missingness, heterozygosity using bcftools view -s (Danecek 2021) with the following parameters: quality (QUAL)<55, heterozygosity (GT="het")>2, missingness (F_MISSING)>0.45, coverage (SUPPORT) >20 and <2. In addition, we filtered out 2 individuals, UCRKL000218 UCRKL000060, from the final VCF file because one had high missingness and one had high heterozygosity respectively.

Structural Variation Analysis

There are 5 structural variants types, DEL as deletion, INS as insertion, DUP as duplication, INV as inversion, and BND as translocation. To visualize the distribution of each structural variant type in the genome and each chromosome, we used geom_bar to create bar plots with x-axis being the type of structural variants and y-axis being the percentage. The distribution of length of each structural type is investigated using geom_histogram. To study the

selection of the allele, we take a look into the minor allele frequency of each SVs type. Minor allele describes the less common allele. Reference allele frequency = 1-(AC/AN), and alternative allele frequency = AC/AN, and minor allele frequency is the smallest value out of the two. We used geom_histogram to visualize the minor allele frequency of all structural variants together and each of them individually.

Spatial Distribution of Structural Variation

This part of analysis is collaborated with Professor Dr. Koenig to see the distribution of structural variants in each chromosome and how they align with centromere. We cut the size of chromosomes into windows and every window is 1 million bp. With the relative size of the genomes and the position of centromere, we can observe the distribution of structural variants on a spatial scale in the chromosome.

Analysis on Gene overlap Deletion Structural Variation

To take the overview of the changes that can affect gene function the most, we further looked into how the fragments overlap with genes and what are the functions of these genes. I used bedtools/2.30.0 coverage (Quinlan 2010) to overlap the deletion structural variation against the gene annotation. However, I observed that the majority of the gene was hundred percent overlap with the deletion structural variant due to the presence of chromosomal sized deletion. In order to proceed the experiment, I validate the quality of genome alignment using samtools coverage (Danecek 2021). All the samples have high coverage except for UCRKL000218 that have coverage lower than 3 which was previously filtered out. Previous study also found structural variants caller can produce chromosome size deletion by fault. Therefore, I filtered out the deletion structural variants that were bigger than 1,000,000 by analyzing the deletion length with logarithmic scale. To continue investigating the deleterious structural variants on gene

expression, I overlapped the deletions with gene annotations. I selected the gene of interest with the gene that has over 75% lost in the genome due to the presence of deletion. I then performed enrichment analysis with ShinnyGO 0.80 using. ShinnyGO outputs enriched functions of genes overlapping the deletions. The software also provides a hierarchical tree of the association of each GO terms.

With the list of genes provided on ShinnyGO 0.80 and the association of each GO term, I further compiled tables that contain gene ID, GO terms for each cluster in the hierarchical tree using Rstudio. I then extracted the nucleotide sequence of each gene within that cluster that contained a deletion. Because of the similarity of barley and rice, with the nucleotide sequence, I blasted the sequence against the rice database (Sakai 2013). The result that has a high bit score will be the potential significant sequence. I also blast the sequence in the *Arabidopsis thaliana* database with the BLASTX: NT query, AA db program. I then recorded the genes with the highest bit score for each hit.

After filtering out the less confident callings, 366,149 structural variations were identified. Deletion is the most observed structural variant, followed by insertion, translocation, duplication and inversion, and each chromosome follows the same pattern. We find out the majority structural variants are most abundant near the beginning and the end of each chromosome which I will discuss in more detail later in the paper. The GO enrichment analysis with the gene that largely overlaps deletion structural variants shows four major overrepresented clustering, Transposon, exonuclease activity, endonuclease activity, and tissue development related gene. All plant development genes that appear during the enrichment to some extent align to Arabidopsis thaliana MAIN and MAIL protein, with an average of 106 bit score.

RESULT

Raw VCF Statistic

Raw VCF statistics per site show us the overall quality of each site from the individual calling. Figure 2 shows the distribution of quality of each site with a maximum possible score of 60. Quality per site is distributed mostly from 50 to 60 for our structural variants. Majority of the calls have very high-quality scores. Figure 3 shows the count of sites of missingness frequency from 0 to 1. When the missingness frequency is 0, there is no missing data for the site. When the missingness frequency is 1, all sites have missing data. The missingness frequency is most abundant around 0-0.5. Figure 4 shows us the distribution of the number of reads supporting the structural variants per site, as known as coverage per site. The coverage is mostly distributed around 0-15.

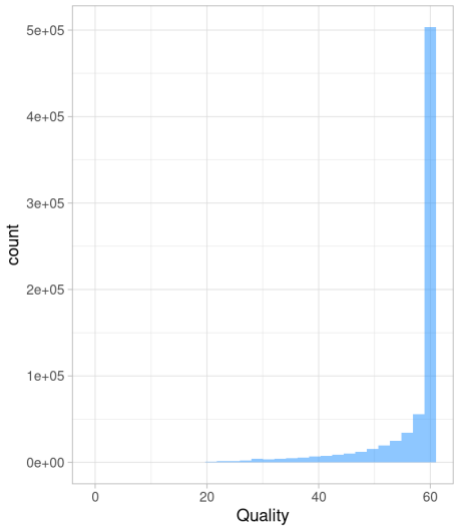


Figure 2

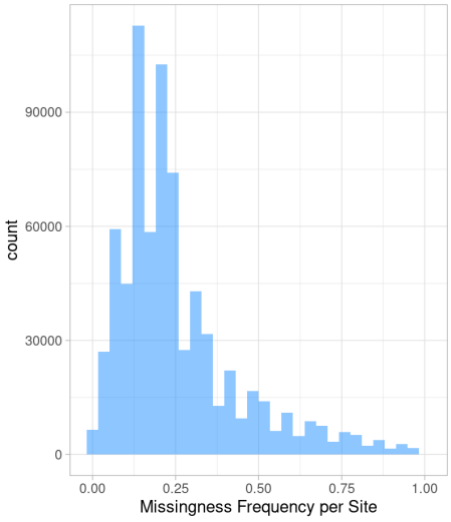


Figure 3

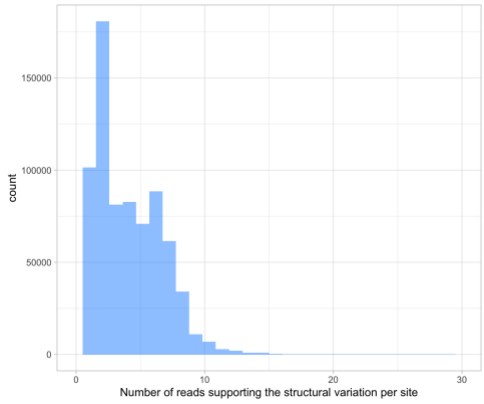
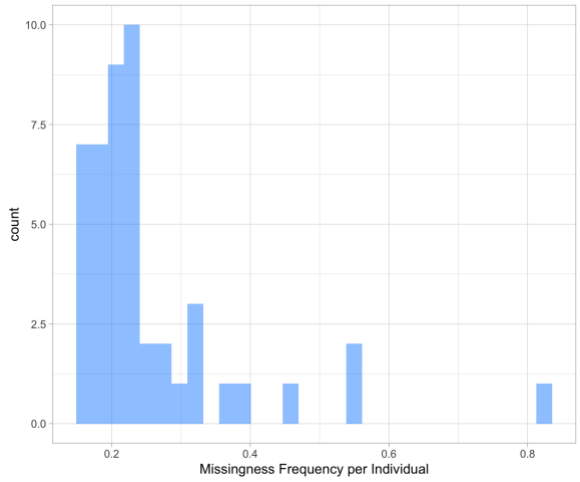


Figure 4

Next, we look at the Raw VCF statistic in the per individual scale. Figure 5 shows the distribution of missingness frequency per individual. There is one individual that has a very high missingness frequency. The individual is UCRKL000218. Figure 6 is the distribution of inbreeding coefficients per individual. Higher inbreeding coefficient means the sample is more pure, and mostly homozygosity. Because barley is a selfing species,low inbreeding coefficient means the sample is not pure and likely contaminated during sample prep. UCRKL000060 has low coefficient and was filtered out.



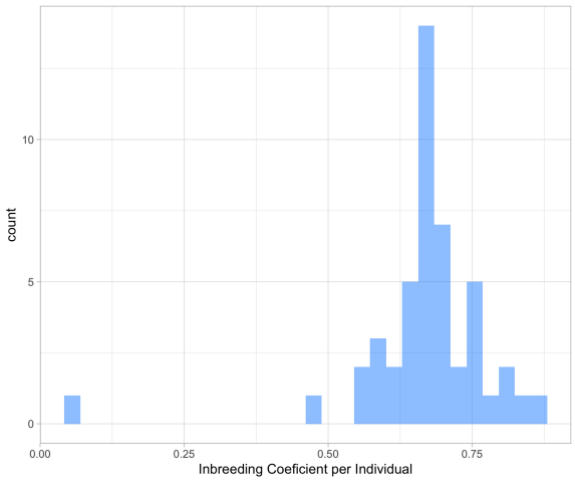


Figure 5

Figure 6

The quality of pac bio long sequence along with reference alignment is relatively confident based on the distribution of quality, coverage, missingness, and inbreeding coefficient. There are 2 out of 47 individuals who need to have further investigation. With the given result, I set up parameters for filtering based on the distribution of the graph. I filtered out sites that have quality score less than 55, count of heterozygosity greater than 2, missingness greater than 0.45, coverage less than 2 and greater than 20. I also filtered out the individual UCRKL000218 and UCRKL000060. After filtering out the less confident callings, 366,149 structural variations were identified.

The analysis of the filtered data shows us the distribution of Structural variants among types. DEL (deletion) INS (insertion) DUP (duplication) INV (inversion). Figure 8 is a barplot of the percentage of each structural variant type in all chromosomes. DEL was most common, followed by INS, DUP and INV. Deletion occupies more than 60% of the structural variants we identified in our experimental population. We observe the same pattern in chromosomes 1 to 7.

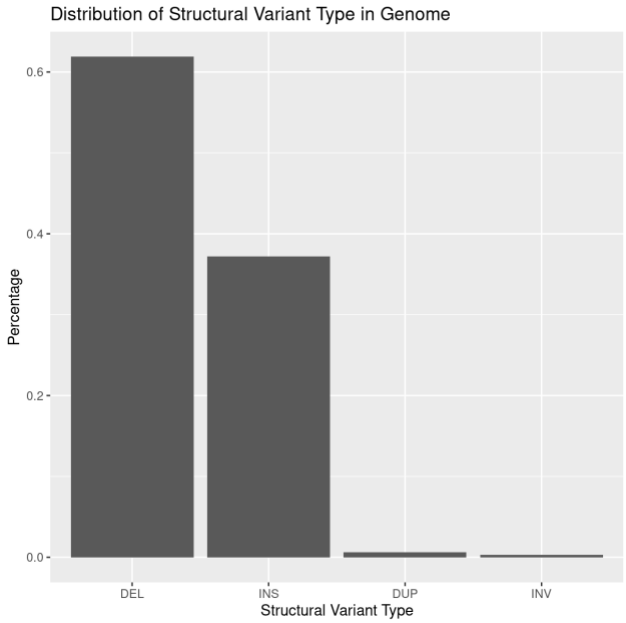


Figure 8

Distribution of Minor Allele Frequency

To understand how common each allele was in the sample, I analyzed the minor allele frequency of all structural variants types. The minor allele is the less common allele in a given population. When the frequency is very small, it suggests that the allele is not favorable in a given population. Figure 9 is a histogram of minor allele frequency across all variants. We can see that there is a peak around 0-0.1. The results suggest that some alleles in the population are unique.

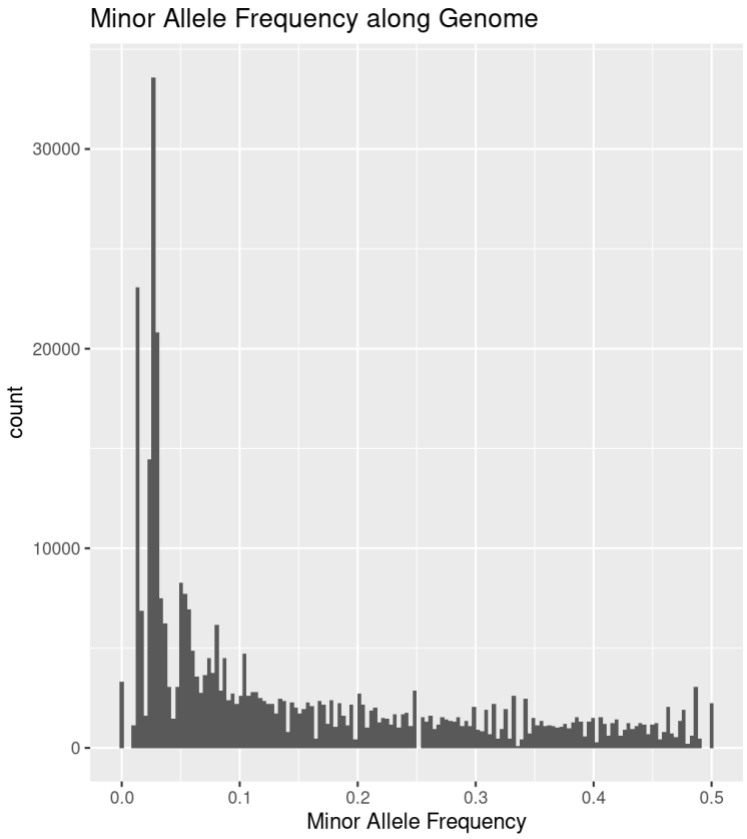


Figure 9

Structural Variants are Abundant Near Centromere and in the Chromosome Arms

We want to investigate the distribution of structural variants in each chromosome and if there is a pattern of them that align with centromere and telomere. This part of analysis is done with Dr. Danial Koenig. Below is a graph with chromosomes 1 to 7. We divided the chromosome into windows of 1 million bp and summed the number of variants we identified in these regions. Figure 10 shows the density of structural variants along chromosomes. The X-axis is the relative size of the genome and the Y-axis is the counts of structural variants in each window. The red line is the position of the centromere. The result shows that structural variants are most abundant near the centromere and at the ends of the chromosome arms.

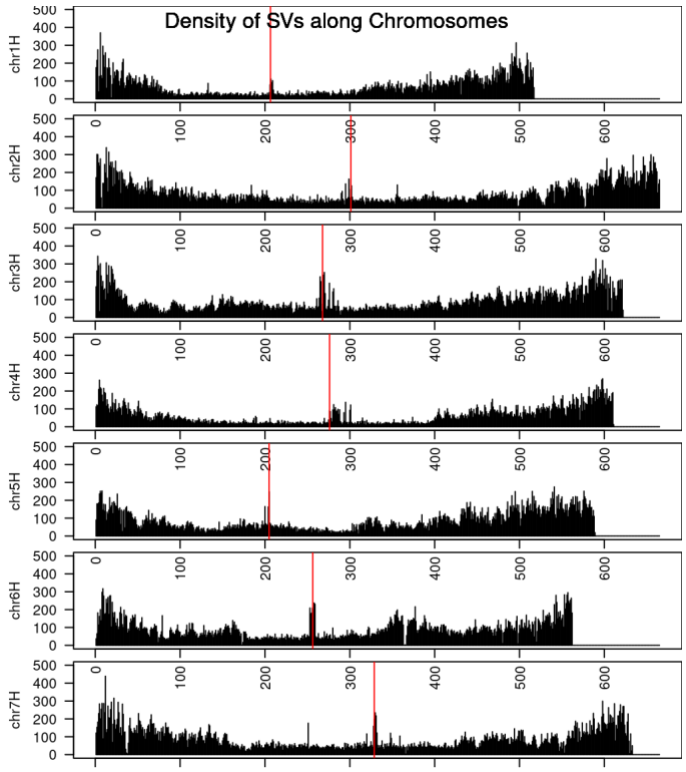


Figure 10

5% mRNA exhibit significant level of overlap with deletion structural variants

We next investigated the overlap between deletions variants and gene sequences. Figure 11 is a barplot after I remove all the mRNAs that have 0 overlap with deletion structural variants. Each bin is the percentage overlap of mRNA by deletion and the y-axis is the count that falls into each bin. Majority of overlap is 100%. The bigger the overlap, the more likely the mRNA will have loss of function. Next, I extract the reads that have overlap bigger than 75%. Out of approximately 38,000 mRNAs in total, there are 1944 mRNAs that overlap bigger than 75%.

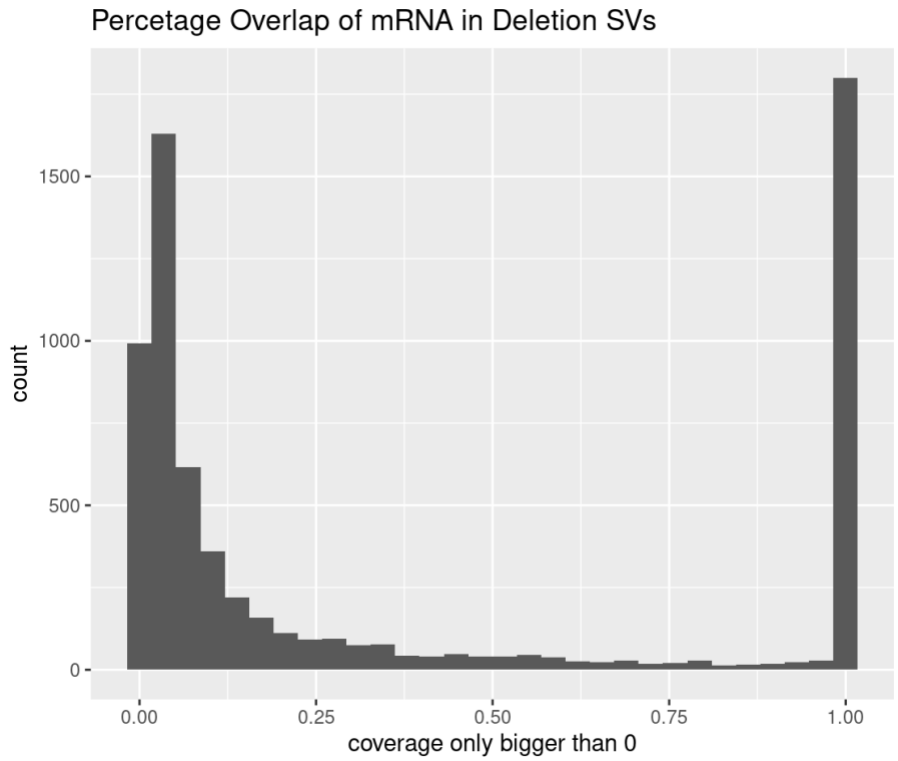
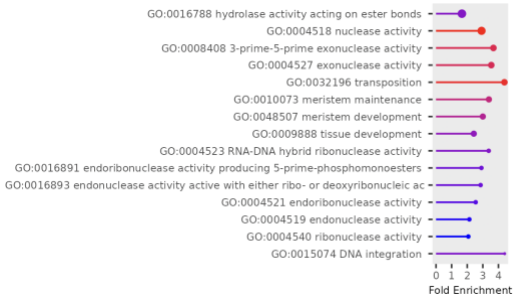


Figure 11

Overrepresented Genes Function

To understand whether genes involved in some biological processes were deleted more than others, we performed enrichment analysis on ShinnyGO 0.80. The enrichment analysis shows the biological process and molecular function that are overrepresented in the high coverage Deletion Structural Variants. Figure 12 shows the number of genes that relate to each GO pathway with different size of dot. The color of the line indicates the negative log value of false discovery rate. The higher the number reflects the more significant the enrichment is. Fold enrichment measures the magnitude of the enrichment. A bigger fold enrichment leading to a stronger enrichment. After generating the list of gene ID that is associated with each GO pathway, I compared each of the pathways using Rstudio. There are a total of 740 gene IDs provided by the enrichment, and only 250 of the genes are unique. I noted that pathways related to endoribonuclease, ribonuclease and endonuclease have the identical list of genes. The

pathways associated with tissue development, meristem development and meristem maintenance have the identical gene list. Transposition and DNA integration share one same gene. Figure 13 is a hierarchical tree that was provided by ShinnyGo showing the correlation between each GO pathway. Pathways that share genes are clustered together. There are three major clustering on the tree. A cluster that relates to movement of DNA sequences such as Transposable elements. A cluster relates to enzymatic activity and nucleic acid cleavage that associate with DNA replication, and repair. A cluster that relates to plant development. The big blue dots indicate the significance of P-value.



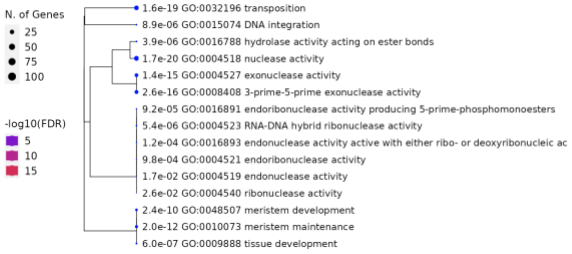


Figure 12 Figure 13

MAIN and MAIL Family Protein

The deletion structural variant that contributes to plant development could contribute to phenotypic differences. There are a total of 52 genes that are in the list. I used the rice and Arabidopsis thaliana database to perform blast analysis on all the genes. The result on the Arabidopsis thaliana protein database shows that all the genes are have high bit score to protein AT1G48120.1, AT1G17930.2, AT1G17930.1, AT2G25010.1, AT2G04865.1, AT2G04865.2. The proteins that have high pit scores all fall into the MAIN and MAIL family, with the average bit score of 106.

DISCUSSION AND CONCLUSION

Here, we present the result of structural variants studied in 45 barley individuals that were found to be genetically diverse. We perform the sequence alignment with the newest reference genome Barley MorexV3. There are 366,149 structural variants filters to our desired threshold. Our findings suggest that more than half of the structural variation is deletion, and the amount is followed by insertion, duplication and inversion. The founding is consistent with other researchers. “deletions were the most prevalent form of SVs, followed by insertions and duplications; inversions were the least common SV identified in this study” (Jobson 2022). The other result regarding the distribution of structural variants in the chromosome is also consistent with previous findings “suggest that structural variants distribution is nonrandom, and there is a higher prevalence of structural variants in centromeric and subtelomeric regions” (Jobson 2022).

The GO enrichment analysis on deletion overlapped genes shows us GO terms that are clustered into three different categories. They are related to movement of DNA sequence, enzymatic activity, and plant development. Those are the pathways that can potentially contribute to barley’s adaptation to different environments. I hypothesize that the gene related to DNA integration and transposition can be important for the movement of transposable elements which can increase the rate of mutation and a chance of having offspring that adapt to the extreme local environment. The gene that is related to enzymatic activity such as one that influences DNA repair, proofreading can contribute to rapid adaptation if being deleted. The other structural variant study in rice has similar findings on deletion. “The analysis runs on deleted temperate *Japonica* (Nipponbare) with core set genes that have basal housekeeping functions and processes such as developmental and metabolic progress, DNA-binding transcription factor activity and so on”(Fuentes 2019). To further validate our dataset, we can use

polymerase chain reaction to validate the deletion and insertion structural variants that have a length up to 0.3kb (Weisweiler 2022).

The protein in MAIN and MAIN like families is being hit very often during the blast analysis on all the plant development genes with protein Arabidopsis thaliana database. A study using a forward genetic approach has found that the presence of plant mobile domain proteins MAIN (MAINTENANCE OF MERISTEMS) and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements, MAIN acts synergistically and redundantly with DNA methylation to silence TEs in Arabidopsis thaliana (Nicolau 2020). PP7L is found to have functions in chloroplast development, high-light tolerance, and salt tolerance (Xu 2019). MAIL1 is important for root development and shows epistasis effect to MAIN (Ühlken 2014). The protein has a big role in proper gene expression and root tissue development in Arabidopsis thaliana and may as well have a role in barley. However, MAIN and MAIL are not studied extensively in Barley and no historical deletion on these genes was documented. Moreover, I did not expect to see the consistent appearance of MAIN and MAIN like family protein among all 52 genes in the list. The repeating protein result is due to the similarity of our gene of interest. My first hypothesis is that the high similarity gene can be caused by the modification of transposable elements overtime on different regions of the genome. My second hypothesis is that the gene of interest comes from the region of chromosomes that are highly repeated, particulty the region near centromeric and telomeric regions of chromosomes.

Now that we know the repetition of some of the gene and their interaction with transposable elements, there are more studies we can do on this dataset to investigate the relationship between structural variants, transposable element and the region where structural variants are found in the genome, and their relative position to coding frame. Future studies can

subset the deletion structural variants by length to investigate the relative position to coding frame because evidence from the rice paper shows “Short deletions peak in the 5′ UTR region, and long deletions are most frequent in the promoter region (Fuentes 2019)”.

Understanding the mechanism of barley adaptation to the wide range of environment, the role of structural variants, transposable elements will help breeder to develop crops that produce more yield, adapt to abiotic stress, provide better nutrient profile, help farmers to use less pesticide, and help us combat any global weather threat in other species.

REFERENCES

Danecek, Petr, et al. “The Variant Call Format and VCFtools.” *Bioinformatics (Oxford, England)*, vol. 27, no. 15, Oxford University Press, 2011, pp. 2156–58, https://doi.org/10.1093/bioinformatics/btr330.

Danecek P, et al. “Twelve Years of SAMtools and BCFtools.” *GigaScience*, vol. 10, no. 2, 29 Jan. 2021, https://doi.org/10.1093/gigascience/giab008.

Fuentes, Roven Rommel, et al. “Structural Variants in 3000 Rice Genomes.” *Genome Research*, vol. 29, no. 5, May 2019, pp. 870–80, https://doi.org/10.1101/gr.241240.118. Accessed 15 May 2022.

Ge, Steven Xijin, et al. “ShinyGO: A Graphical Gene-Set Enrichment Tool for Animals and Plants.” *Bioinformatics*, edited by Alfonso Valencia, vol. 36, no. 8, Dec. 2019, pp. 2628–29, https://doi.org/10.1093/bioinformatics/btz931.

Harlan, Harry V., and Mary L. Martini. “A Composite Hybrid Mixture 1.” *Agronomy Journal*, vol. 21, no. 4, Apr. 1929, pp. 487–90, https://doi.org/10.2134/agronj1929.00021962002100040014x. Accessed 19 Apr. 2021.

Huala, Eva, et al. “The Arabidopsis Information Resource (TAIR): A Comprehensive Database and Web-Based Information Retrieval, Analysis, and Visualization System for a Model Plant.” *Nucleic Acids Research*, vol. 29, no. 1, Jan. 2001, pp. 102–5, www.ncbi.nlm.nih.gov/pmc/articles/PMC29827/.

Jobson, Emma, and Robyn Roberts. “Genomic Structural Variation in Tomato and Its Role in Plant Immunity.” *Molecular Horticulture*, vol. 2, no. 1, Mar. 2022, https://doi.org/10.1186/s43897-022-00029-w. Accessed 25 Mar. 2022.

Landis, Jacob B., et al. “Natural Selection Drives Emergent Genetic Homogeneity in a Century-Scale Experiment with Barley.” *BioRxiv (Cold Spring Harbor Laboratory)*, Cold Spring Harbor Laboratory, Sept. 2023, https://doi.org/10.1101/2023.09.22.557807. Accessed 8 May 2024.

Li, Heng. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” *Bioinformatics*, edited by Inanc Birol, vol. 34, no. 18, May 2018, pp. 3094–100, https://doi.org/10.1093/bioinformatics/bty191.

Marius Weisweiler, et al. “Structural Variants in the Barley Gene Pool: Precision and Sensitivity to Detect Them Using Short-Read Sequencing and Their Association with Gene Expression and Phenotypic Variation.” *Theoretical and Applied Genetics*, vol. 135, no. 10, Springer Science+Business Media, Aug. 2022, pp. 3511–29, https://doi.org/10.1007/s00122-022-04197-7. Accessed 8 May 2024.

Mascher, Martin, et al. “Long-Read Sequence Assembly: A Technical Evaluation in Barley.” *The Plant Cell*, vol. 33, no. 6, Mar. 2021, pp. 1888–906, https://doi.org/10.1093/plcell/koab077. Accessed 14 Aug. 2021.

Nicolau, Melody, et al. “The Plant Mobile Domain Proteins MAIN and MAIL1 Interact with the Phosphatase PP7L to Regulate Gene Expression and Silence Transposable Elements in Arabidopsis Thaliana.” *PLOS Genetics*, vol. 16, no. 4, Public Library of Science, Apr. 2020, pp. e1008324–24, https://doi.org/10.1371/journal.pgen.1008324. Accessed 8 May 2024.

Quinlan, Aaron R., and Ira M. Hall. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics*, vol. 26, no. 6, Jan. 2010, pp. 841–42, https://doi.org/10.1093/bioinformatics/btq033.

RStudio Team. “RStudio | Open Source & Professional Software for Data Science Teams.” *Rstudio.com*, 2023, www.rstudio.com/.

Sakai, Hiroaki, et al. “Rice Annotation Project Database (RAP-DB): An Integrative and Interactive Database for Rice Genomics.” *Plant and Cell Physiology*, vol. 54, no. 2, Jan. 2013, pp. e6–0, https://doi.org/10.1093/pcp/pcs183.

Sakellariou, Michalia, and Photini V. Mylona. “New Uses for Traditional Crops: The Case of Barley Biofortification.” *Agronomy*, vol. 10, no. 12, Dec. 2020, p. 1964, https://doi.org/10.3390/agronomy10121964.

Sedlazeck, Fritz J., et al. “Accurate Detection of Complex Structural Variations Using Single-Molecule Sequencing.” *Nature Methods*, vol. 15, no. 6, Apr. 2018, pp. 461–68, https://doi.org/10.1038/s41592-018-0001-7. Accessed 28 Apr. 2020.

Sutton, Tim, et al. “Boron-Toxicity Tolerance in Barley Arising from Efflux Transporter Amplification.” *Science*, vol. 318, no. 5855, Nov. 2007, pp. 1446–49, https://doi.org/10.1126/science.1146853. Accessed 15 Feb. 2022.

Ühlken, Christine, et al. “MAIN-LIKE1 Is a Crucial Factor for Correct Cell Division and Differentiation InArabidopsis Thaliana.” *The Plant Journal*, vol. 78, no. 1, Mar. 2014, pp. 107–20, https://doi.org/10.1111/tpj.12455. Accessed 21 Mar. 2022.

Visioni, Andrea, et al. “Advancing the Conservation and Utilization of Barley Genetic Resources: Insights into Germplasm Management and Breeding for Sustainable Agriculture.” *Plants*, vol. 12, no. 18, Jan. 2023, p. 3186, https://doi.org/10.3390/plants12183186. Accessed 21 Jan. 2024.

Wickham, Hadley. “Ggplot2: Elegant Graphics for Data Analysis.” *Google Books*, Springer International Publishing, 2016, books.google.com/books/about/ggplot2.html?id=RTMFswEACAAJ.

Xu, Duo-Rong, et al. *Extrachloroplastic PP7L Functions in Chloroplast Development and Abiotic Stress Tolerance*. no. 1, May 2019, pp. 323–41, https://doi.org/10.1104/pp.19.00070. Accessed 26 May 2023.

Yuan, Yuxuan, et al. “Current Status of Structural Variation Studies in Plants.” *Plant Biotechnology Journal*, vol. 19, no. 11, July 2021, pp. 2153–63, https://doi.org/10.1111/pbi.13646. Accessed 18 Mar. 2022.