9. The Metabolic Model of the *Quercus suber* (Cork Oak Tree) Leaf

Hüseyin DEMIRCI[1], Oscar DIAS[1], Inês CHAVES[2], Célia Miguel[2], Miguel ROCHA[3], Isabel ROCHA[2]

1 - Centre of Biological Engineering, University of Minho, 4710-057, Braga, Portugal

2 - ITQB NOVA, Av. da República, 2780-157 Oeiras, Portugal

3 - Department of Informatics, University of Minho, 4710-057, Braga, Portugal

The cork oak tree, Quercus suber, is an important renewable resource from which wine stoppers and many other natural products are derived. Portugal is the main producer of cork which approximately manifactures half of the world's total consumption. The recently sequenced genome of Quercus suber [1] has 953 Mb size containing about 79.000 genes. It is important to have a better under- standing of the genomics and metabolomics of the tree to increase resistance to abiotic and biotic stresses and to obtain high quality cork. This information can be used as a supporting parameter for decision-making in cork production since normally it is not possible to evaluate the quality of cork before the tree is 40 years old.

In this work we present a metabolic genome scale model for the leaf of the Quercus suber. We have used the Merlin software [2] to reconstruct the draft metabolic model. The enzymes are annoted using UniProtKB and SwissProt databases [3]. For the annotation we have used previously annotated plant en- zymes from species such as Quercus, Arabidopsis thaliana, Oryza sativa subsp. japonica, Vitis vinifera, Zea mays and Solanum tuberosum using the Blast e- value threshold e−10. If there was no suitable plant species, we annotated the enzyme with any other organism with e-value < e−10. As it is a well studied model organism, the majority of the enzymes have been annoted with Arabidopsis thaliana. More than 1600 annotations were manually curated with the help of Merlin's environment. The major pathways including Glycolysis/ Gluconeogenesis,

carbon fixation, TCA cycle, and production of amino-acids and other biomass precursors have been checked manually. The biochemical reactions were checked to be chemically and electrically balanced. Mostly, H+ and H2O may be missing in some databases therefore the reactions were updated according to MetaNetX, KEGG, BiGG databases when related information is found. Also the gaps in the pathways have been investigated and removed by adding/correcting necessary equations.

The biomass composition has been defined using protein, carbohydrates, lipids, cofactors, DNA and RNA components using similar approaches in previous plant models such as AraGEM and Tomato [4, 5]. The required drains for photons and inorganic compounds for CO2,H2O,O2 are defined in the model to describe the uptake/secretion of these compounds. The sources for Nitrogen, Phosphate, Sulphur are also defined similarly with the help of Ammonia, Nitrate, Orthophosphate, Sulphate and Hydrogen sulfide.

The obtained Cork model consists of 3269 reactions, 2934 metabolites and 7531 genes of which 405 are transporters. The transport genes have been identi- fied using the Triage tool of Merlin. The compartment prediction has been using Loc3Tree [6] protein localization prediction system. Chloroplast, cytoplasm, endoplasmic reticulum, golgi apparatus, mitochondrion, nucleus, peroxisome, plasma membrane, plastid, and vacuole are the predicted locations inside the plant cell. For the reversibility of the directions, we forced the Kegg's reaction directions according to the the yeast model reactions. Later we have manually curated the directions to guarantee the growth of biomass precursors.

The validity of the model has been checked for biomass production, using simulation tools such as Optflux [9]. Also, inhouse model validity tools have been used for the control of biomass precursors. The derived Cork model is able to grow biomass and produce Oxygen under photosynthetic conditions with the help of photons where CO2 is the main carbon source. On the other hand when the main carbon source is defined as glucose, the model can simulate the respiration, producing CO2 and H2O and consuming Oxygen. We observe non-zero flux at about 1445 reactions which is also a comparable result with previous studies. We conclude that our leaf model is capable of simulating both photosynthetic (light) and respiration (dark) reactions. To the best of our knowledge, this is the first metabolic model of a tree species. We believe that this model will bring new insights for Quercus suber studies such as the formation process of cork. Also, it will be a basis for metabolic models of other plant species. Considering plants as resources of many valuable natural metabolites, these models can bring significant biotechnological applications.

References

[1] The draft genome sequence of cork oak, Ramos Antonio Marcos et al., Scientific Data, 5, 2018.

[2] Reconstructing genome-scale metabolic models with merlin,Oscar Dias, Miguel Rocha, Eug´enio C. Ferreira, Isabel Rocha; Nucleic Acids Research, Volume 43, Issue 8, pp. 3899–3910, 2015.

[3] The UniProt Consortium, UniProt: the universal protein knowledgebase , Nucleic Acids Res. 46: 2699 (2018). [4] AraGEM, a Genome-Scale Reconstruction of the Primary Metabolic Network in Arabidopsis, Cristiana Gomes de Oliveira Dal'Molin, et al., Plant Physiology, 152 (2):579-589, 2010.

[5] A genome-scale metabolic network reconstruction of tomato (Solanum ly- copersicum L.) and its application to photorespiratory metabolism, Yuan H. et al., Plant J., 85(2):289-304, 2016.

[6] Loc3Tree: Protein Subcellular Location Prediction Sytem, Goldberg T. et al. Nucleic Acid Research, 2014.

[7] Responses to lightintensity in a genome-scale model of rice metabolism, Poolman, M.G. et al., Plant Physiol. 162, 1060–1072, 2013.

[8] Assessing the metabolic impact of nitrogen availability using a compart- mentalized maize leaf genome-scale model, Simons M. et al., Plant Phys- iol., 166(3):1659-74, 2014.

[9] OptFlux: an open-source software platform for in silico metabolic engi- neering, Isabel Rocha et al., BMC Systems Biology 4:45, 2010.