



PPP Annual Report 2019

PPP projects which are under supervision of the "Topsectoren" must report annually on the scientific content and financial progress. This form is used to report the progress of the content of the project. PPP projects that finish in 2019 should make use of a different form: "PPP-final report."

The annual report will be published on the TKI / topsector website. Therefore, please ensure that there is no confidential information in the annual report.

The PPP-annual report must be sent, at the latest, by the 1st of March 2020 to the "TKI's": info@tkitu.nl or info@tki-agrifood.nl. For Wageningen Research, the report has to be sent to the "Topsector secretary" of your respective institute.

General information	
PPP-number	KV 1605-004
Title	A de novo sequencing catalogue of structural variation in different <i>B. oleracea</i> morphotypes
Theme	Duurzame Plantaardige Productie
Implementing institute	Plant Breeding, Wageningen University and Research
Project leader research (name + e-mail address)	Guusje Bonnema: Guusje.Bonnema@wur.nl Richard Finkers: Richard.finkers@wur.nl
Coordinator (on behalf of private partners)	Dr. Henk Huits, Bejo zaden
Project-website address	https://topsectortu.nl/nl/de-novo-sequencing-catalogue-b-oleracea KOL: https://www.wur.nl/nl/Onderzoek-Resultaten/Onderzoeksprojecten-LNV/Expertisegebieden/kennisonline/A-de-novo-sequencing-catalogue-B.oleracea.htm
Start date	1 oktober 2017
Final date	31 december 2021

Approval by the coordinator of the consortium

The annual report must be discussed with the coordinator of the consortium. The "TKI's" appreciate additional comments concerning the annual report.

Assessment of the report by the coordinator on behalf of the consortium:	<input checked="" type="checkbox"/> Approved <input type="checkbox"/> Not approved
Additional comments concerning the annual report:	A few small setbacks but the project is well on its way and will deliver very interesting and useful results for the companies

Summary of the project

Problem definition	<i>Brassica oleracea</i> is a species that includes many diverse economically important crops with extreme phenotypes, like cabbages, kohlrabi, cauliflower, broccoli and kales. In 2014 the first reference genomes were published, of a white cabbage and a rapid cycling kale like plant, which was recently followed by a broccoli genome. The availability of these reference genomes, plus resequencing data generated in a previously funded 1000 genome TKI project, illustrated that the genomes of these morphotypes vary considerably. In addition, inter-crossing morphotypes to transfer traits across crops is often hampered by reduced recombination. Structural variations between the genomes of these different crops could influence recombination behaviour, but this could not be studied due to a lack of high quality de novo reference genomes and appropriate segregating populations.
--------------------	--

Project goals	<p>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 <i>B. oleracea</i>, <i>B. rapa</i> and <i>B. nigra</i> somehow facilitated these phenomena.</p> <p>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 re-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 <i>B. oleracea</i> 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</p>
---------------	--

Results	
Planned results 2019	<p>For 2019, the plan was to generate nanopore and bionano data and to integrate this with Illumina and PacBio data to assemble the genomes of the five different morphotypes.</p> <p>In addition, we planned to generate RNA-seq data to support the genome annotation. For material development, the plan was to inter-cross the different F1 plants to generate so called four-way progeny, being progeny of crosses with 4 different parental genomes. We aimed for five four-way crosses, with each cross having 4 different parents, plus their reciprocals and to sow these to collect leaflets for DNA sequencing.</p> <p>In addition, we intended to find a means to create around 1000 genome libraries for resequencing of the four-way cross progenies for an affordable budget.</p>
Achieved results 2019	<p><i>Material development:</i> 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, and in spring 2019, these F1 plants are intercrossed to produce progeny of four-way crosses. These seeds were available summer 2019. 8 of the 10 progenies contained sufficient seeds, however crosses with one parent a broccoli x cauliflower F1, failed. These will be redone spring 2020.</p> <p>Progenies of all 8 four-way crosses (4 combinations, with each reciprocal) have been sown by ENZA and seedling DNA was delivered January 2020. We contacted Igenomix and decided to use their pooled library system to generate template for resequencing the four-way cross progenies.</p> <p>In 2019, all 5 morphotypes are grown in a greenhouse in Wageningen and different tissues of plants grown under control - or stress conditions are harvested, RNA is isolated and pooled for RNA-seq. Data were generated December 2019</p> <p><i>Bioinformatic analyses (ChengCheng Cai):</i> Nanopore data (35-60x) have been generated of the five morphotypes and together with PacBio and Illumina data this resulted in updated contigs. In addition, we planned to generate Bionano data to link contigs and generate scaffolds. This service</p>

	<p>was seriously delayed, and instead of getting the data august 2019, we received these around December. The hybrid assemblies were of high quality for kale, cabbage, broccoli and kohlrabi (see table 1 below) , however cauliflower data were insufficient and need to be redone. We provided new plant materials and await the results.</p> <p><i>Table 1: Bionano hybrid scaffolding.</i></p> <table border="1"> <thead> <tr> <th>Statistics</th> <th>Broccoli</th> <th>Kale</th> <th>Kohlrabi</th> <th>White</th> </tr> </thead> <tbody> <tr> <td>Contig N50 (Mb, input)</td> <td>9.34</td> <td>13.26</td> <td>8.02</td> <td>6.32</td> </tr> <tr> <td>Scaffold N50 (Mb, output)</td> <td>30.61</td> <td>30.34</td> <td>30.55</td> <td>33.29</td> </tr> <tr> <td>Scaffold + Unscaffolded N50 (Mb, output)</td> <td>30.61</td> <td>30.34</td> <td>30.55</td> <td>31.65</td> </tr> <tr> <td>Longest contig (Mb, input)</td> <td>30.85</td> <td>35.95</td> <td>25.92</td> <td>28.55</td> </tr> <tr> <td>Longest scaffold (Mb, output)</td> <td>47.13</td> <td>48.71</td> <td>46.11</td> <td>48.33</td> </tr> <tr> <td>Contig count (input)</td> <td>375</td> <td>389</td> <td>332</td> <td>373</td> </tr> <tr> <td>Scaffolded count (output)</td> <td>41</td> <td>37</td> <td>40</td> <td>37</td> </tr> <tr> <td>Scaffolded + Unscaffolded count (output)</td> <td>233</td> <td>253</td> <td>164</td> <td>178</td> </tr> <tr> <td>Contig size (Mb, input)</td> <td>553.16</td> <td>560.87</td> <td>551.72</td> <td>558.01</td> </tr> <tr> <td>Scaffolded size (Mb, output)</td> <td>547.68</td> <td>555.95</td> <td>549.13</td> <td>557.60</td> </tr> <tr> <td>Scaffolded + Unscaffolded size (Mb, output)</td> <td>559.21</td> <td>569.41</td> <td>558.49</td> <td>566.93</td> </tr> </tbody> </table> <p>In addition, repeat annotations were finished for the four morphotypes (~50% TEs for each genome). The gene prediction is ongoing and RNA-seq data are being analysed to support gene structure. The annotation results will be ready spring 2020.</p>	Statistics	Broccoli	Kale	Kohlrabi	White	Contig N50 (Mb, input)	9.34	13.26	8.02	6.32	Scaffold N50 (Mb, output)	30.61	30.34	30.55	33.29	Scaffold + Unscaffolded N50 (Mb, output)	30.61	30.34	30.55	31.65	Longest contig (Mb, input)	30.85	35.95	25.92	28.55	Longest scaffold (Mb, output)	47.13	48.71	46.11	48.33	Contig count (input)	375	389	332	373	Scaffolded count (output)	41	37	40	37	Scaffolded + Unscaffolded count (output)	233	253	164	178	Contig size (Mb, input)	553.16	560.87	551.72	558.01	Scaffolded size (Mb, output)	547.68	555.95	549.13	557.60	Scaffolded + Unscaffolded size (Mb, output)	559.21	569.41	558.49	566.93
Statistics	Broccoli	Kale	Kohlrabi	White																																																									
Contig N50 (Mb, input)	9.34	13.26	8.02	6.32																																																									
Scaffold N50 (Mb, output)	30.61	30.34	30.55	33.29																																																									
Scaffold + Unscaffolded N50 (Mb, output)	30.61	30.34	30.55	31.65																																																									
Longest contig (Mb, input)	30.85	35.95	25.92	28.55																																																									
Longest scaffold (Mb, output)	47.13	48.71	46.11	48.33																																																									
Contig count (input)	375	389	332	373																																																									
Scaffolded count (output)	41	37	40	37																																																									
Scaffolded + Unscaffolded count (output)	233	253	164	178																																																									
Contig size (Mb, input)	553.16	560.87	551.72	558.01																																																									
Scaffolded size (Mb, output)	547.68	555.95	549.13	557.60																																																									
Scaffolded + Unscaffolded size (Mb, output)	559.21	569.41	558.49	566.93																																																									
Planned results 2020	<p>For 2020 we anticipate to finalise the genome assemblies to obtain chromosome-level reference genomes (decide whether we have sufficient omics data, including genetic mapping data to anchor scaffolds, or we will still need to add HiC data) and genome annotation.</p> <p>Main challenge for 2020 is to reveal Structural Variations between our reference genomes and to generate genetic maps and linking these to respective reference genomes or generate a pan genome for this purpose. We plan to have developed analysis tools in 2020 and make progress in the analyses (but anticipate that this continues in 2021).</p>																																																												

<p>Deliverables/products in 2019 (provide the titles and /or a brief description of the products/deliverables or a link to a website.</p>
<p>Scientific articles: PhD ChengCheng Cai wrote first draft of a paper describing genetic diversity based on SBG data of 940 genomes (previous TKI project). This is relevant, as this increases insight in relatedness of the morphotypes we work on in this TKI project, and the resequencing data of that project will be included in this running project. We aim to submit paper before summer 2020.</p>
<p>External reports:none</p>
<p>Articles in professional journals/magazines:none</p>
<p>(Poster) presentations at workshops, seminars, or symposia. ChengCheng Cai attended the PAG meeting in San Diego January 2020. He presented the poster: Genome assembly of five different B.oleracea morphotypes.</p>
<p>TV/ radio / social media / newspaper:</p>
<p>Remaining deliverables (techniques, devices, methods, etc.):none</p>

