



PPP-final report

PPP projects which are under supervision of the "Topsectoren" must file a final report concerning the total project period. This form is used to report the content of the project. There is a separate form for the financial reporting.

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

The PPP-final 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 1604-002
Title	<i>The Capsicum Genome Initiative (CGI)</i>
Theme	Sustainable Plant Production
Implementing institute	Cluster Applied Bioinformatics, department of Bioscience, Wageningen University and Research
Project leader research (name + e-mail address)	Dr. ir. Sander Peters (sander.peters@wur.nl)
Coordinator (on behalf of private partners)	Dr. Aat Vogelaar (a.vogelaar@rijkswaan.nl)
Project-website address	https://www.wur.nl/nl/Onderzoek-Resultaten/Onderzoeksprojecten-LNV/Expertisegebieden/kennisonline/KV-1604-002-Capsicum-Genome-Initiative.htm
Start date	31-03-2018
Final date	01-03-2020

Approval by the coordinator of the consortium

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

Consortium

Mention any changes in the composition of the project partners:	No changes
---	-------------------

Summary of the project

Problem definition	This TKI project started in March 2018 applying exiting new genomics technologies, including Genome Mapping, 10X Genomics, and state-of-the-art NGS. We aim to reconstruct and annotate new reference genomes of wild Capsicum species and mine the genetic diversity of gene bank
--------------------	--

	<p>collections such as from the Centre of Genetic Resources the Netherlands (CGN) and Katsetsart University Thailand. With the latest technical improvements, such as PacBio 'Sequel' technology and BioNano 'Saphyr', large insert sequencing for genomes such as <i>Capsicum</i> becomes economically feasible, while assessment of genetic diversity of a large panel of genomes can very well be supported by Illumina sequencing technology. The recently developed technology from 10X is an innovative technology that will be of further support for remapping to uncover inaccessible parts of the genome as well as megabased diploid assemblies preserving phasing details such as SNVs, indels, and structural rearrangements. In addition, genome wide physical profiling of reference genomes with BioNano technology has proven successful for extended scaffolding in our lab, solving many assembly problems for complexed sequences of crop genomes. The high quality <i>Capsicum</i> reference genome will serve as a golden standard to assess the genetic diversity of a large panel of pepper accessions from genebank collections such as from CGN (Centre for Genetic Resources the Netherlands) and Katsetsart University. The knowledge gained in the Capsicum Genome Initiative will allow breeders to have insight in genetic and genome structure diversity that is necessary to advance on precision breeding, reducing their R&D costs and better respond to economic, environmental, social and health issues inherent to agricultural crop production.</p>
Project goals	<p>The overall aim of the Capsicum Genome Initiative is to disclose the genetic information that underlies economically important traits in cultivated crops and related wild species of pepper (<i>Capsicum annuum</i>). This precompetitive project therefore has both applied as well as of fundamental importance. We will integrate sequence information and high throughput genome wide physical mapping. The deliverables that are anticipated within the project are; (1) list of selected <i>Capsicum</i> accessions; (2) all raw sequence data; (3) all genome mapping and gDNA crosslinking data. (4) all processed/cleaned sequence data; (5) read alignments of accessions against closest phylogenetic related reference genome; (6) contig assemblies and scaffold sequences of reference genomes; (7) SNP list from accessions compared to closest related reference genome; (8) phylogenetic analysis for all accessions based on whole genome SNPs (9) Genome structure comparisons based on the optical mapping data output; (10) Seed materials available under SMTA by CGN (upon request for seed lots from CGN, a small handling fee will be requested).</p>

Results					
Planned results in the original project plan					
	Deliverable/Milestone	Description	WP no.	Yr/mth	
	Consortium				
	D0.1	Consortium agreement	0	2017/09-2018/01	
	M0	TKI grant	0	2018/01	
	Accessions, DNA				
	D1.1	Selection and list of accessions	1	2018/03	
	D1.2	gDNA for 3 reference species and 100 (max) accessions	1	2018/08	
	M1	Partner agree to list of accessions	1	2018/03	
	M2	gDNA QC sequencing and genome mapping grade	1	2018/05	
	De novo sequencing, resequencing, optical mapping				
	D2.1	All PacBio sequence data 3 ref genomes	2	2018/10	
	D2.2	All Illumina sequence data 100 (max) draft accessions	2	2018/12	

	D2.3	Optical map for 3 reference genomes	2	2018/12
	D2.4	Dovetail Chicago lib sequences -> 10X lib sequences	2	2018/17
	M3	PacBio Sequencing completed	2	2018/12
	M4	Illumina sequencing completed	2	2018/18
	M5	Irys runs for genome mapping completed	2	2018/12
	Bioinformatics			
	D3.1	NGS hybrid assembly 3 ref genomes	3	2019/05
	D3.2	<i>De novo</i> genome map constructions 3 ref genomes	3	2019/06
	D3.3	Hybrid NGS Genome Map assembly 3 ref genomes	3	2019/07
	D3.4	Advanced genome reconstruction	3	2019/10
	D3.5	Read Mapping & SNP calling	3	2019/10
	D3.6	SV & colinearity assessment	3	2019/10
	D3.7	FTP service	3	2018-2019
	M6	Hybrid NGS assembly completed	3	2019/05
	M7	Hybrid NGS Genome map assembly completed	3	2019/07
	M8	Structural variation analysis completed	3	2019/11
	Communication & Datamanagement			
	D4.1	Project meetings	0	2018/02, 2018/ 2019/11
	D5.2	Final report 7 Publication of the CGI project	0	2019/12
	D5.3	Data distribution completed	0	2019/12
	M10	Consortium paper	0	2020
Achieved results	<ul style="list-style-type: none"> • Regeneration of 114 <i>Capsicum</i> accessions, including accessions from the Centre of Genetic Resources, The Netherlands (CGN) and Katsetsart University (Th) genebank collections • Phenotypic characterization with ECPGR descriptors and photographic images • Seed collection and distribution to the project partners under SMTA (22 accessions that first did not set seed have been regrown) • Visual and laboratory sanity check for plant material and seed material • Seed materials available under SMTA and distributed by CGN • DNA isolation and sequence library construction from 103 selected accessions including (3 accessions for de novo sequencing and 100 accessions for draft sequencing) • QC check for DNA preps and library preps-> Go decision • Chloroplast contamination ckeck is lower than 6% -> go decision • IsoSeq library preparation from 4 tissues (leaves, roots, flowers, fruits) for 3 de novo accessions and QC check -> go decision • Pacbio Sequel data production meeting the output requirement of 150Gb -> go decision • Genome mapping for <i>C. chacoense</i> completed • Genome mapping for <i>C. annuum</i> Early Cal Wonder completed 			

	<ul style="list-style-type: none"> • Genome mapping for <i>C. galapagoense</i> completed • NGS, de novo genome map and hybrid assembly/scaffolding for <i>C. chacoense</i>, <i>C. annuum</i> and <i>C. galapagoense</i> completed • NGS, de novo genome map and hybrid assembly/scaffolding for <i>C. galapagoense</i> completed • NGS, de novo genome map and hybrid assembly/scaffolding for <i>C. annuum Early Cal Wonder</i> completed • Pacbio IsoSeq sequencing for <i>C. annuum</i>, <i>C. chacoense</i> and <i>C. galapagoense</i> completed • Construction of chromosome scaffolds completed • BUSCO analysis for 3 reference genomes to verify the completeness of the reference genomes have been carried out -> QC check approved -> Go decision for structural annotation • Sequence mapping and SNP calling for 100 Capsicum draft accessions against 3 reference genomes using GATK toolkit have been completed. • Mapping of transcriptome data to 3 reference genome completed • Gene model building for structural annotation have been carried out. • Gene calls for 3 reference genomes have been produced. • Structural annotation has been completed, results are output in GFF format. • Phenotypic descriptors, photographic images, genome and transcriptome sequence resources, assemblies, read mapping, SNP calls, and structural annotation data have been made available either on HD or on the sftp resource. • This project report • The end meeting in planned for March – April 2020
<p>Explanati on of changes relative to the project plan</p>	<p>Based on these findings we have proposed to use 10X + BioNano Genomics scaffolding for Capsicum <i>de novo</i> reference genome reconstruction -> favoured by all project partners.</p>

What was the added value created by the project for:	
<p>Participating "Knowledge Institutes" (scientific, new technologies, collaboration)</p>	<ul style="list-style-type: none"> • The project created a new sequence resource for follow up projects.

	<ul style="list-style-type: none"> • We have established a collaboration with Cornell University. Based on the assembly comparisons of HiC and BioNano scaffolds from <i>S. lycopersicum</i> cv. Heinz that have been cross-checked with a third platform technology using 10X Genomics we have detected and analysed inconsistencies between HiC and BioNano scaffolds that are resolved in favour of BioNano genome scaffolds. • We plan to draft a consortium paper describing the genetic diversity and phylogenetic relationships of the accessions that have been profiled in this project. A description of target genes related to economically important traits in pepper is anticipated.
Participating private partners (practical application of the results, within which period of time?)	
Society (social, environment, economy)	<p><u>Economic impact</u> The implication of the genome bioinformatics technology developed in this project is that genome structures of complex crop genomes have been established with unsurpassed precision. This allows assessment of genome wide structural variation and identification of structurally compatible donor and recipient genomes for breeding. The knowledge gained in this project allows breeders to; (1) advance on sequenced based precision breeding; (2) reduce their R&D costs; (3) and better respond to economic, environmental, social and health issues inherent to agricultural crop production.</p> <p><u>Scientific impact</u> Genome structure analysis has mainly been dependent on genome assemblies and construction of traditional genetic and physical maps, which is a labour intensive and very time consuming approach. Furthermore, genome sequence assemblies and maps are often inaccurate due to complex repeat sequences and genetic and physical mapping ambiguities of markers. In this project long-range genome mapping information has been cost-effectively produced with innovative technology, aimed at improving contiguity of genome sequence assemblies, identifying structural variation, and haplotyping. The resource provides valuable insight to develop even more effective strategies for future genome projects.</p> <p><u>Societal impact</u> The production of new innovative crops is a knowledge-intensive process requiring cooperation between industry, knowledge institutes and government. The innovation targeted in this project has strengthened the competitiveness of the Dutch breeding industry, but will promote employment in the Research and Agricultural</p>

	sectors thereby contributing to wealth and well-being in and outside the Netherlands.
Possibly other stakeholders (spin-offs)	

Follow-up	
Did the PPP result in one or more patents (first filings)?	No, the project is entirely precompetitive
Are there any follow-up projects planned? If yes, explain. (Contract research resulting from this project, additional funding, or new PPP projects)	Yes, at the end meeting of this project a plan for a follow up project will be presented

Deliverables/products during the entire course of the PPP (provide the titles and/or a brief description of the products/deliverables or a link to a website.)
<p>There is a grace period after the project finishes on March 1st 2020 as described in the Consortium Agreement. Until that time project results will not be shared with the public. Thereafter a scientific paper for this project is foreseen.</p> <p><u>Scientific articles:</u></p>
<u>External reports:</u>
<u>Articles in professional journals/magazines:</u>
<p><u>(Poster) presentations at workshops, seminars or symposia.</u></p> <p>The capsicum project is presented at the Centre of Molecular Biological Research (CBMSO) in Madrid</p>
<u>TV/ radio / social media / newspaper:</u>

Remaining deliverables (techniques, devices, methods, etc.):

- Intermediate results have been presented and discussed in the CGI progress meeting in November 28, 2018 in Wageningen.
- The Kennis on line link describing the project can be found here:

<https://www.wur.nl/nl/Onderzoek-Resultaten/Onderzoeksprojecten-LNV/Expertisegebieden/kennisonline/KV-1604-002-Capsicum-Genome-Initiative.htm>