PPP Project Annual Report 2018
The PPP-projects that have been established under the direction of the top sectors must submit an annual report on their technical and financial progress. This format is to be used for reporting the technical progress. A separate format ('PPP final report') is available for PPP-projects that have been completed in 2018.

The annual reports will be published in full on the websites of the TKIs/top sector, excluding the blocks 'Approval coordinator/consortium' and 'Planning and progress'. Please ensure that no confidential matters are left in the remaining blocks.

The PPP Project Annual Reports must be submitted to the TKI's before March 1st 2019. For Wageningen Research this will be coordinated via a central point.

General information	
PPP number	KV 1605-004
Title	A de novo sequencing catalogue of structural variation in different *B. oleracea* morphotypes
Theme	Duurzame Plantaardige Productie
Executive knowledge institution(s)	Wageningen University and Research Plant Breeding
Research project leader (name + e-mail address)	Guusje Bonnema: Guusje.Bonnema@wur.nl Richard Finkers: Richard.finkers@wur.nl
Coordinator (on behalf of private parties)	Henk Huits, Bejo zaden
Government contact person	
Total project size (k€)	894
Address projectwebsite	https://topsectortu.nl/nl/de-novo-sequencing-catalogue-b-oleracea
Start date	1 oktober 2017
End date	31 december 2021

Approval coordinator/consortium *The annual report should be discussed with the coordinator/the consortium. The TKIs appreciate being informed of possible feedback on the annual report.*	
The coordinator has assessed the annual report on behalf of the consortium:	√ approved □ rejected
Possible feedback on the annual report:	

Planning and progress (if there are changes to the project plan, please explain)	
Is the PPP going according to plan?	Content wise yes, however there have been some changes in the scientific staff conducting the project. Theo Borm was first ill and after recovery switched jobs. A fellowship PhD ChengCheng Cai, with expertise in genome assembly, is analysing the data. In 2019 we will get support of a bioinformatics postdoc and technician
Have there been changes in the consortium/project partners?	no

Is there a delay and/or deferred delivery date?	no
Are there any substantive bottlenecks? Provide a brief description	Main challenge was to generate segregating populations, based on four way crosses. This is a challenge, but sofar all is progressing as planned
Are there any deviations from the projected budget?	No

Short content description/aim PPS
What is going on and how is this project involved?
What will be delivered by the project and what is the effect of this?

Brassica oleracea is an economically important crop exhibiting enormous diversity in its appearance (ranging from cauliflower to kohlrabi and from cabbage to kale) and uses (ranging from fodder to vegetables to ornamentals). Despite this enormous diversity, *B. oleracea* truly remains one species - morphotypes can be easily interbred.
There is accumulating evidence that morphotype specific meiotic aberrations happen frequently - based on microscopic observations and skewed segregations, and that there are morphotype specific structural variations – based on observations in sequencing data.
The aim of this project is to generate de novo reference genomes of five homozygous genotypes representing diverse morphotypes (cabbage, cauliflower, kale, kohlrabi and broccoli) and analyse structural variations (SVs) between these genomes. To reveal whether these SVs affect both the rate and the positions of recombination we will develop four way crosses and assess recombination in their progeny. We will also investigate whether the relatively recent genome triplication in the common ancestor of *B. oleracea, B. rapa* and *B. nigra* somehow facilitated these phenomena.
To accomplish this, not only a variety of different sequence data will be generated of selected genotypes representing morphotypes, but also intermorphotypic 'F1's will be intercrossed as parents to create four way mapping populations. These populations will be skim-sequenced, allowing us to accurately anchor sequence scaffolds genetically, and to directly study the effect that structural differences between the genomes may have on suppression of meiotic recombination in the intermorphotypic F1s. Microscopic analyses of meioses in intermorphotypic F1's will provide additional evidence for the presence of structural differences between genomes. Reference genomes will be compared to the reconstructed non-triplicated common ancestor of the Brassiceae, to date rearrangements and to verify if any of the three *B. oleracea* subgenomes is more amenable to rearrangements. The approach in which we combine de novo sequencing, bioinformatics, genetics and cytological analyses is novel, with a potentially large scientific impact as well as an immediate practical impact on breeding strategies and thus crop improvement

Results in 2018/ so far
Give a short description of the high-lights and project deliverable in 2018 / so far

Material development:
The partners in this project, Bejo and ENZA, provided the five inbred lines of a cabbage cauliflower, broccoli, kale and kohlrabi. Bejo successfully generated 10 F1 progenies, representing all possible combinations of two parents. At this moment, February 2019, these plants are almost flowering and will be intercrossed to produce progeny of four way crosses. These seeds will be available in autumn.
Seedlings of the 5 parental genotypes have been sown in autumn 2018 for DNA isolations. This DNA was sent to Genomescan to generate Illumina and PacBio sequence data, and for kale and white cabbage also 10xdata. Presently seeds are sown again for HMW DNA isolation for Nanopore sequencing and for isolation of mRNA of diverse organs, developmental stages and treatments (stresses). RNA seq data will be generated to support the annotation.
Cytogenetic analyses:
From several accessions representing a number of morphotypes , flower buds from field grown plants have been harvested to study meiotic behaviour (Alexandre Pele). This resulted in remarkable observations. Accessions of most morphotypes displayed not the expected 9 bivalents at diakenesis, however this was dissolved in metaphase I and all gametes obtain 9 chromosomes. However in both kohlrabi and cauliflower, univalents and multivalents persist until the end of metafase I resulting in 5-10% aneuploid gametes. These data need to be considered when studying meioses of the F1 plants from biparental crosses. And as a result it is advised to do cytological observations at telophase I and II instead of in metaphase I.

Bioinformatic analyses (ChengCheng Cai):
We have generated Illumina and Pacbio sequencing data for all of the five morphotypes, with an average depth of ~228X Illumina and ~9.38X Pacbio data. We filtered low quality reads for Illumina data and trimmed adaptors for both Illumina and Pacbio data before genome assembly (Table 1 and 2). Contigs were constructed with the combination of these two types of clean reads, with contig N50 ranging from 191Kb to 666Kb. The quality of these assembled contigs was evaluated using BUSCO, with a percentage of complete BUSCOs ranging from 90.80% to 96.50% (Table 3).
Additional 10X genomics data were generated for kale (62.92Gb) and white cabbage (66.29Gb). These data were used to construct scaffolds, with a scaffold N50 of 626Kb for kale and 520Kb for white cabbage (Table 4).

Number of delivered products in 2018 *(in an appendix, please provide the titles and/or description of the products or a link to the products on public websites)*			
Academic articles	Reports	Articles in journals	Introductions/workshops
0	0	0	0
Titles/ description of the most important products in 2018 (5 at max) and their target group			

Appendix: Names of the products or a link to the products on a public website including the link to the project summary on Kennisonline

Table 1. Summary of Illumina sequencing data for the five morphotypes.

morphotypes	**raw data (Gb)**			**clean data (Gb)**		
	450bp	**600bp**	**total**	**450bp**	**600bp**	**total**
Broccoli	54.74	91.6	146.34	52.75	87.92	140.67
Cauliflower	48.21	115.51	163.72	46.35	111.1	157.45
Kale	42.14	119.23	161.37	40.87	110.67	151.54
Kohlrabi	56.52	81.64	138.16	54.03	79.1	133.13
White	61.97	70.02	131.99	59.77	66.48	126.25

Table 2. Statistics of Pacbio subreads for the five morphotypes.

morphotypes	ReadNum	BaseNum (Gb)	Mean (bp)	MAX (bp)	N50 (bp)	N90 (bp)	MIN (bp)
Broccoli [a]	748,966	8.31	11,099	129,789	20,268	6,211	50
Broccoli [b]	754,500	8.31	11,017	129,789	20,118	6,149	50
Cauliflower [a]	547,866	4.98	9,091	93,109	20,551	4,059	50
Cauliflower [b]	551,144	4.98	9,036	88,836	20,398	4,021	50
Kale [a]	662,152	7.16	10,811	123,127	17,504	6,252	50
Kale [b]	666,843	7.16	10,734	123,127	17,408	6,193	50
Kohlrabi [a]	400,833	4.23	10,561	119,486	20,991	5,599	50
Kohlrabi [b]	403,810	4.23	10,483	119,486	20,820	5,541	50
White [a]	636,310	5.83	9,159	88,263	19,834	4,087	50
White [b]	640,295	5.83	9,102	88,263	19,686	4,048	50

a: subreads before removing adaptors
b: subreads after removing adaptors

Table 3. Contig statistics for the five morphotypes.

Statistics	Broccoli	Kale	White	Cauliflower	Kohlrabi
Max (bp)	5,754,080	4,835,892	1,916,164	1,889,646	1,574,289
N10 (bp)	2,137,072	1,888,001	887,619	665,318	609,543
N20 (bp)	1,437,074	1,102,917	604,713	476,461	397,267
N30 (bp)	1,090,303	845,033	444,167	354,775	308,102
N40 (bp)	848,745	655,605	353,980	275,180	242,397
N50 (bp)	666,345	488,747	276,005	220,105	191,158
N60 (bp)	501,331	384,221	218,002	167,640	147,061
N70 (bp)	378,768	280,866	165,432	123,959	110,209
N80 (bp)	263,918	190,922	114,713	88,709	82,307
N90 (bp)	140,041	96,293	67,085	55,791	50,941
Min (bp)	2,773	2,965	4,408	4,254	4,603
Contig Number	2,181	2,824	3,563	3,967	4,327
Contig Length (bp)	546,570,702	547,665,109	523,381,064	488,254,544	490,044,588
GC content (%)	36.94	36.77	36.80	36.77	36.88
PacBio data depth (X)	14.57	10.85	10.15	8.20	7.08
Complete BUSCO	96.50%	96.40%	94.10%	91.30%	90.80%

Table 4. Scaffold statistics for kale and white cabbage genomes.

Statistics	Kale		White	
	Contig	Scaffold	Contig	Scaffold
Max (bp)	4,835,892	4,835,892	1,916,164	2,812,867
N10 (bp)	1,888,001	2,217,401	887,619	1,561,788
N20 (bp)	1,102,917	1,355,674	604,713	1,043,601
N30 (bp)	845,033	984,645	444,167	822,097
N40 (bp)	655,605	786,397	353,980	658,905
N50 (bp)	488,747	625,566	276,005	519,967
N60 (bp)	384,221	455,248	218,002	396,873

N70 (bp)	280,866	348,332	165,432	284,518
N80 (bp)	190,922	238,748	114,713	201,050
N90 (bp)	96,293	120,927	67,085	110,193
Min (bp)	2,965	2,965	4,408	4,408
Total Number	2,824	2,515	3,563	2,394
Total Length (bp)	547,665,109	549,425,552	523,381,064	530,869,688
Gap Length (bp)	0	2,179,351	0	7,548,662