Marginal lands for Growing Industrial Crops

D3.3– Catalogue of universal QTLs (QTLome) for fiber quality

Due date of deliverable: M42
Actual submission date: M52

Lead beneficiary

Wagenigen University, Netherlands

Responsible Author

Luisa Trindade WU

Additional Authors

Type			**Dissemination Level**		
R	Document, report	☒	**PU**	Public	☒
DEM	Demonstrator, pilot, prototype	☐	**CO**	Confidential, only for members of the consortium (including the Commission Services)	☐
DEC	Websites, patent fillings, videos, etc.	☐			
OTHER		☒			

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement No. 727698.

Description

D3.3 is the outcome of Sub-task 3.2.1 of the MAGIC proposal, which entails the "identification of universal markers for fibre quality breeding using forward genetics". This task was achieved through the creation of a catalogue of conserved genomic loci that control fiber (cell wall) quality across model (biomass) crops and orphan species suited for cultivation on marginal lands. The catalogue was constructed by projecting known loci that control fibre/biomass quality in model plant species to other (orphan) crops through the use of genome synteny (i.e. conserved gene presence and order across multiple genomes; see "Strategy" section). Because of their genomic conservation among multiple crops, the loci identified represent both universal genomic regions controlling fiber/biomass quality across plant species, as well as universal markers, whose allelic variation in novel species adapted for marginal lands can be assessed quickly through targeted sequencing technologies at the moment of performing breeding research. The latter aspect also ensures that the universal QTLs identified in Subtask 3.2.1 can significantly speed up the improvement of novel promising crops for marginal lands, by skipping (part of) the cost- and time-consuming pre-breeding research to characterize the genetic architecture of traits of interest. The catalogue of universal QTLs is publicly stored at a safe data repository of the Dutch university system, and can be accessed via this link: https://figshare.com/s/d418ac8dc0681bff4f09.

Strategy

The first step in the realization of the catalogue of universal quantitative trait loci (QTLs) for fiber (cell wall) quality was the collection of known cell wall QTLs mapped across independent studies from scientific literature. In total, 610 cell wall QTLs from 8 species that span 5 different plant families (*Arabidopsis thaliana*, *Populus trichocarpa*, *Eucalyptus grandis*, *Glycine max*, *Miscanthus sinensis*, *Sorghum bicolor*, *Zea mays*, and *Oryza sativa*) were retrieved. These QTLs control several different traits related to cell wall quality and have been mapped to their physical genomic position. As amply reported in scientific literature, both fibre and biomass quality essentially depends on the composition/quality of plant cell walls, which is the reason why cell wall quality QTLs were selected as a first step in the activities of Subtask 3.2.1.

The second step in the realization of the catalogue consisted in the identification of all the syntenic regions (i.e. regions showing conservation of gene type and gene order between different species) across a set of 200 angiosperm genomes retrieved from online databases. These genomes represent angiosperm diversity and contain a high number of promising species for marginal lands, including 19 of the crops included in the "MAGIC crops" database. This step was performed using

bioinformatic tools for large-scale synteny computation, such as the Diamond and the MCScanX programs.

The third step in the realization of the catalogue consisted in the creation of a database of functionally-characterized cell wall genes from *Arabidopsis thaliana* and *Zea mays* (model species for plant and biomass research, respectively), to be used for the identification of candidate cell wall genes within the universal QTL loci. These genes were retrieved from scientific literature, and in total ~2000 cell wall genes have been identified. The identified genes have then been characterized in terms of sequence and protein domain(s) variation, and used as "seed" genes in a search for their homologs across all the 200 angiosperm genomes included in this study. This gene search was conducted using different bioinformatic platforms, like BLAST and HMMER. As result, ~340000 cell wall genes have been identified across the 200 angiosperm genomes.

After the completion of the three steps above, the 610 cell wall QTLs have been mapped onto the identified syntenic regions of the 200 angiosperm genomes. In this way, the genes from all the species syntenic to cell wall QTL regions have been identified and connected together in a synteny network. Finally, the 340000 cell wall genes have also been mapped on the synteny network, to identify the cell wall genes falling in regions syntenic to known cell wall QTLs. The synteny network just described has then been filtered to identify groups of genomes (species) among which high and low-fragmented synteny of the initial cell wall QTLs was taking place. In this way, several subnetworks restricted to specific groups of species have been identified, which have then in turn been clustered into universal "meta-syntenic" cell wall QTLs. These universal QTLs constitute the catalogue included in this deliverable and a precious source of information for breeding research in biomass crops.

The catalogue: universal QTLs and markers for fibre and biomass quality

The excel file attached to this report (https://figshare.com/s/d418ac8dc0681bff4f09) constitutes the core of the deliverable 3.3, and reports all the universal "meta-syntenic" QTLs identified. In total, 159 universal "meta-syntenic" QTLs have been detected across 200 angiosperm genomes and by using an initial set of 610 cell wall QTLs. These universal QTLs are grouped by QTL ID in column A of the excel file and contain a total of 408818 genes from 74 different plant species, including several biomass (orphan) crops with potential for cultivation on marginal lands. Of these genes, 25571 are cell wall genes, which have been identified starting from the set of functionally-characterized cell wall genes from *Arabidopsis thaliana* and *Zea mays*. These cell wall genes include critical regulators and structural elements at the basis of cell wall biosynthesis, and are therefore good candidates at the basis of the variation in cell wall quality detected in the plant populations where the initial 610 cell wall QTLs were mapped. Overall, all the meta-syntenic QTLs and all the candidate

cell wall genes included in the catalogue represent respectively the “universal” QTLs and the universal candidate genes at the basis of fibre and cell wall quality, whose detection was the main goal of Subtask 3.2.1.

For each gene included in the catalogue of universal QTLs, the catalogue indicates whether the gene is a cell wall gene and, if yes, its broad functional role within cell wall synthesis (i.e. cellulose synthesis, hemicellulose synthesis, pectin synthesis, lignin synthesis, transcription factor), as well as the specific function of each gene within each broad cell wall biosynthetic pathways (columns G and F, respectively). Moreover, a unique Gene ID and the plant species of each gene in the universal QTLs are also reported (columns B and C, respectively). Thanks to this information, the sequences of interesting candidate genes within universal QTLs from target species that are under improvement for cultivation on marginal lands can be easily retrieved from online genetic repositories (e.g. NCBI, Phytozome, Gramene, TAIR,…). In turn, gene sequences can be used for the construction of primers to be used for the targeted sequencing of the candidate genes in extant populations of crops of interest. In this way, the candidate cell wall genes at the basis of fibre and cell wall quality represent actual universal molecular markers that can be used to directly assess the allelic variability and the breeding potential of target conserved loci in novel crops. This is critical, as allows for a significant shortening of the breeding cycles needed to establish new commercial varieties of orphan crops suitable for marginal lands. In addition, the ongoing dropping of genotyping costs and improvement of sequencing technologies will make the procedure just described easily affordable for any plant species and for any research institute in a very near future, representing an important novel tool for the community of biomass crops breeders.