



PLAN OF ACTION

First draft

Abstract

The goal of this project is to characterize the gene(s) responsible for the significant difference in flowering time between the late flowering species Jersey Kale (*Brassica oleracea*) and the rapid flowering species (TO1000, or *Brassica oleracea*[3]).

Rik Frijmann, Daphne van Ginneken, Esther Kockelmans, Nino
Vrolijk

22-09-2019

Table of contents

Introduction

Preliminary research

Research question

Method

Products

Project organization

Contact information

Planning

Risc analysis

Project boundaries

References

Introduction

This document is an overview of agreements among the project group members themselves, and between the project group and internal- and external managers. Examples of these agreements are the platforms on which products are delivered. In addition, it contains a schedule that describes the duration of the project. To clarify certain problems that may arise while completing the project, an elaborate risk analysis is performed as well. In this analysis, potential risks are treated and discussed in advance, in an attempt to mitigate these risks as much as possible. This project supports research conducted to identify why there are so many phenotypic differences between woody and herbaceous (cabbage) species.

Brassica, or Kale, is a genus of closely related plants showing great morphological differences. One of the most stark differences from the group is the Jersey Kale (*Brassica oleracea*), which grows an impressive wooden stem measuring at over a meter tall, often over a couple of years. This is in contrast to most of the other species in the genus, which are more akin to herbs or flowery bulbs, usually only growing for a year¹. In this project, the primary focus is to identify the genes responsible for the significant differences in flowering time between the Jersey Kale and its herbaceous counterparts.

¹ 1 University of Illinois Extension: Watch Your Garden Grow
<https://web.extension.illinois.edu/veggies/cabbage.cfm> Accessed September 17, 2019.

Preliminary research

This project builds on preliminary research on the same topic.² A crossbreed of a heterozygote Jersey kale and a homozygote rapidly flowering kale (TO1000) was made by manual cross-pollination. This resulted in an F1 generation which displayed a wide array of different flowering times, ranging from 113 days to up to 533 days.

The F1 generation was split into four groups based on flowering time, creating four distinct groups: early flowering (113-135 days), intermediate flowering (154-163 days), late flowering (176-239 days) and not flowering (240-533 days). The group 'not flowering' was later included, because these plants were simply unexpectedly late flowering. These groups were subjected to a BSA (Bulk Segregated Analysis), sequencing every member of a group in a bulk. This was then subjected to joint genotyping. As genes are passed on randomly, it is expected that this will lead to an overexpression in the dataset of the SNPs responsible for the flowering time phenotype of their respective groups.

Image 1 shows one of the preliminary results when comparing the groups amongst each other. Locations of SNPs occurring significantly more often than expected at random show high peaks, which in turn shows that most of the SNPs likely responsible for the length of flowering time in this sample occur on chromosome 6.

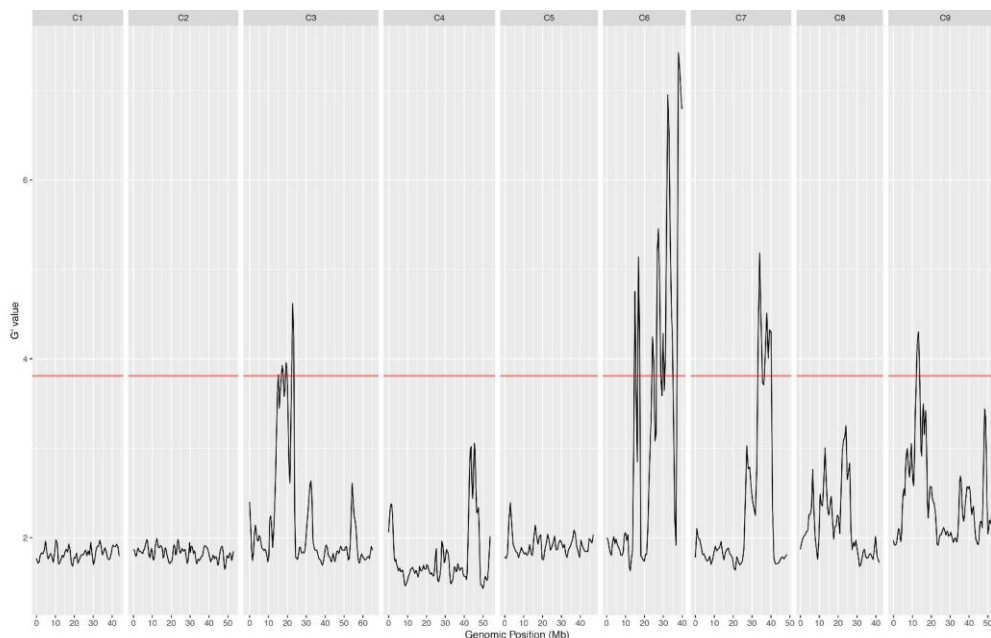


Image 1 One of the preliminary results comparing the groups. SNPs occurring more than expected show high peaks. Most of the SNPs likely responsible for flowering time occur in chromosome 6.

SNPs located on genes and their frequency were used in a later functional enrichment analysis. This revealed a few pathways of particular interest, as only early flowering or late flowering groups

² David vs Goliath: searching for genetic markers underlying growth form, life-history and correlating features using a crossing experiment in cabbage (*Brassica oleracea*).
<https://www.universiteitleiden.nl/en/research/research-projects/science/ibl-david-vs-goliath-searching-for-genetic-markers-underlying-growth-form-life-history-and-correlating-features-using-a-crossing-experiment-in-cabbage-brassica-oleracea>. Accessed September 16th, 2019

showed overexpression in these particular pathways. As the latest step, the genes of these pathways were colour coded with the type of SNP occurring on the gene. Resulting in overviews of SNPs in genes of the pathways, one of which is seen in image 2.

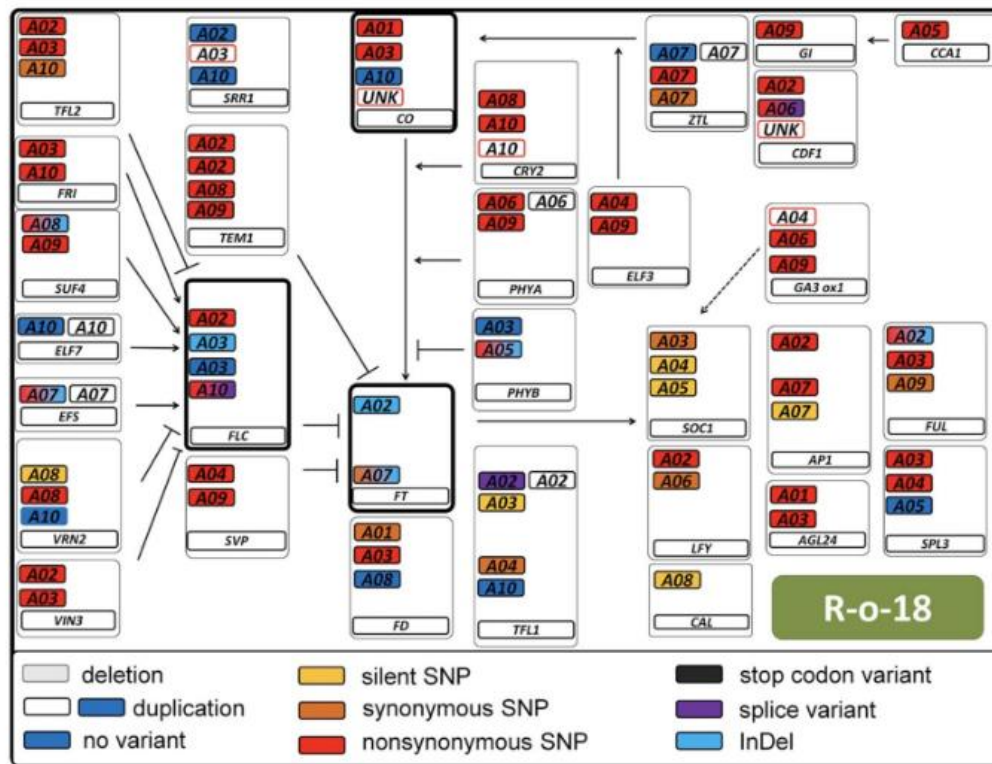


Image 2 Pathway for a different species of Brassica, responsible for the trait flowering time. The genes and their splicing variants are colour coded based on the type of SNP present in the genes.

Identification of genes responsible for flowering time

The complete pathways responsible for the trait flowering time have been determined during the preliminary research. During this project, the aim is to use these pathways of interest to scan the genomes of both organisms of the P-generation for mutations and the frequency thereof.

'Sequence from genes of interest' to 'gene sequence of second parent' steps:

During this step, the IDs of all genes in a KEGG pathway of interest will be used to make up a list of gene IDs. For every ID in the pathway, the sequence of that gene will be pulled from the well annotated genome of TO1000. Then, using BLAST for every gene, the corresponding gene variants of the Jersey kale parent will be identified.

Sequence comparison and SNP-analysis steps:

For this step, the project group will receive the next files: a GFF3 file, the reference genome and Jersey Kale genome in FASTA format, and a list with genes involved in the flowering time process. The GFF3 file is used to identify the locations of the flowering time genes. These coordinates can be used to trace the genes in the reference FASTA genome and in the Jersey Kale FASTA genome. The next step is to splice out those sequence and to compare these sequences of reference genome and Jersey Kale genome in order to find SNPs. And more specific, to find the location and frequency of those SNPs. And also to find out if a SNP translates to a different amino acid or codon.

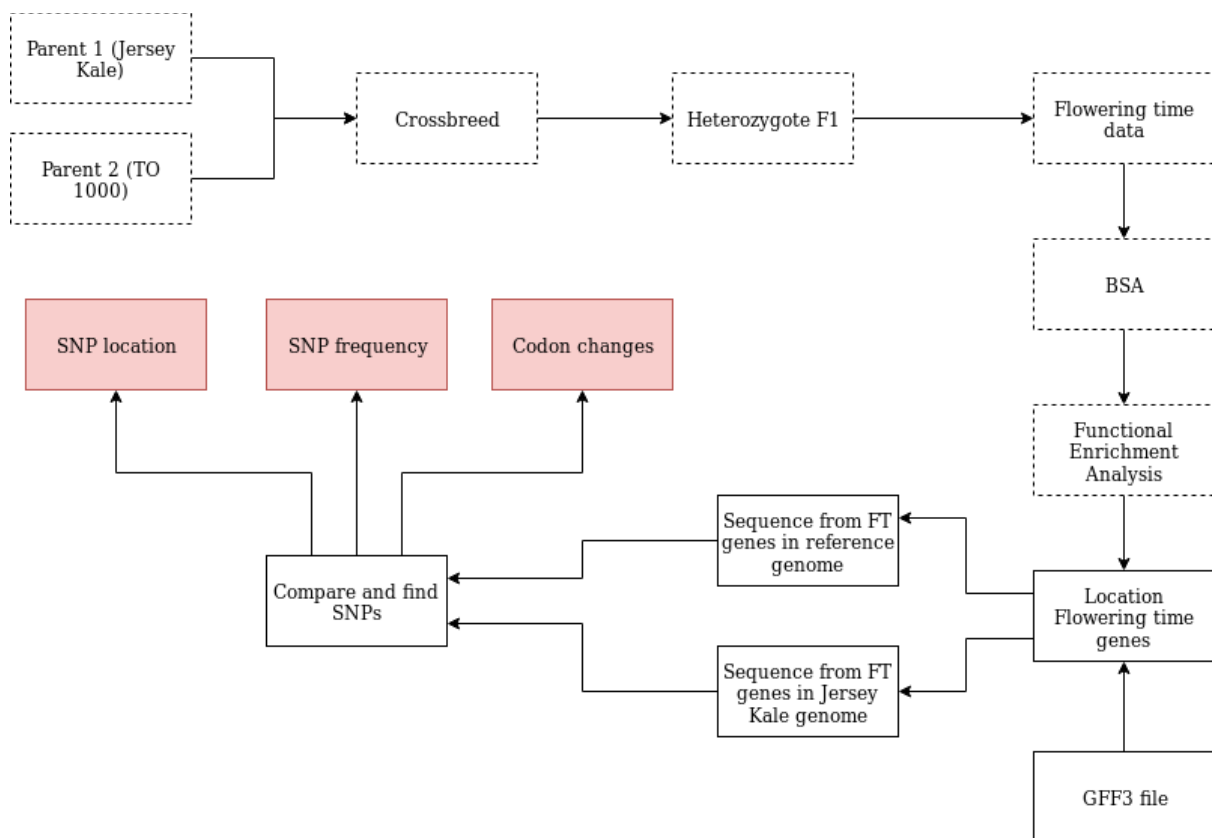


Image 3. Flowchart from the entire project. Steps outlined in dotted lines have been completed during preliminary research.

In the second semester, the group members will give a presentation on the current status of the project. The final presentation will take place at the end of the second semester, also given by one or two members of the group. During the third semester, the project group will hand in the final report.

Research question

The goal of this project is to characterize the gene(s) responsible for the significant difference in flowering time between the late flowering species Jersey Kale (*Brassica oleracea*) and the rapid flowering species (*TO1000*, or *Brassica oleracea*³).

³ Assembly *Brassica oleracea* var. *oleracea* (wild cabbage) . NCBI website.
https://www.ncbi.nlm.nih.gov/assembly/GCF_000695525.1/. Accessed September 16th, 2019.

Method

In preparation for the project each member of the group has to create a GitHub account and Email their own SSH (public) key to the managers. Using this key, the managers will be able to give grant each member access to the cloud. The unique GitHub account will be used as a central platform to save relevant documents as well.

First we compare the list of flowering time genes with the GFF3 file to find the locations of those genes. With these locations we can find the sequences of those genes in the reference genome and in the Jersey Kale genome. By comparing these sequences we can find SNPs.

Products

The products, as the project group will provide for the researchers at Naturalis Leiden, are described in this chapter.

SNP data:

By the end of the project, the project group will produce a list of SNPs in genes related to the flowering process, the frequency of occurring and the types of SNPs. This list of SNPs will be in flat text format.

Scripts and documentation:

The product outlined under 'SNP data' will be generated with the use of custom written scripts. These scripts will be used to keep the research reproducible. Therefore, the project group will provide extensive documentation on the use of the scripts, so any researcher or programmer using the scripts can do so with minimal additional time investment.

Presentations:

The project group will give a presentation with the aim of updating the entire research group 'Understanding Evolution' and a class of students on the status of the project. Before the end of the project, another presentation will be given, with the aim of informing the group 'Understanding Evolution' and a class of students on the final results. Both presentations will consist of two segments, one 10-15 minute presentation and discussion of the data, and one 10 minute segment for questions.

End report:

At the end of the project, the project group will provide a detailed report on the products and results.

Project Organization

This project is a collaboration between three separate groups: Hogeschool Leiden, a student group with the name BPEXA, and researchers from Naturalis. The group that is formed following this collaboration is called: Understanding Evolution.

The people involved in the project and a brief overview of their tasks is represented in table 1.

Table 1: Brief overview of all project members and their tasks.

Member	Affiliation	Role	Responsibility
Frijmann, Rik	Student group	Internal project leader, student	Primary contact for student group, oversight, main body of work
Ginneken, Daphne van	Student group	Student	Main body of work
Kockelmans, Esther	Student group	Student	Main body of work
Lens, Frederic	Naturalis, Leiden	Project researcher	Expert
Pijpe, Jeroen	Hogeschool Leiden	Student guide	Primary contact Hogeschool Leiden, final authority for conflict resolution
Vos, Rutger	Naturalis, Leiden	Project researcher	Primary contact Naturalis, Leiden; expert
Vrolijk, Nino	Student Group	Student	Main body of work

Primary contacts:

The primary contact for every organization is expected to check their email on at least two different days of the week, during office hours (9 am to 5 pm). One of these days should be the day of the weekly meetings. This measurement is to ensure there are enough contact possibilities throughout the week.

Any member that is not a primary contact, is expected to check their email on the day of meetings, or once a week during office hours - whichever occurs most often.

Oversight:

This task makes a project member responsible for maintaining an overview of all tasks that are being performed regarding the project. The individual is also expected to provide schedules for the weekly meetings between the student group and the Hogeschool Leiden.

Main body of work:

This member is responsible for the execution of the tasks as outlined in this plan of action.

Expert:

This member is an expert in a certain facet of the project. They are available for very specific problems. Only to be contacted during set times.

Final authority for conflict resolution:

This member has the final authority over the status of participation of members with the role 'student' to this project. They will enforce the behavior rules as outlined at the start of the project to the student group. They can be contacted when any member of the project encounters a situation with another member of the project, that they can not properly or safely resolve themselves.

Contact information

Rik Frijmann

- +316 24145123
- s1101245@student.hsleiden.nl
- Github: hfrijmann

Daphne van Ginneken

- +316 41172728
- s1099617@student.hsleiden.nl
- Github: dvginneken

Esther Kockelmans

- +316 31179951
- s1101570@student.hsleiden.nl
- Github: Estherrr1997

Frederic Lens

- +316 83172643
- frederic.lens@naturalis.nl

Jeroen Pijpe

- pijpe.j@hsleiden.nl

Rutger Vos

- +316 27085806
- rutger.vos@naturalis.nl
- Github: rvosa

Nino Vrolijk

- +316 23687852
- s1108075@student.hsleiden.nl
- Github: Ninjow

Planning

Table 2: Overall planning for the project.

All deadlines are expected deadlines, unless underlined dates. Underlined dates are hard deadlines.

Week	Task	Deadline	Responsible
1	Get in touch with the project leader	2-9-19	Rik
2	Learn more about the research	13-9-19	Everyone
3	Make plan of action	<u>22-9-19</u>	Everyone
	- Intro	-	Esther
	- Research Question	-	Esther
	- Method	-	Daphne
	- Products	-	Rik
	- Project Organization	-	Nino
	- Planning	-	Daphne
	- Risk Analysis	-	Nino
	- Project Boundaries	-	Rik
4	Literature research	27-9-19	Everyone
	Understanding the data	27-9-19	Everyone
	- GFF3 file	-	Everyone
	- Reference genome FASTA	-	Everyone
	- Jersey Kale genome FASTA	-	Everyone
	- List of Flowering Time genes	-	Everyone
5	Work on script to find location Flowering Time Genes in GFF3 file	-	Daphne
	Work on script to trace genes	-	Esther, Nino
	Work on script to compare sequences	-	Rik
6	Work on script to trace genes	-	Esther, Nino
	Find location Flowering Time Genes in GFF3 file	11-10-19	Daphne

	Work on script to compare sequences	-	Rik
7	Arrange midway presentation at Naturalis	<u>14-10-19</u>	Daphne
	Trace genes in the reference genome	18-10-19	Esther
	Trace genes in the Jersey Kale genome	18-10-19	Nino
	Work on script to compare sequences	-	Rik
8	Research how the SNP can affect protein translation	1-11-19	Daphne
	Compare sequences ref. genome with JK genome	1-11-19	Rik
9	Make powerpoint for first presentation	8-11-19	Esther
	Make a list of SNPs	8-11-19	Rik
	- Location of each SNP	-	Nino
	- Frequency of each SNP	-	Esther
	- Effect on protein translation	-	Daphne
10	Practise First presentation	15-11-19	Nino and Daphne
11	First presentation	?	Nino and Daphne
12			
13	Arrange end presentation at Naturalis	<u>6-12-19</u>	Nino
14			
15	Start writing the final report:	20-12-19	Everyone
	Summary	-	Nino
	Introduction	-	Daphne
	Research Question	-	Esther
	Methods	-	Rik
	Procedure	-	Nino
	Findings	-	Daphne
	Conclusion	-	Esther
	Discussion	-	Rik
16	Make powerpoint for Final presentation		Esther
	Read and grammarcheck final report		Everyone

17	Practise Final presentation		Nino and Daphne
	Give Naturalis our Final report for feedback		Rik
18	Turn in final report at Hogeschool Leiden	<u>26-1-20</u>	Everyone
	Final presentation	?	Nino and Daphne
19			
20			

Risk Analysis

In order to complete this project as well as possible, it is required to identify the risks of problems that may occur while working on the project. For every potential risk, the group members have thought about a solution or how to prevent a risk. This is done using a risk analysis. The risk factor is calculated by giving a score on a scale from 1 to 5 on the probability that the risk occurs, the impact of a risk and the chance that the risk will be discovered. For example: a score of 1 means that the probability of occurrence of the risk is very low, the impact is not significant and the chance that the risk will be discovered is very high. A score of 5 means that the probability that the risk will occur is likely, the impact is severe and the probability of discovering is minimal. Risks with a risk-factor less than 10 will be ignored. For risks with a risk-factor of 10 or higher, the group needs to come up with preventive measures.

Table 3: The risks with their calculated risk-score (risk-factor (total) = chance that the risk will occur * impact * chance that the risk will be discovered).

Description of the risk	Chance that the risk will occur	Impact	Chance that risk will be discovered	Risk-factor (total)
A computer will break down leading to a loss of data or progress	3	4	1	12
We are not able to perform BLAST searches against the F1	2	3	2	12
A member of the project group is not able to continue BPEXA	1	4	1	4
There are conflicts between group members	3	3	3	27
Members of the group miss deadlines	2	3	1	6
A group member is ill for a long time and can not participate in the project	1	3	1	3
A group member is absent at the presentation(s)	1	4	1	4

A group member is absent more than two times at the group meetings (and has to quit the project)	2	4	1	8
The final report contains more than 5 grammar mistakes per 500 words	2	5	3	30
The group does not start on time working on the products (presentation, final report)	2	4	2	16
Communication does not go well/miscommunication	2	3	3	18

Risks with a risk-factor above 10:

- A computer will break down -> loss of data.
 - Solution: the project group will always work in the drive environment. If a project member is unable to work in this synchronised environment, they will be required to upload their work to the drive on the first opportunity they are presented.
- We are not able to perform BLAST searches against the F1.
 - Solution: the project will move forward, attempting other means of reducing the list of potential candidates. To account for this possible change in approach, this step will be assigned extra time to be finished.
- There are conflicts between group members.
 - Solution: The group will attempt to resolve conflicts internally with an open discussion. Should the discussion not resolve the issue, the matter will be discussed on the weekly meetings with the student guide serving as moderator. If that secondary discussion does not resolve the conflict, the student guide will take charge of the situation.
- The final report contains more than 5 grammar mistakes per 500 words
 - Solution: before turning in any products, an extensive grammar and spelling check will be performed by at least 2 of the project members. The minimum amount of time for this action is 2 days.
- The group does not start on time working on the final products (presentation, final report)
 - Solution: a detailed planning will be made, leaving time between the expected date that products will be finished, and the expected date the products are expected to be finished. This extensive planning step, in combination with time to spare, should mitigate any oversight problems causing the project group to start late.
- Communication does not go well/miscommunication

- Solution: to encourage clear and open communication, the group has assigned one student the task to keep general oversight of activities, and to actively engage in conversation about any potential points of friction. In addition, a discussion point will be made on weekly meetings between the student group and the student guide to discuss any points of friction any member of the group may experience.
- Additional solution: common point of friction is a lack of oversight, and differing points of view on the status of a product. To avoid these points of friction, project members will ensure other project members have reliable access to their work. As such, all project members are required to work in the drive environment whenever possible, or to upload the most recent version to the drive environment whenever they finish working for the day.

Project Boundaries

It is important to define clear boundaries in which the project will be completed. These boundaries will ensure that the project group and the project manager know exactly what to expect from one another.

No lab work:

The student group will not work in a laboratory. Whenever data generated by a laboratory is required, this task will be ignored or outsourced to a team provided by Naturalis. The measure taken is discussed with the primary contact for Naturalis beforehand.

No additional programming languages:

Altogether, the project group has expertise in a variety of programming languages. However, the scope of this project requires only the application of a few specific and powerful programming languages, well-suited to the task. As such, the project group will not use programming languages outside the following list:

- Python 2 and/or Python 3, including the BioPython module.
- BASH
- R

No additional tasks beyond the original scope of the project:

The project is well defined, with clear starting material and a clear objective. Beyond these tasks, as outlined in this document, the project group will perform no additional tasks for the head researchers at Naturalis. Should the work provided by the project group open up additional avenues of research that the head researchers require to be explored, this task may be proposed to the project group. The project group will evaluate the additional burden on time and resources, and accept the task if both time and resources permit the full execution of the additional task in the time frame the project group works for Naturalis.

References

1 University of Illinois Extension: Watch Your Garden Grow

<https://web.extension.illinois.edu/veggies/cabbage.cfm> Accessed September 17, 2019.

2 David vs Goliath: searching for genetic markers underlying growth form, life-history and correlating features using a crossing experiment in cabbage (*Brassica oleracea*).

<https://www.universiteitleiden.nl/en/research/research-projects/science/ibl-david-vs-goliath-searching-for-genetic-markers-underlying-growth-form-life-history-and-correlating-features-using-a-crossing-experiment-in-cabbage-brassica-oleracea>. Accessed September 16th, 2019

3 Assembly *Brassica oleracea* var. *oleracea* (wild cabbage) . NCBI website.

https://www.ncbi.nlm.nih.gov/assembly/GCF_000695525.1/. Accessed September 16th, 2019.