Forschung hautnah:

Wissenschaftliches Schülerpraktikum vergeben durch den

Förderverein der BiologieOlympiade e.V.

Am Max-Planck-Institut für molekulare Pflanzenphysiologie

In Potsdam-Golm

Arbeitsgruppe: Prof. Dr. Alisdair Fernie

Betreuer: Dr. Saleh Alseekh

Thema: Quantitative Trait Loci for Metabolic Traits in Tomato

Martha Genzer

Klassenstufe: Abitur 2022

Schule: Gymnasium am Tannenberg in Grevesmühlen

Zeitraum: 05.09 – 30.09.2022

Index

1. Introduction

1.1 Personal Introduction

My mother and father, who are both farmers, awakened my interest in biology. In school, I liked all the natural sciences and participated in different competitions like the Math-Olympiad and “chemie - die stimmt!”. When I got older, I became especially interested in biology. In my future, I also want to work in this field and therefore I will study biochemistry at Heidelberg.

I just finished my abitur with mathematics and biology as my advanced courses. My school emphasizes STEM-subjects. Luckily, my teacher for biology informed me about the Biology-Olympiad. I participated in four national rounds and finally won a silver medal at the International Biology-Olympiad 2022 in Armenia. By reaching the third round of the national Biology-Olympiad 2021/22, the “Förderverein der BiologieOlympiade e.V.” gave me the opportunity to do an internship at the Max Plank Institute of Molecular Plant Physiology in Potsdam-Golm in cooperation with Dr. Alseekh. This internship makes it possible for me to understand scientific work from doing experiments in the lab to evaluating the data to writing about it.

1.2 Institutional Introduction

Like all the Max Planck Institutes, the Max Planck Institute of Molecular Plant Physiology carries out fundamental research. The Institute describes its work as “investigating fundamental processes of plant physiology and development and the interaction of plants with the environment.”[1] I worked in the group of Prof. Alisdair Fernie, who researches central metabolism. My project group mentor Dr. Alseekh focused on genetics in crop metabolism. In his group they combine the quantitative genetic approaches and metabolomics techniques to understand the genetic architecture underlying naturally occurring plant metabolic variance in crop species such as tomato (*Solanum lycopersicum*), beans (*Phaseolus vulgaris L.*) and lupine (*Lupinus albus L.*).

The Institute is located at the Potsdam Science Park together with two other Max Planck Institutes, the University of Potsdam and two Frauenhofer-Institutes. There work about 410 employers from all parts of the world.

Figure 1.1: The Max Planck Institute of Molecular Plant Physiology and the foil tunnel

1.3 Abstract

The aim of the project I worked on during my internship was to analyze primary and secondary metabolites in leaves of a set of backcross inbred lines (BILs) population developed by crossing the *Solanum lycopersicum* with the wild tomato *Solanum pennellii*. I extracted the metabolites from the leaves and analyzed the content of the extracts by liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS). The metabolite data (phenotypic data) was then used to do a quantitative trait loci (QTL) mapping. Quantitative trait loci analysis is a powerful tool for the association between phenotypic traits (e.g metabolites) and specific genome regions. With this analysis, two candidate genes were identified for some metabolic traits. In this report, I will introduce metabolomics and the methods used here.

2. Project-referring information

2.1 Background information

Leaf materials from tomato backcrossed inbred lines (BIL) were used for this project. The lines were generated as the following. First, the wild tomato species *S. pennellii* and domesticated *S. lycopersicum* were crossed to generate the F1 plants. Then the F1 hybrid plants were backcrossed to *S. lycopersicum* to obtain BILs, backcross processing was done two times. After six rounds of self-pollinating of this 2nd BC line, 1500 (BC2S6) BILs were generated. Samples (fruit and leaf materials) from different seasons were harvested for the BIL population. The

BILs population was previously genotyped using single primer enrichment technology (SPET)[2]. Some of the BILs were selected and resequenced in order to have detailed information about the genome. During my internship in the group, I worked with leaf materials from the 40 resequenced BILs, each line with three biological replicates. In total, I have extracted and analyzed 120 samples for two metabolic platforms and performed QTL mapping. The schematic overview of the population is indicated in figure 2.1.

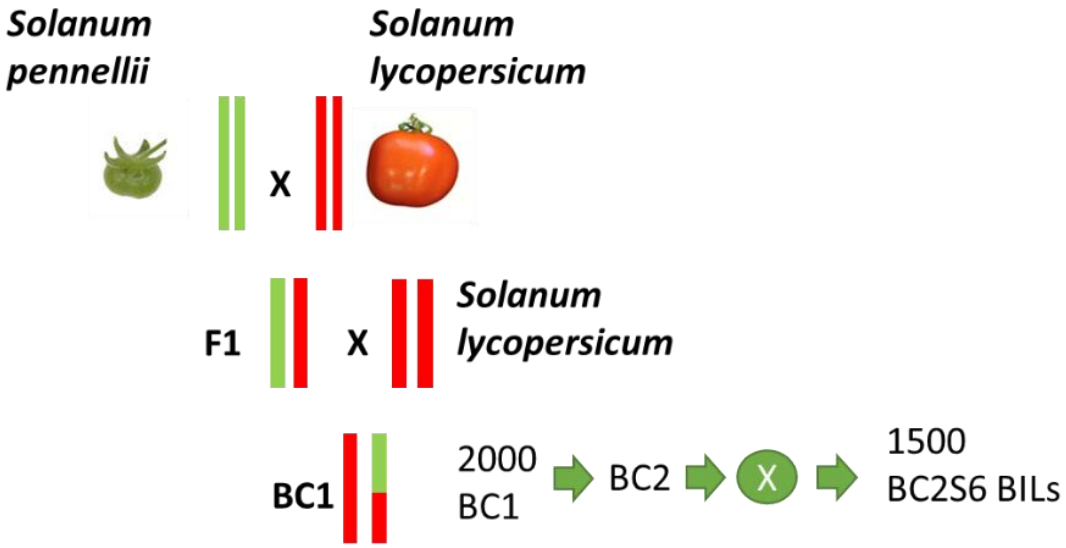

Figure 2.1: The schematic overview of the BILs population.

2.2 Quantitative trait loci (QTL) approach

In order to identify the region in the tomato genome associated with the accumulation of primary and secondary metabolites, single nucleotide polymorphisms (SNPs) across 120 lines were used for mapping.

The QTL results were displayed as a Manhatten plot with $-\log_{10}$(p values) plotted against the position in the genome. The threshold for the Manhatten plot was detected by Bonferroni correction. Each dot in the Manhatten plot represents a SNP marker and a high logarithm of odd (LOD) score represents the high possibility of correlation of a marker and the phenotype of interest. Therefore, the significant SNP indicates the position of the genome region, where the responsible gene encoding for the trait of interest is located within a certain interval.[3]

2.3 Methods to analyze metabolites

To analyze the metabolites in leaves, we used gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS) approaches in order to detect primary and secondary metabolites, respectively.

In the GC-MS approach, while detecting primary metabolites it is essential to have volatile molecules. Therefore, the derivatization process was necessary. To do so, *N*-Methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) was added to each extract. This process is called silylation which is responsible for creating volatile compounds.

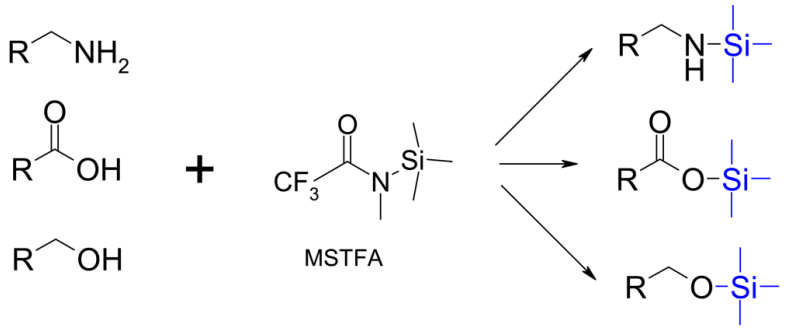

Figure 2.2: Reaction of amines, carboxylic acids, and aldehydes from metabolites with silyl groups from MSTFA

In the GC, an auto-sampler with an automatic needle takes 1 µL of the liquid sample and injects it into a capillary column. The sample is heated and vaporized. The column separates the components because there is a mobile phase, the inside walls of the column, and a stationary phase, the carrier gas. Depending on their polarity and boiling point, the components are transported through the column at different speeds and at different times. The oven increases the temperature so that components with a low boiling point pass the column first and components with a high boiling point late. When the components finally elute from the column, a detector converts the amount of each component into an electric signal. The time when the detector senses a component is called retention time (RT). The analysis results are obtained as chromatograms.

The LC works with a similar method, the liquid sample is carried by a liquid eluent through the column. In a reverses phase column, the inside of the column contains nonpolar long chain alkyl groups on silica particles and forms the stationary phase. The mobile phase is the eluent, which composition changes from a higher concentration of a polar eluent to more of a nonpolar eluent. Therefore, the retention time depends on the chemical affinity of components with their interaction between the mobile and stationary phases.

The gas chromatography and liquid chromatography machines are connected with mass spectrometers. In mass spectrometry, the molecules are ionized and often fragmented in the ion source. Then the ions accelerated and subjected to a magnetic field in some mass analyzers. A detector measures the deflection, which depends on the mass/charge (*m/z*) ratio of ions. The components can be identified by their fragments mass and retention time.[4]

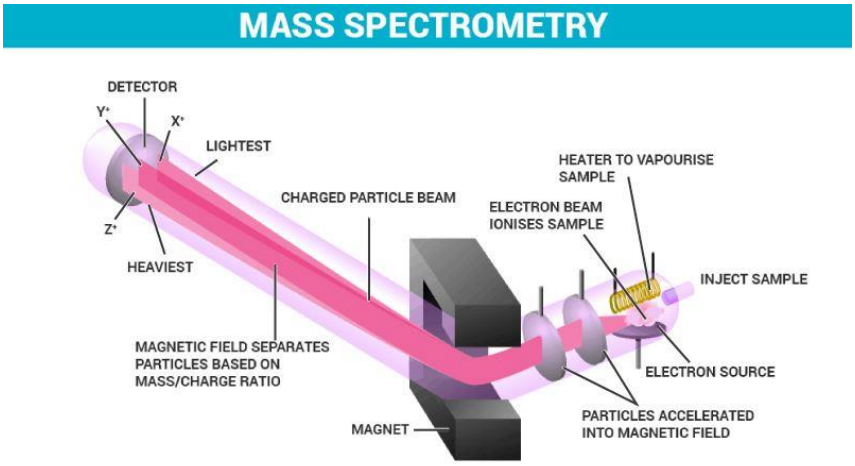

Figure 2.3: Scheme of the mass spectrometry

The data from the GC-MS and LC-MS at the end are presented in a three-dimensional structure, the retention time reflects the chromatography (how much time the molecule spends to pass the column), the mass spectrum (the ions and their fragments) measured by the mass analyzer, and the third dimension is the intensity of the ions which reflects the relative quantity present in the original sample (the tomato samples/lines in my case). I used the intensity of the metabolites (phenotypic data) measured by GC-MS and LC-MS and did the statistical association genotypic data (sequencing and molecular markers across the tomato lines). This method is called QTL mapping which allows the identification of quantitative trait locus (QTL), a genomic region that influences a phenotypic trait (metabolite content in the tomato lines).

3. Material and methods

3.1 Grinding and Aliquoting

The experiments started by grinding the leaf samples and transferring materials to screw tubes. I have grinded about 120 samples (40 lines with three biological replicates). The samples must be all the time cooled in liquid N_2 or dry ice to stop the enzymatic activity. After that, around 50 mg (+/- 5 mg) of the leaf material was weighed into a 2 ml tube and a metal bead was added

to each sample. I made 3 replicates for each line. Then, the leaves were pulverized with the grinder to a fine powder.

3.2 Extraction

Metabolite extraction for 120 leaves samples was conducted using the general protocol from the lab with methyl tert- butyl ether (MTBE) and methanol-water (MeOH).

For the metabolite extraction, an MTBE/MeOH buffer was used. This buffer was prepared by mixing 100 ml methanol with 64 µl Ribitol (25 mg/ml, as a standard), 500 µl Isovitexin (1 mg/ml), 300 ml MTBE and 36 µl PC (5 mg/ml). The second buffer was prepared with the ratio of 3:1 water: methanol (150 ml water and 50 ml methanol). In order to keep the enzymatic activity the buffer was kept on dry ice while performing metabolite extraction. Afterward, 1 ml of extraction buffer was added to the frozen leaf samples and vortex until it was fully re-suspended. Then the samples were incubated for 10 minutes at a 4 °C orbital shaker. Afterward, the samples were incubated for another 5 minutes in an ultra-sonication bath. 500 µl of the water: methanol (3:1) mixture was added into each tube and mixed well. Next, the tubes were centrifuged at full speed for 5 min in a tabletop centrifuge at 4 °C. The second buffer (water: methanol) caused a phase separation. 450 µl of the upper green apolar phase was transferred to a new tube for lipid analysis, 240 µl of the transparent polar phase to a new tube for secondary metabolite analysis and 120 µl of the transparent phase for primary metabolites. Then the samples were dried and concentrated by putting them in a SpeedVac and after 3 hours samples were collected and placed at -20°C.

3.3 GC-MS and LC-MS samples preparation

Before runing the GC-MS analysis, samples should be derivatized. The derivatization protocol was performed as the following: first of all, 120µl concentrated polar samples which were kept at -20°C were dried in a speed vac for about 15-20 min in order to get rid of any water droplets. 40 µl of methoxyaminhydrochlorid prepared with 20 mg/ml pyridine were added to each sample and then samples were incubated at 37 °C for 2 hours at 950 rpm. After the samples were shortly centrifuged and 70 µl MSTFA mix was added. This mix contains 5 ml MSTFA and 100 µl FAME. The tubes were incubated for another 30 min at 37 °C, samples were centrifuged again and 90 µl transferred into GC-MS glass vials.

For the secondary metabolites measurements, samples were dissolved in 200 µl of the methanol-water mix (v/v 1:1). After incubating for a few minutes at room temperature, samples were vortexed and sonicated in an ultra-sonication bath for 5 minutes. Thereafter, the samples

were vortexed again and centrifuged for 5 minutes at 14000 rpm. Finally, 90 µl was transferred into LC-MS glass vials.

3.4 Chromatography and mass analysis

The primary metabolites were analyzed using the GC-MS. In the system, helium was used as a carrier gas, and a 30-m MDN-35 capillary column was used for chromatography separation. After sample injection, the GC oven holds the temperature at 85°C for 3 minutes, and then the temperature increases to 15°C per minute for 25 minutes.

The secondary metabolites samples were analyzed with a UHPLC-MS/MS (UHPLC: Waters, ACQUITY UPLC System; MS: Thermo Scientific, Q Exactive Plus) on a reverse phase C18 column held at 40 °C running with a flow of 400 µL/min with gradual changes of eluent A (0.1% formic acid in water UHPLC grade) and B (0.1% formic acid in acetonitrile UHPLC grade). The mass spectra was obtained in a mass range of 100-1,500 m/z in a full MS scan and all ion fragmentation (AIF) induced by high-energy collisional dissociation (HCD) of 40 keV. Due to limited capacity while running the samples only the negative ionization mode was used. The obtained chromatograms were subjected to data processing (section 3.5) for annotation and quantification.

3.5 Data-processing

For the primary metabolites, I have performed peak peaking manually using the pipeline and the library that the group already established. The retention time shift was corrected and *m/z* for several ions were checked and compared to the reference library. Peak picking and the retention time shift correction were done using Xcalibur software (Thermo Fisher Scientific).

For LC-MS/MS data both targeted and non-targeted approaches were applied. The processing of chromatograms, peak detection, and integration were performed using RefinerMS (version 5.3; GeneData) and Xcalibur software (Thermo Fisher Scientific). For the analysis, the pipeline contained a noise filter, a retention time alignment, a peak detector and an isotope cluster. The noise filter removed the background signals, which were not from the prerequisite metabolites. The RT-alignment removed shifts in the retention time. The peak detector detected peaks to quantify their intensities.

After this processing, the data was exported into an excel sheet with the sample IDs, mass charge ratio and retention time (*m/z*). Then an algorithm compared the RT and *m/z* ratio library

with known RTs and *m/z* ratios from common metabolites and listed the identities (so the names of the metabolites) to the peaks.

The aim of the project was to the identification of variability of primary and secondary metabolites and ultimately to map genes involved in metabolite biosynthesis pathways. Although a large amount of data was generated by the metabolomics approaches and data processing, in this report I am indicating only some examples. Although, several methods are available to identify QTL, in this project I used a simple QTL mapping method called marker regression. For the analysis of QTL, marker regression considers each marker individually, split the individuals into groups, according to their genotypes at the marker, and compares the group's phenotype averages[5]. Furthermore, Manhatten plots were generated to indicate the chromosomal position of SNP markers for each metabolic trait. Then, I searched for the possible candidate genes associated with the metabolic traits by looking into the tomato genome.

4. Results

In total 120 samples including 40 tomato lines (BILs) each with three replicates were analyzed using both LC-MS and GC-MS machines for measuring the secondary and primary metabolites, respectively. After, chromatograms were evaluated to identify and determine the intensities of the metabolites across the samples. The GC data was normalized by ribitol and the average of the three replicates was used for the next step. Normalized data was uploaded into the MetaboAnalyst program to generate the figures (see figure 4.1).

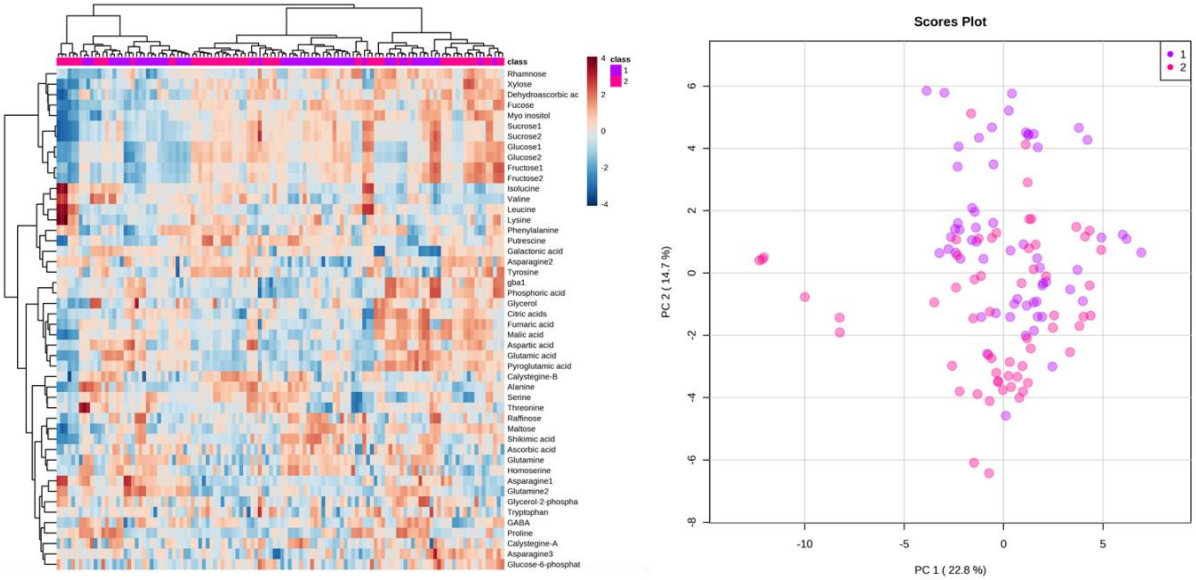

Figure 4.1: Heat map and PCA plot for GC data

Therefore, principal component analysis (PCA) illustrates the distribution of samples within 2 distinct batches. Principal component 1 covers 21, 7 % of the variance of the data, PC 2 covers 14%. The plot does not show clear clusters. That means there is no big difference between the two batches of samples. However, PCA shows quite a large variation between different tomato lines (BILs). The heat map depicts the magnitude of the metabolite changes in the different samples, where the red and blue color reflects the increase and decrease in metabolites content, respectively. Next, I used this data to do QTL mapping using an R script. Single QTL mapping method analysis was used, which postulates the presence of a single QTL. Each point on a dense grid across the genome is contemplated and contributes to the location of the putative QTL. The grid is built by genetic markers, in this case SNPs. A SNP is a single nucleotide polymorphism, which means a genomic variant at only a single base position. An issue of a single QTL is that regions between two SNPs must be inferred based on the SNPs. Manhattan plots show which SNPs correlate with a specific locus in the genome. The dots represent the LOD scores of the SNPs. A LOD score (“logarithm of the odds”) is a statistical estimate. It

shows the probability that two markers are located near each other on a chromosome. The LOD scores were logarithmized on the base 10 and multiplied by -1. The threshold for significance was calculated with a Bonferroni correction. This reduces the chance of false-positive results by dividing the standard significance level of 5 % by the number of hypotheses. The number of hypotheses equals the number of samples. This calculation revealed a threshold of 5.

For every identified component, a Manhattan plot was generated. Next, I want to show an example of interesting plots (significant associations for certain metabolites). The Manhatten plots from sucrose, glucose and fructose demonstrated a high correlation between the abundance of these sugars to SNPs on chromosome 5. The box plot comparing the intensity between the genotypes (different lines) shows differences in the metabolite levels and having SNP at certain genomic positions. The best significant SNP was identified as SNP_SSL2.50CH05_2996884. We searched for genes correlated to sugar at this region on chromosome 5 in the tomato genome and found solyc05g008600 as a candidate gene, which is annotated as a fructose-bisphosphate aldose. This enzyme is important for the Calvin-Benson-Cycle, which generates sucrose as an end product, that can be transported out of the leaves to the fruit. The expression of the gene was checked from RNA seq data that the group already has, and the result showed that the candidate gene (solyc05g008600) is highly expressed in wild tomato species (*S. pennellii)* compared to the cultivated tomato variety (*S. lycopersicum)*.

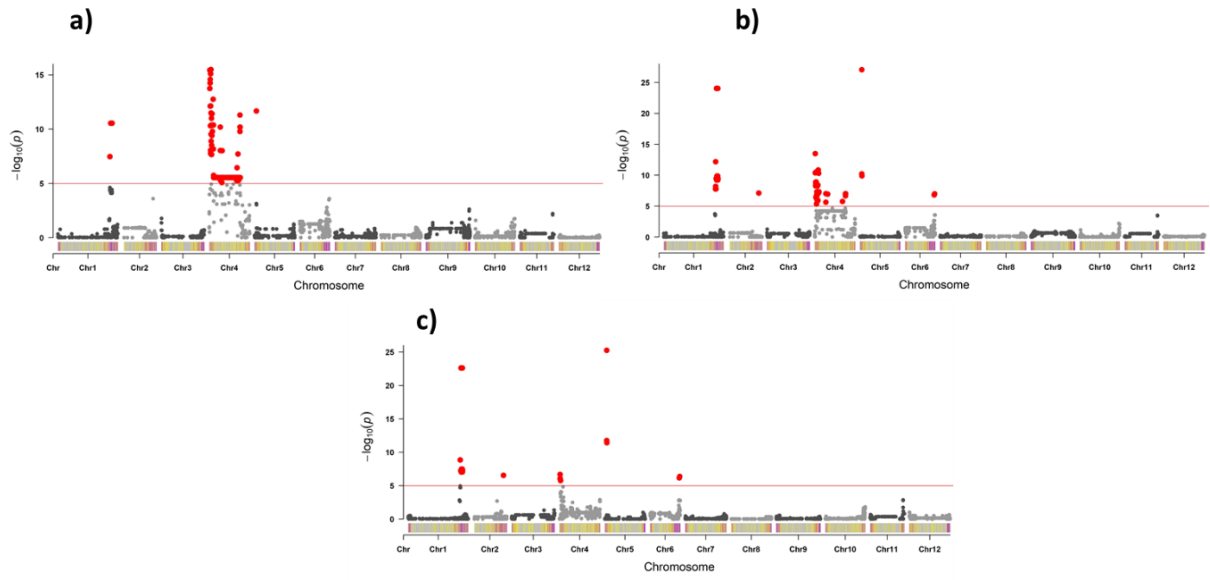

Figure 4.2: The Manhatten plots indicate the significant QTL on chromosome 5 for fructose (a) glucose (b) and sucrose (c). The threshold of $-\log_{10}$ (p) was set to 5 representing Bonferroni corrected statistical threshold.

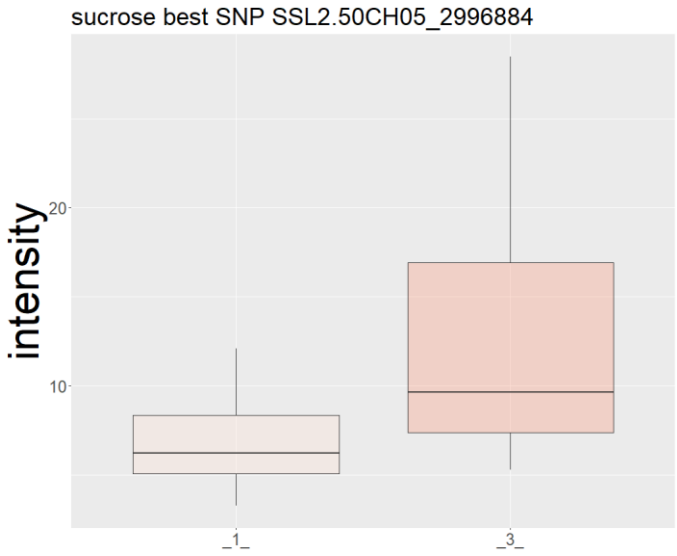

Figure 4.3: Box plot comparing the intensity from LEA determinate BILs (1) and the intensity from *S. pennellii* BILs (3)

To explore the metabolic diversity in the samples, I generated PCA plots using the LC-MS data with MetaboAnalyst. The data transformation was done as a log transformation to the base 10. The data was normalized by a median, which means that the plot is mean-centered and every variable is divided by its standard deviation.

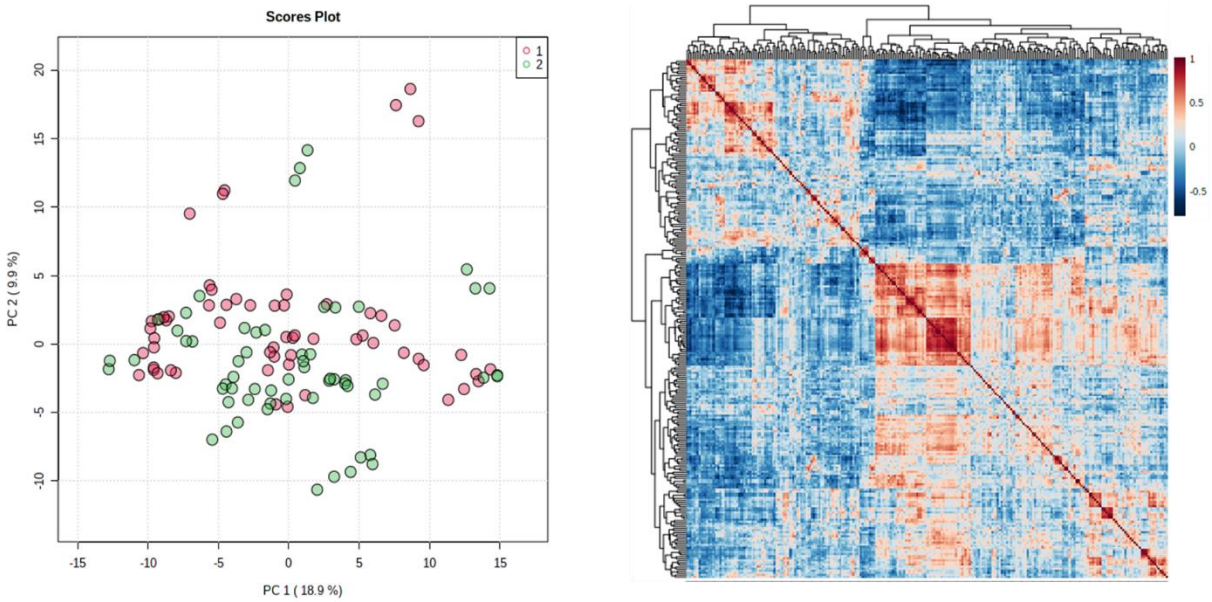

Figure 4.4: Auto-centred PCA plots, correlation spearman normalized by Isovitexin

I also performed a QTL analysis with the LC-MS data, using a similar approach to the GC-MS data above. In the Manhattan plot of the flavonoid Neohesperidin, we found an interesting correlation. The abundance of this metabolite is associated significantly to SSL2.50CH09_70593765 SNP. In this region I found a candidate gene Solyc09g091510 on chromosome 9, which is related to the chalcone synthase. Chalcones are belonging to the

flavonoid group and can be used as anti-cancer agents. Neohesperidin is also a chalcone, so the candidate gene might be responsible for the abundance of this compound.

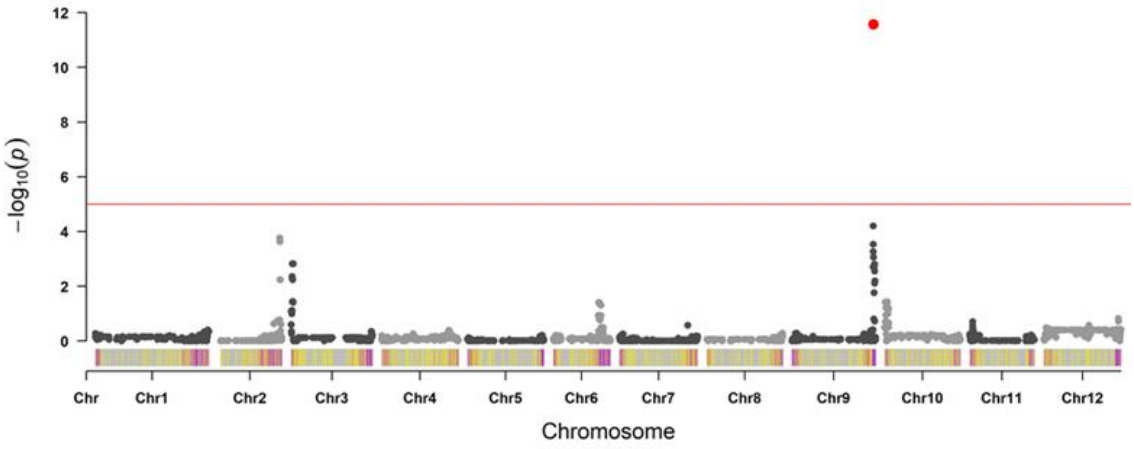

Figure 4.5: The Manhattan plot of Neohesperidin which is giving significant QTL on chromosome 9

5. Discussion

Plants produce many diverse metabolites classified mainly as primary and secondary metabolites. These compounds play many different roles in plant growth and development and import in plant response to the surrounding environments such as abiotic and biotic stresses[6].

In my project, I have analyzed a small set of selected backcross inbred lines (40 lines with three replicates) to explore the genetic diversity of these lines. The result I obtained is very interesting and therefore further analysis would be necessary to verify and extend the data analysis.

During my internship, I learned and used both GC-MS and LC-MS methods to do metabolomics profiling. Following recent protocols in the lab such as "Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices" from Alseekh, S., Aharoni, A., Brotman, Y. et al. ensure reproducible results[7,8]. Further, I also performed QTL mapping and identified candidate genes that might be important for primary metabolites biosynthesis such as sucrose, glucose and fructose and secondary metabolites biosyntheses such as Neohesperidin.

Sugars are formed in the Calvin-Benson cycle and have a great important role in a plant's growth and development. In previous studies, genes were identified that regulated the sugar level in leaves and fruits. For example, in tomato, the QTL LIN 5 modifies the fruit sugar

composition by influencing apoplastic invertases[9]. My results here add information and a new candidate gene that might be contributed to sugar metabolism in the plant. For secondary metabolites such as neohesperidin which is a precursor of neohesperidin dihydrochalcone built through alkaline hydrogenation. Neohesperidin dihydrochalcone is a low-calorie sweetener which is up to 1800 times sweeter than sucrose. It is used in foodstuffs like sweets and drinks, but also in pharmaceutical products to suppress bitterness[10]. The results obtained in this project could help to design and select lines that might be important in breeding programs for tomatoes.

6. Side Project: Biological validation by overexpression

To prove the candidate genes that we found, biological validation is needed. This is usually done by performing CRISPR-Cas9, overexpression or virus-induced gene silencing (VIGS) to observe phenotypic changes in the plant. Nevertheless, there was not enough time for me to do the biological validation for the genes that I detected after the data analysis, so I helped to perform an overexpression assay in another related project. The PCR and gel extraction was already done when I joined the project. I helped with the Gateway cloning including BP reaction, *E. coli* transformation, colony PCR, plasmid isolation, LR reaction and *Agrobacterium* transformation.

The PCR multiplied the gene and added the attB1 and attB2 adapter primers to the gene of interest. Then, the quality and yield of the attB-PCR product was tested by performing an agarose gel electrophorese. Afterward, the gel was purified with a kit. The kit contained all the necessary chemicals for purification. After cell lysis and the clarification of the lysate, the DNA was separated and bound to a plasmid column with a silica membrane. Finally, the pure DNA was eluted. This was used for gateway cloning. Therefore, the enzyme BP clonase II was added to the PCR product and a donor vector. The enzyme catalyzed the reaction from the gene with the attB ends from the PCR and the donor vector with attP sides to the entry clone with the insert and attL ends and a byproduct with attR ends.

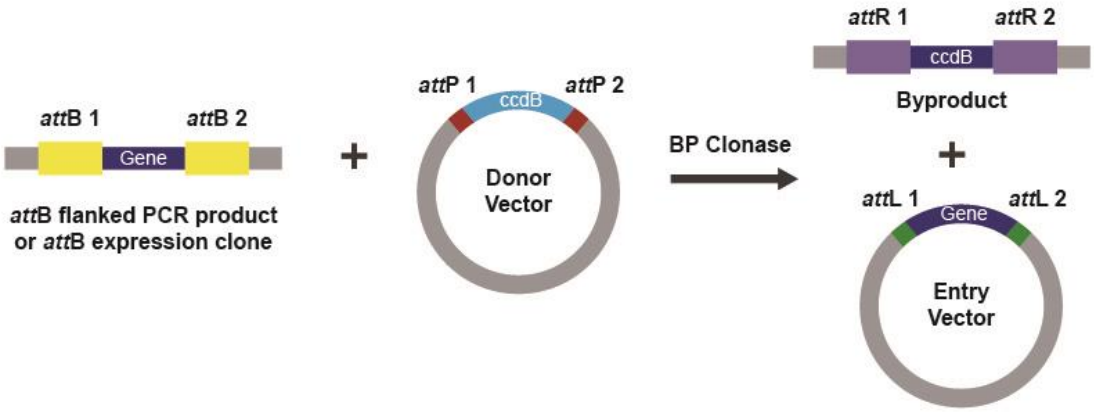

Figure 6.1: Scheme of the BP reaction of Gateway cloning

This entry vector was incubated in chemically competent *E.coli* cells. After 30 minutes, a heat shock was performed to insert the vector into the bacteria. LB growth medium was added to the cells, the vials were shaken for one hour and then the bacteria were spread on selective plates to form colonies. Plates that have selective antibiotics were incubated at 37°C overnight. Positive colonies checked by colony PCR were purified after gel electrophoresis. Overnight inoculation was performed with a liquid medium that has selective antibiotics. Then plasmid isolation was performed, and plasmid DNA was sent for sequencing. After sequencing the LR reaction was performed. The LR enzyme works the opposite way of the BP enzyme and catalyzes the reaction from the entry clone with the gene and attL ends and the destination vector with attR ends to an expression clone with attB ends and the gene and a byproduct with attP ends. The same process was applied as it was done in the BP reaction and plasmid DNA

was sent for sequencing. Following to LR reaction, *Agrobacterium* transformation was performed and the constructs were delivered to the greenhouse team for the plant injection.

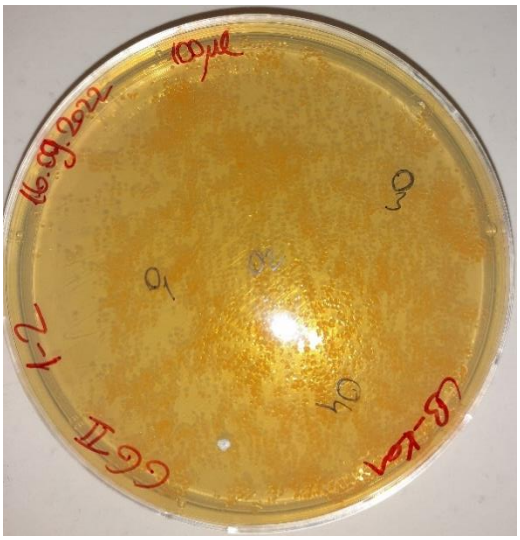

Figure 6.2: LB-Kanamycin resistance plate with *E.coli* cells

7. Personal Conclusion

During my internship, I learned how scientists work. The experience of standing in a lab doing experiments was new and a possibility to see processes like PCR in reality aside from talking about it in school.

8. Acknowledgement

I would like to thank Dr. Saleh Alseekh for explaining to me all the things and making this internship possible. Thanks to Esra Karakas for preparing this project for me and guiding me through it. Finally, I would like to thank Mustafa Bulut and Regina Wendenburg for helping me and answering my questions. ☺

9. Bibliography

Picture Sources:

I photographed picture 1.1 and 6.2. The figures in the result section were created using the online tool MetaboAnalyst.

- Logos: https://tse3.mm.bing.net/th/id/OIP.NF4V8Zu0OwYn_lIJG3Nh7QHaDs?pid=ImgDet&rs=1 and https://tse2.mm.bing.net/th/id/OIP.MTEl-QNDyPKEy22GPfjbpgHaDT?pid=ImgDet&rs=1 accessed on 27.09.2022 at 16.42 o'clock
- Figure 2.2: MDPI and ACS Style. Villas-Bôas, S.G.; Smart, K.F.; Sivakumaran, S.; Lane, G.A. Alkylation or Silylation for Analysis of Amino and Non-Amino Organic Acids by GC-MS? *Metabolites* 2011, *1*, 3-20. Metabolites | Free Full-Text | Alkylation or Silylation for Analysis of Amino and Non-Amino Organic Acids by GC-MS? (mdpi.com)
- Figure 2.3: https://s3-ap-southeast-1.amazonaws.com/subscriber.images/chemistry/2018/02/12114729/Mass-spectrometry-diagram.jpg accessed on 27.09.2022 at 16.32 o'clock
- Figure 6.1: https://www.molecularcloud.org/assets/img/resource/html-3.jpg accessed on 26.09.2022 at 14.38 o'clock

References:

1. Home | Max-Planck-Institut für Molekulare Pflanzenphysiologie (mpg.de) (26.09.2022, 09.28)

2. Schmidt, M.H., Vogel, A., Denton, A.K., Istace, B., Wormit, A., van de Geest, H., Bolger, M.E., Alseekh, S., Maß, J., Pfaff, C., Schurr, U., Chetelat, R., Maumus, F., Aury, J.M., Koren, S., Fernie, A.R., Zamir, D., Bolger, A.M., and Usadel, B. (2017). De Novo Assembly of a New Solanum pennellii Accession Using Nanopore Sequencing. The Plant cell 29, 2336-2348.

3. Quantitative Trait Locus (QTL) Analysis. Cecelia M. Miles, Ph.D. & Marta Wayne, Ph.D. © 2008 Nature Education Quantitative Trait Locus (QTL) Analysis | Learn Science at Scitable (nature.com).

4. R Waddell Smith, Mass Spectrometry, Encyclopedia of Forensic Sciences (Second Edition), Academic Press, 2013, Pages 603-608, ISBN 9780123821669, https://doi.org/10.1016/B978-0-12-382165-2.00250-6

5. Broman, K.W., and Sen, S. (2009). A Guide to QTL Mapping with R/qtl. (Springer).

6. Wang, S., Alseekh, S., Fernie, A.R., and Luo, J. (2019). The structure and function of major plant metabolite modifications. Molecular plant 12, 899-919.

7. Alseekh, S., Aharoni, A., Brotman, Y. *et al.* Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods* 18, 747–756 (2021). https://doi.org/10.1038/s41592-021-01197-1 .

8. Mohamed Salem, Michal Bernach, Krzysztof Bajdzienko, Patrick Giavalisco. A Simple Fractionated Extraction Method for the Comprehensive Analysis of Metabolites, Lipids, and Proteins from a Single Sample. Jove June 2017. A Simple Fractionated Extraction Method for the Comprehensive Analysis of Metabolites, Lipids, and Proteins from a Single Sample | Protocol (jove.com).

9. Fridman E, Zamir D. Functional divergence of a syntenic invertase gene family in tomato, potato, and Arabidopsis. Plant Physiol. 2003 Feb;131(2):603-9. doi: 10.1104/pp.014431.

10. Annett Braune, Wolfram Engst, and Michael Blaut: Degradation of Neohesperidin Dihydrochalcone by Human Intestinal Bacteria. Journal of Agricultural and Food Chemistry 2005 53 (5), 1782-1790. DOI: 10.1021/jf0484982

10. Appendix

The complete data tables can be requested from Dr. Saleh Alseekh (Alseekh@mpimp-golm.mpg.de).