Towards a genome-scale model of the Cork Oak tree

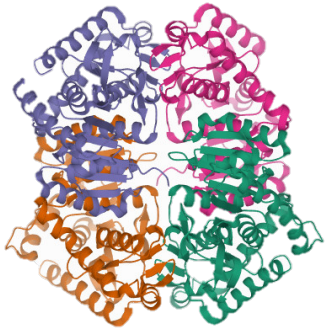
Oscar Dias
17/11/2020

GSM Models

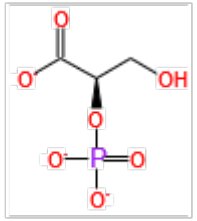
DNA → RNA → ENZYME → REACTION

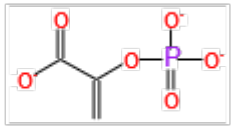
Transcription

Translation

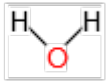


Catalysis

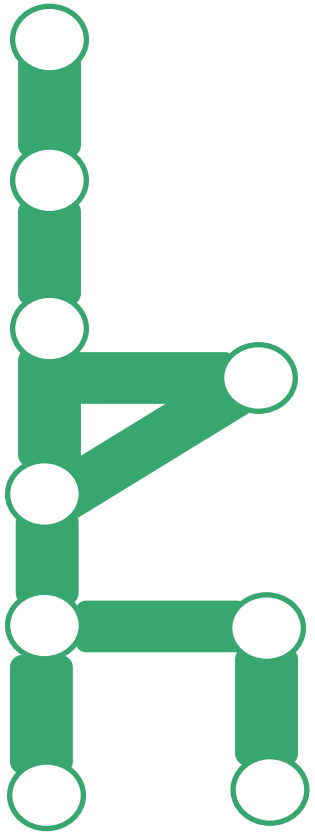




+



GPRs
Compartments
Gap Filling
Transport reactions



Genome annotation

Metabolic Network Assembly

GSM Models

Stoichiometric Matrix

$$\begin{pmatrix} S_{11} & \dots & S_{1n} \\ \dots & \dots & \dots \\ S_{m1} & \dots & S_{mn} \end{pmatrix} \times \begin{pmatrix} v_1 \\ \dots \\ v_j \end{pmatrix} = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}$$

Flux Vector

Further Constraints

$$\begin{pmatrix} 0 \\ \dots \\ -10 \end{pmatrix} \leq \begin{pmatrix} r_1 \\ \dots \\ v_j \end{pmatrix} \leq \begin{pmatrix} 999 \\ \dots \\ 999 \end{pmatrix}$$

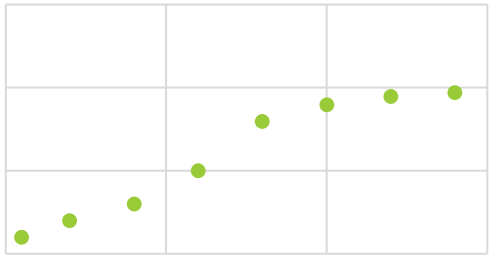
Biomass Formulation

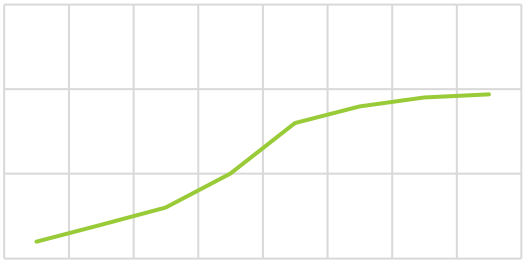
DNA + RNA + Protein + Lipid + ... ⟶ Biomass

FBA

$$maximize \rightarrow z$$
$$subject\ to: S \times v = 0$$
$$\alpha_j \leq v_j \leq \beta_j$$

Experimental Data x Phenotype Prediction





Conversion into a Stoichiometric model

Model validation and Phenotype Prediction

Quercus suber - Cork Oak

- Forest tree from the Mediterranean region
- Cork production
- Huge economic impact
- Biotic and abiotic stress:
  - Climate changes (drought, heat waves)
  - Diseases
  - Fires

http://corkoakdb.org/organism/1

Goals

1. Reconstruction of a high-quality GSM model for *Q. suber*
2. Integrate transcriptomics data to obtain tissue-specific models
3. Merge the tissue-specific models into a diel multi-tissue model
4. Evaluate the metabolic behavior of *Q. suber in stress conditions*

Methodology

Genome sequence

- Assembly accession GCF_002906115.1
  - 20-DEC-2017
  - CEBAL

BLAST

- SWISS-PROT
- TrEMBL
- e-Value threshold: e^{-30}

Genome annotation

- **Reviewed:**
  - A: *Quercus suber*
  - B: *Quercus*
  - *C: Arabidopsis thaliana*
  - D: *Vitis vinifera*
  - E: *Nicotiana tabacum*
  - F: *Solanum tuberosum*
  - G: *Oryza sativa subsp. Japonica*
  - H: *Zea mays*
  - I: Other organism
- **Unreviewed:**
  - A: *Arabidopsis thaliana*
  - B: *Oryza sativa subsp. japonica*
  - C: *Zea mays*
  - D: default

Methodology

- **Leaf and Inner Bark:**
  - PRJNA392919 (CEBAL)
- **Phellogen:**
  - PRJEB33874 (FCUL)

Workflow

RNA-SEQ dataset retrieval	FASTQ quality check	FASTQ quality trimming	Aligning trimmed reads against reference genome	Counting mapped reads	Filtering and normalizing read counts

Tools

Quercus suber leaf tissue RNA-SEQ transcriptomics dataset	*FastQC*	*Sickle* windowed adaptive trimming	*Bowtie2*	*FeatureCounts* program of the *Subread* package	*edgeR* library

I/O

Input	2 *FASTQ* files containing paired-end reads	2 *FASTQ* files + trimming parameters	2 trimmed *FASTQ* files + reference genome in *FAA* or *FNA* format	*SAM* file + genome annotation file in *GTF* or *SAF* format	Raw read counts in *TXT* format
2 *FASTQ* files containing paired-end reads	Extensive quality report + estimation of trimming parameters	2 trimmed high-quality *FASTQ* files	*SAM* file containing the reads aligned to the reference sequences	Read counts in *TXT* format	Preprocessed read counts in *CSV* format
Output					

Methodology

Transcriptomics Data

Q. suber GSMM

TROPPO

Specific GSMMs

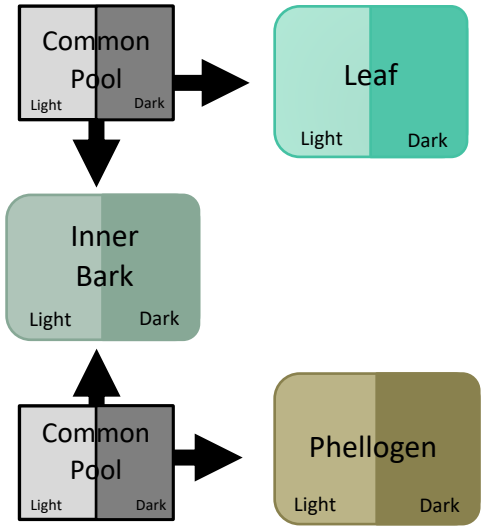
Diel multi-tissue GSMM

Merlin
metabolic models reconstruction using genome-scale information

Leaf

Inner bark

Phellogen

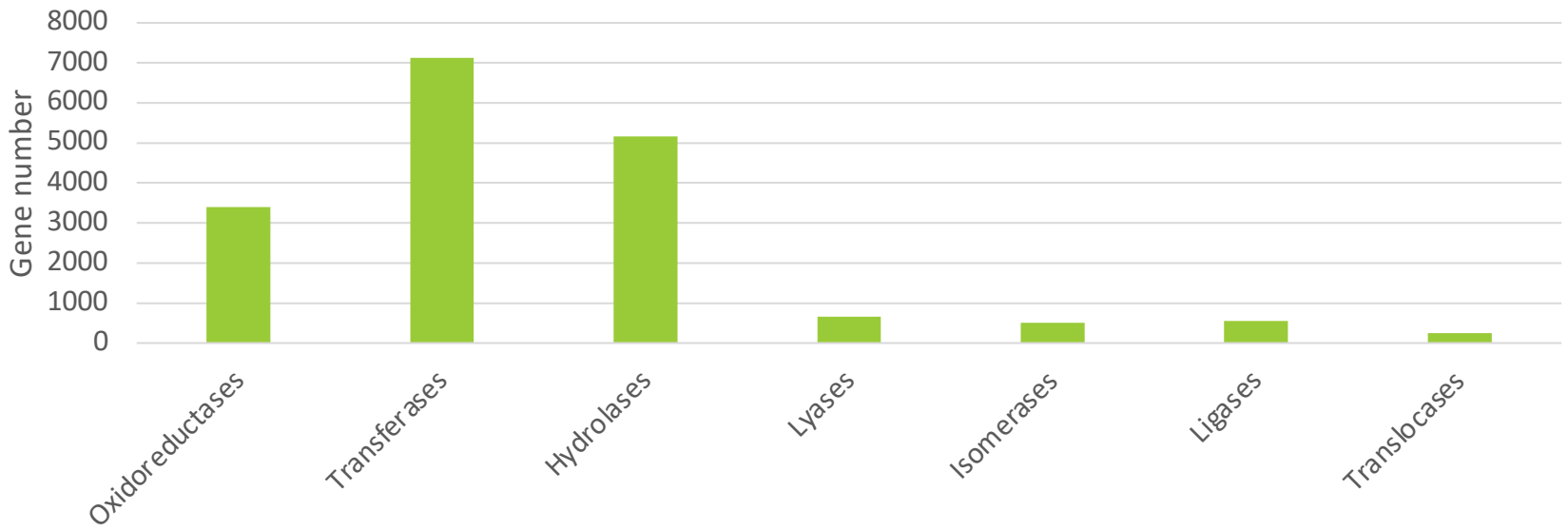


Diel multi-tissue GSMM

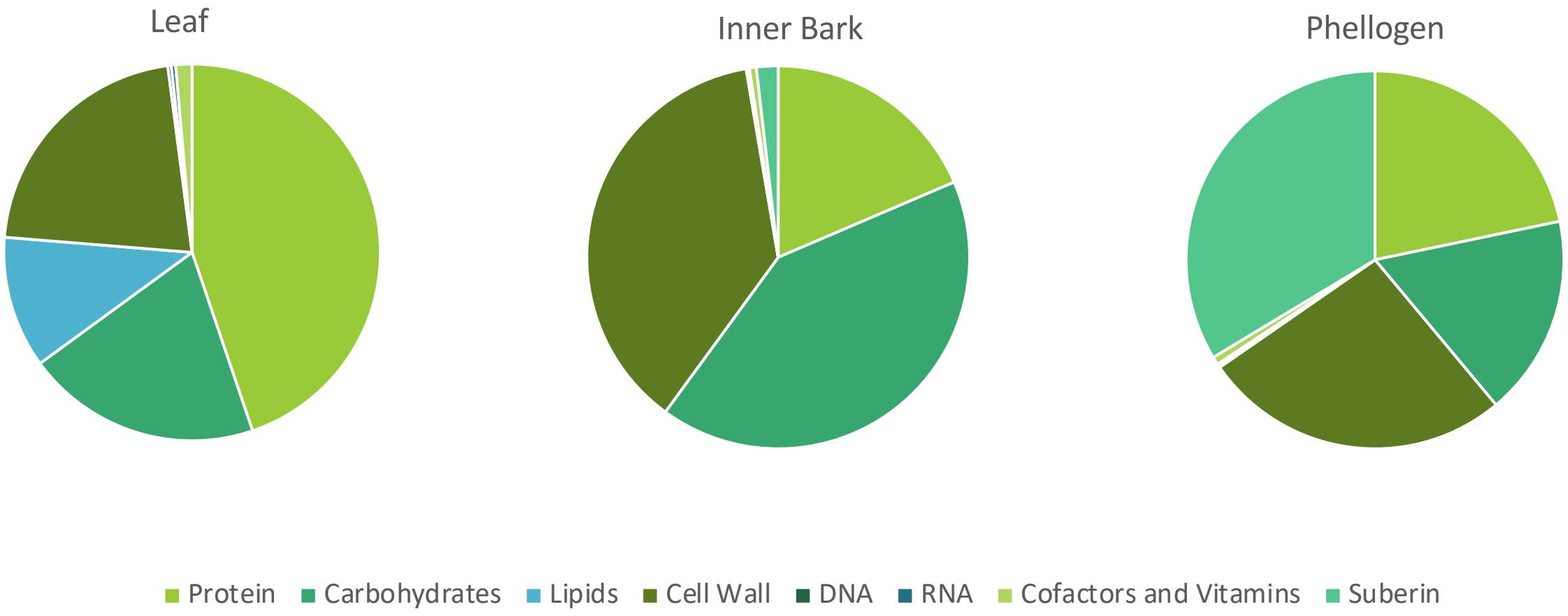
Genome annotation

59614 genes

17654 metabolic genes

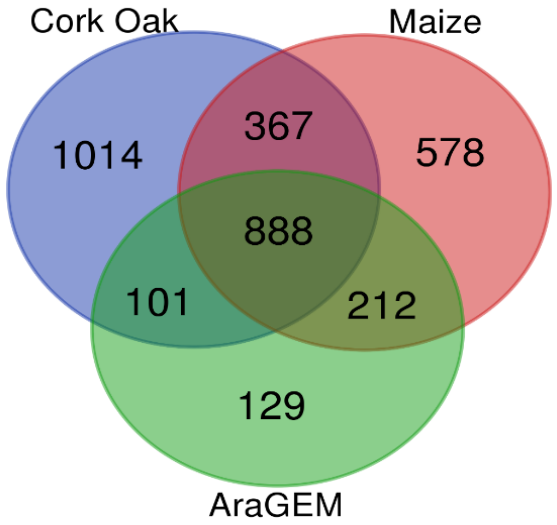


Biomass composition



Model Comparison

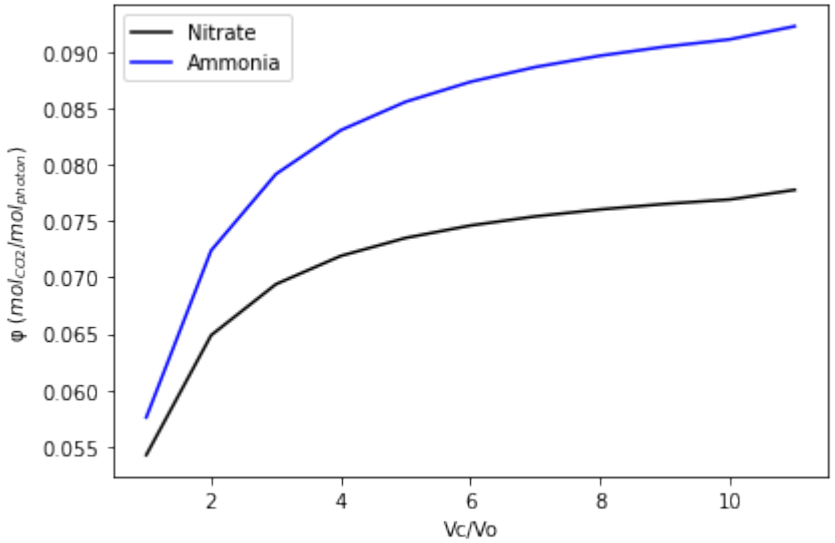
	Cork Oak (This work)	AraGEM (2013)	Maize (2015)	Tomato (2015)	Rice (2017)
Genes	7752	2857	-	3410	3602
Metabolites	8059	2769	6458	2143	1136
Reactions	7561	2857	6250	1998	1330
Compartments	8	5	9	5	4



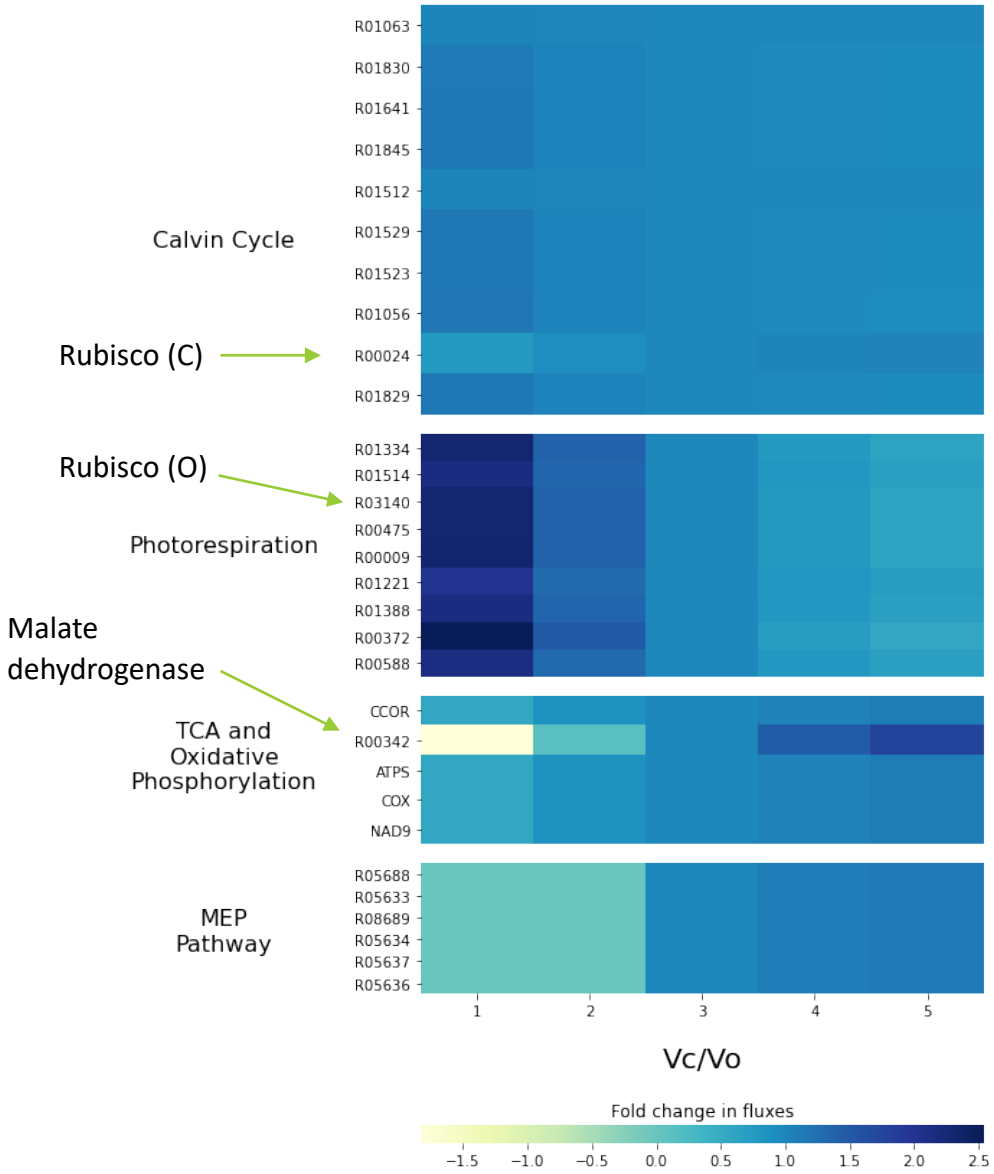
Leaf-GSMM

Assimilation Quotient (mol_{CO2}/mol_{O2})

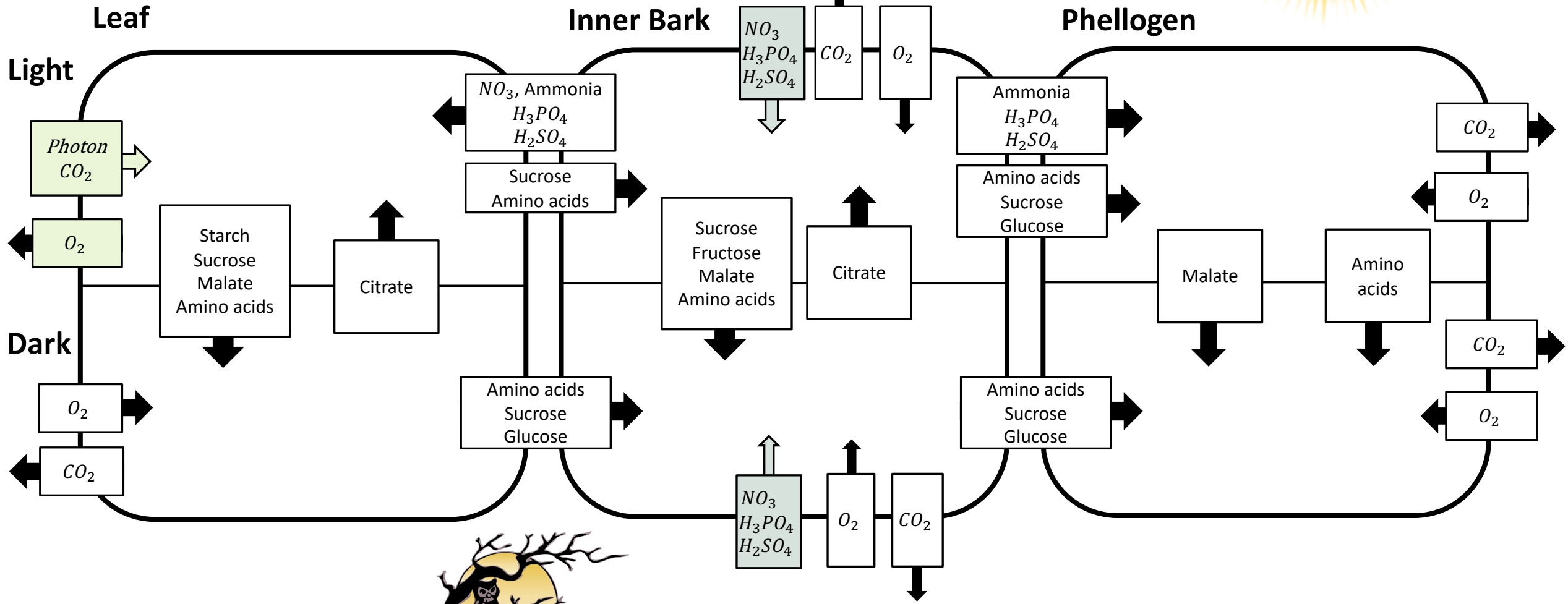
Nitrogen Source	Cork Oak GSMM	*Oryza sativa* GSMM
Nitrate	0.75	-
Ammonia	0.93	0.87-0.98



Quantum yield Φ (mol_{CO2}/mol_{photon})	
Autumn	Summer
0.084-0.089	0.051-0.075



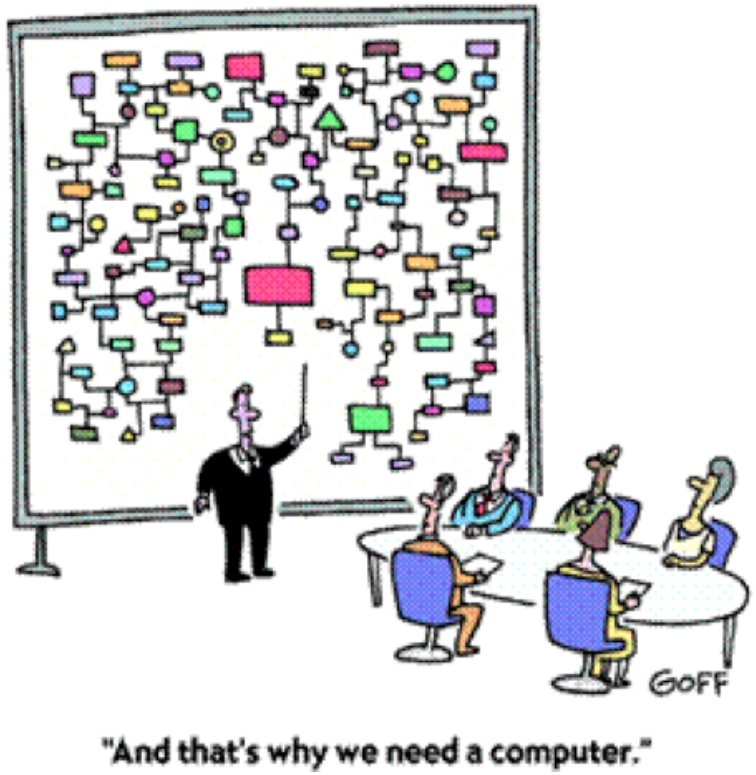
Diel Multi-tissue GSMM



Conclusions and Future Perspectives

- Validated GSM model containing plant secondary metabolism pathways
- Tissue specific models based on transcriptomics data
- Diel Multi-tissue GSM model
- Possible applications:
  - Extensive analysis of utilization of different nitrogen and phosphorus sources
  - Application of dynamic simulations (dFBA) to evaluate the growth and metabolism over time
  - Study of interactions with eventual pathogenic organisms

Thank you!



"And that's why we need a computer."

Emanuel Cunha
Miguel Silva
Diogo Lima
Hüseyin Demirci
Miguel Rocha
Isabel Rocha
Oscar Dias

https://www.ceb.uminho.pt/biosystems

CENTRE OF BIOLOGICAL ENGINEERING

BioData.pt

elixir PORTUGAL