

HIP-BB1.3 'Molecular markers for maturity and yield'

Project lead: Dr. Richard Finkers, Dr. Christian Bachem, Wageningen U&R, Plant Breeding

Highlights. A first prototype has been developed showing the sequence organisation of a complex locus over more than one assembled genome (Figure 1). This prototype has been developed by the close interaction of the bioinformatics PhD student and molecular focuses PostDoc on the project by developing a specification document in close collaboration with their respective PI's. After some background reading, a preliminary Functional Specifications document was drafted for development of an online tool that will allow non-technical staff to investigate the haplotype diversity of loci in potato using various datasets. At this time, the stSP6A and stSP3D loci were also established as the areas in which to focus our efforts (though the developed tool should work for any loci), with stSP6A linked to initiating tuberization and stSP3D linked to initiating flowering.

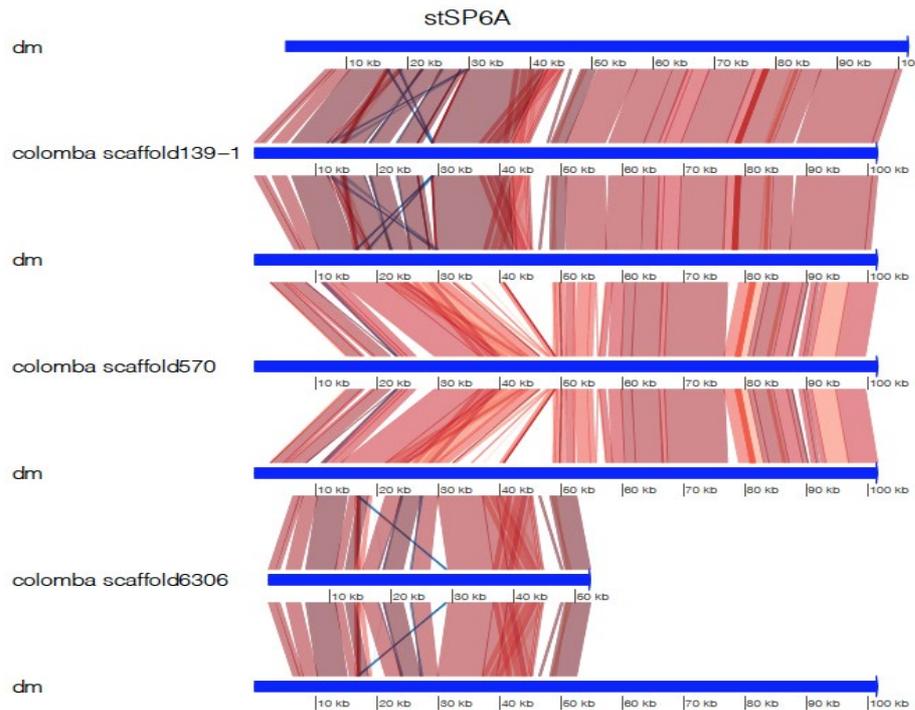


Figure 1. Sequence based organisation of four scaffolds of Columba vs the genome reference sequence of DM. scaffold570 has deletions vs the genome reference. Scaffold 6306 is shorter, either due to a large structural variation, or because the scaffold is broken in the assembly.

As the technical requirements of the tool were refined, preliminary testing and validation of some of the underlying packages also began. Time was spent investigating and comparing the available tools and workflows for on-the-fly haplotyping polyploid WGS data. These tools included WhatsHap, SDHap and HapTree and validated in close collaboration with the PostDoc. As some of the most promising tools under very active development, a final decision on the best workflow will be made in early 2020.

Time was also spent on prototyping how to display and compare the various haplotypes in an informative way. Scripts were written using the R package genoPlotR and applied to 6 potato varieties that were already haplotyped by NRGene. From this synteny images were produced that showed the similarity between each of the haplotypes and a reference sequence (DM) across various distances, centered around the gene of interest. Attached you will find one of these synteny plots for reference. These plots were then used by other team members to begin investigating the effects of each variety's combination of haplotypes on initiating flowering and tuberization.

Bottlenecks: Het project heeft een langzame start gekend door ziekte van sleutel betrokkenen. Dit is inmiddels opgevangen, en budget is al eerder doorgeschoven. Hierdoor zijn er geen verdere knelpunten.

Planning: In 2020, zal de in het specificatie document vastgelegde specificaties verder uitgewerkt worden tot een tool, waarmee de variatie in de verschillende genen automatisch bekeken kan worden. En daardoor bijdragen aan de uiteindelijke deliverable, welke gefocust is op het kunnen bekijken van variatie in grote sets van resequencing data en geassembleerde genomen.

Products: Door het gereedkomen van het specificatie document zijn we op schema.