



HAL
open science

TRANS-C³ - The Transcriptome of downy oak (*Quercus pubescens* Willd.) in Response to Climate Change

Xi Liu, Béatrice Lloriod, Stefano Caffarri, Erwan Corre, Jean-Philippe Mévy

► To cite this version:

Xi Liu, Béatrice Lloriod, Stefano Caffarri, Erwan Corre, Jean-Philippe Mévy. TRANS-C³ - The Transcriptome of downy oak (*Quercus pubescens* Willd.) in Response to Climate Change. Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM), Jul 2018, Marseille, France. hal-01857993

HAL Id: hal-01857993

<https://hal.science/hal-01857993v1>

Submitted on 17 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TRANS-C³ - The Transcriptome of downy oak (*Quercus pubescens* Willd.) in Response to Climate Change



Xi LIU¹, Béatrice LORIOD², Stefano CAFFARRI³, Erwan CORRE¹, Jean-Philippe MEVY⁴

¹ CNRS, Sorbonne Université, FR2424, ABIMS platform, Station Biologique, 29680, Roscoff, France - xi.liu@sb-roscoff.fr, erwan.corre@sb-roscoff.fr

² Transcriptomique et Génomique Marseille-Luminy (TGML), Campus Universitaire de Luminy 13009 Marseille, France U1090 Inserm, Aix-Marseille Université - beatrice.loriod@inserm.fr

³ CEA-CNRS, UMR 6191 Biologie cellulaire et moléculaire et des plantes et bactéries, Institut de Biologie Environnementale et Biotechnologie (IBEB), Laboratoire de génétique et biophysique des plantes (LGBP), Faculté des Sciences de Luminy, Case 901, 163 av de Luminy13288 Marseille cedex 9 - stefano.caffari@univ-amu.fr

⁴ IMBE - UMR CNRS 7263 / IRD 237 Equipe Diversité et Fonctionnement : des Molécules aux Ecosystèmes (DFME), Aix-Marseille Université, Centre Saint-Charles - Case 4 - Jean-Philippe.mevy@imbe.fr

1. Background

The TRANS-C3 project fits into the context of global change and aimed to understand the response of forest trees to the aridification of the Mediterranean climate. For this purpose, researches have been conducted using downy oak model at the O3HP (Oak Observatory at the OHP) site, where the drought prediction is simulated by a reduction in rainfall amount to about 30%. Precisely the objective was to understand *in situ* plant response to climate change both in terms of gene expression, metabolic footprint and the impact on an essential physiological process: photosynthesis.

Keywords: RNA-seq, Genome-guided *de novo* transcriptome assembly, Gene expression, Oak, Global change, Cross-cutting approach, Photosynthesis, Metabolomics

2. Samples

Tissue
Leaves from *Quercus pubescens* Willd.

40 Samples and 4 conditions
2 seasons (summer – spring) x 2 treatments (water stress – control) x 10 trees

Sequencing
Illumina Next-Seq 500: 150 bp paired-end x 40 libraries (≈50M reads/sample)

3. Assembly

Preprocessing

Trimmomatic^[1]: adapter removing, quality trimming and filtering
% of paired reads surviving vary from 89% to 93%, with no notable difference between the samples: 1.7Md pair reads pass the cleaning

In silico reads normalization

Trinity^[2]: normalization by set then global normalization
• %selection by read set vary from 16% to 26% (360M pair reads in total), e.g. reads selected after normalization in summer samples > spring samples ⇒ summer samples display higher transcriptomic diversity
• 86M paired reads after final step of global normalization

De novo assemblies

Assembly	Nb Seq	Nb Residue	Average	N50
Trinity	124070	53333443	429.87	465
CLC <i>de novo</i> ^[3]	458607	203955898	444.73	452
BinPacker ^[4]	396590	478221122	1205.83	2102
Trinity genome-guided	530080	315371132	594.95	780

Table 1 | comparison of different assembly metrics. Reads were assembled in FR orientation, K-mer was set or estimated at 25. BinPacker was used by activating option -d. HiSta2^[5] and Samtools^[6] were used in upstream for Trinity genome-guided *de novo* assembly. 155986 sequences in Trinity genome-guided assembly possesses the TPM value superior to 1 cross all the samples.

- Trinity *de novo* should have generated more than 124070 sequences
- BinPacker generated longer sequences and assembled more residues than the other assembler tested
- Trinity genome-guided using genome from close species *Q. robur* assembled an higher number of transcripts than other assemblers

Pseudo-mapping reads back to the assemblies

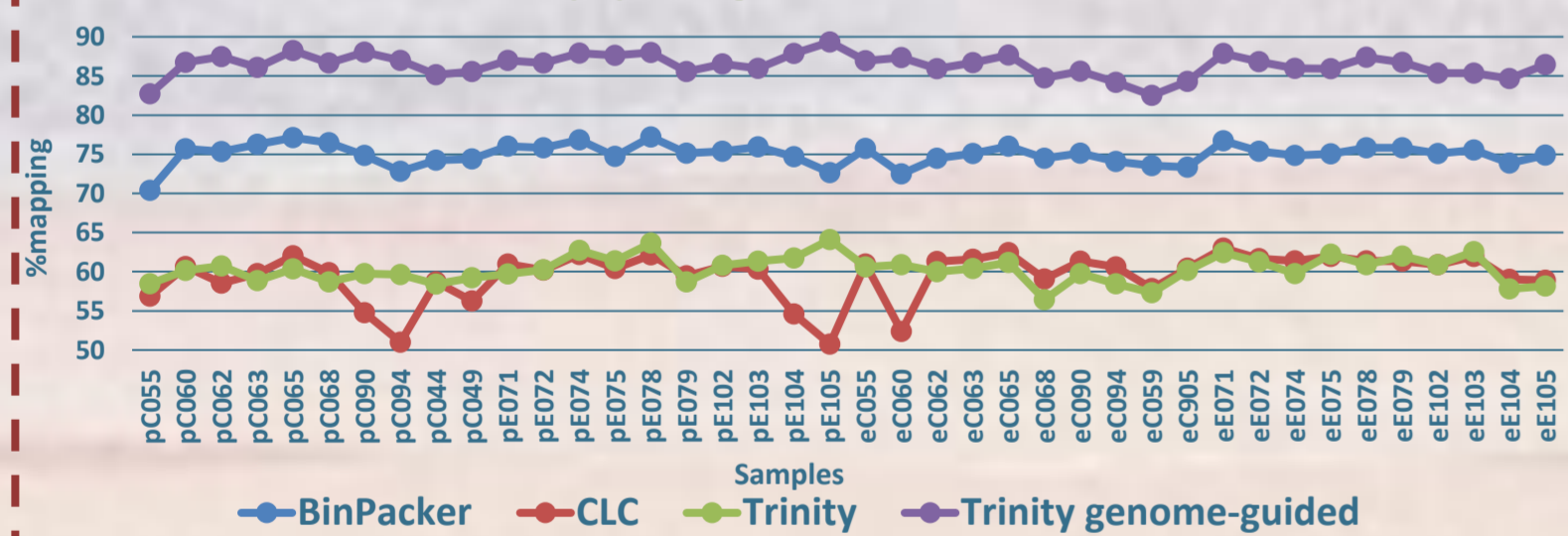


Figure 1 | Comparison of pseudoalignment rate between different assemblies. Kallisto^[7] analysis with FR strand-specific paired-end pseudoalignment option. We also mapped reads back to a public reference transcriptome, but the pseudo-mapping rate was just around 40%.

Trinity genome-guided *de novo* assembly displays the best pseudoalignment rate compare to the other *de novo* assemblers for the diploidic species assembly.

Trinity genome-guided *de novo* assembly evaluation

Contig Ex90N50 Statistic: Ex90N50 is 1544bp and corresponds to 34572 transcripts
BUSCO^[8]: C:87.9%,S:41.7%,D:46.2%,F:7.8%,M:4.3%,n:1440
BUSCO analysis result displays a good completeness (1440 BUSCO groups searched) and completeness (87.9% complete BUSCOs orthogroups) of transcriptome, but displays a high number of duplicated groups (46.2%).

⇒ Trinity genome-guided *de novo* assembly was chosen as the reference transcriptome for further analyses.

4. Differential analysis

Samples	eC	eE	pC	pE
eC	/	15	3122	3529
eE	31	/	3460	3709
pC	2312	2613	/	2
pE	2569	1949	11	/

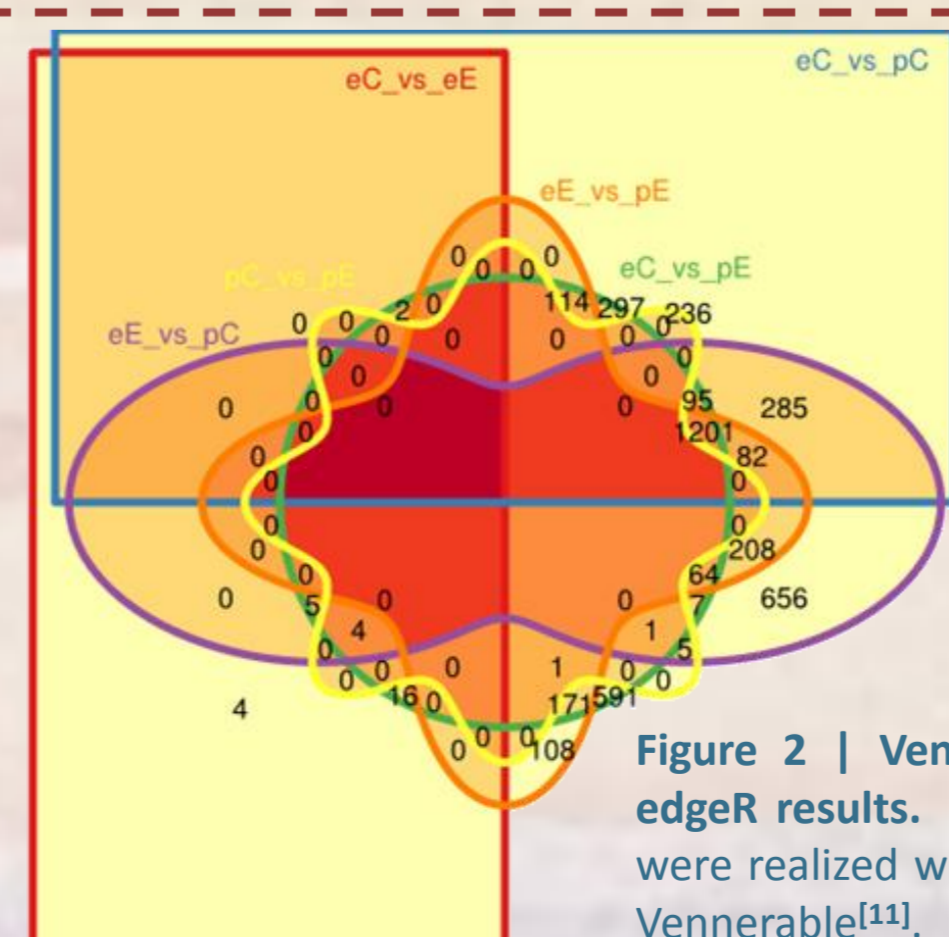


Figure 2 | Venn diagram of edgeR results. Venn diagram were realized with R packages Vennrable^[11].

Table 2 | Differentially expressed genes (DEG) in response to the aridification. The thresholds of statistical tests were fixed at $\text{Padj} < 1e^{-3}$ and $\log_2\text{FC} \geq 2$. 5724 and 4153 DEG were detected by DESeq2^[9] and edgeR^[10] respectively (detail shown in black and grey). Abbreviations of the conditions: e and p for summer and spring, C and E for control and water stress.

Genetic expression in response to the aridification varies with the season, according to edgeR results, 4 transcripts from 42 were expressed independently to the season.

5. Water stress vs control in summer: annotation with GO-enrichment

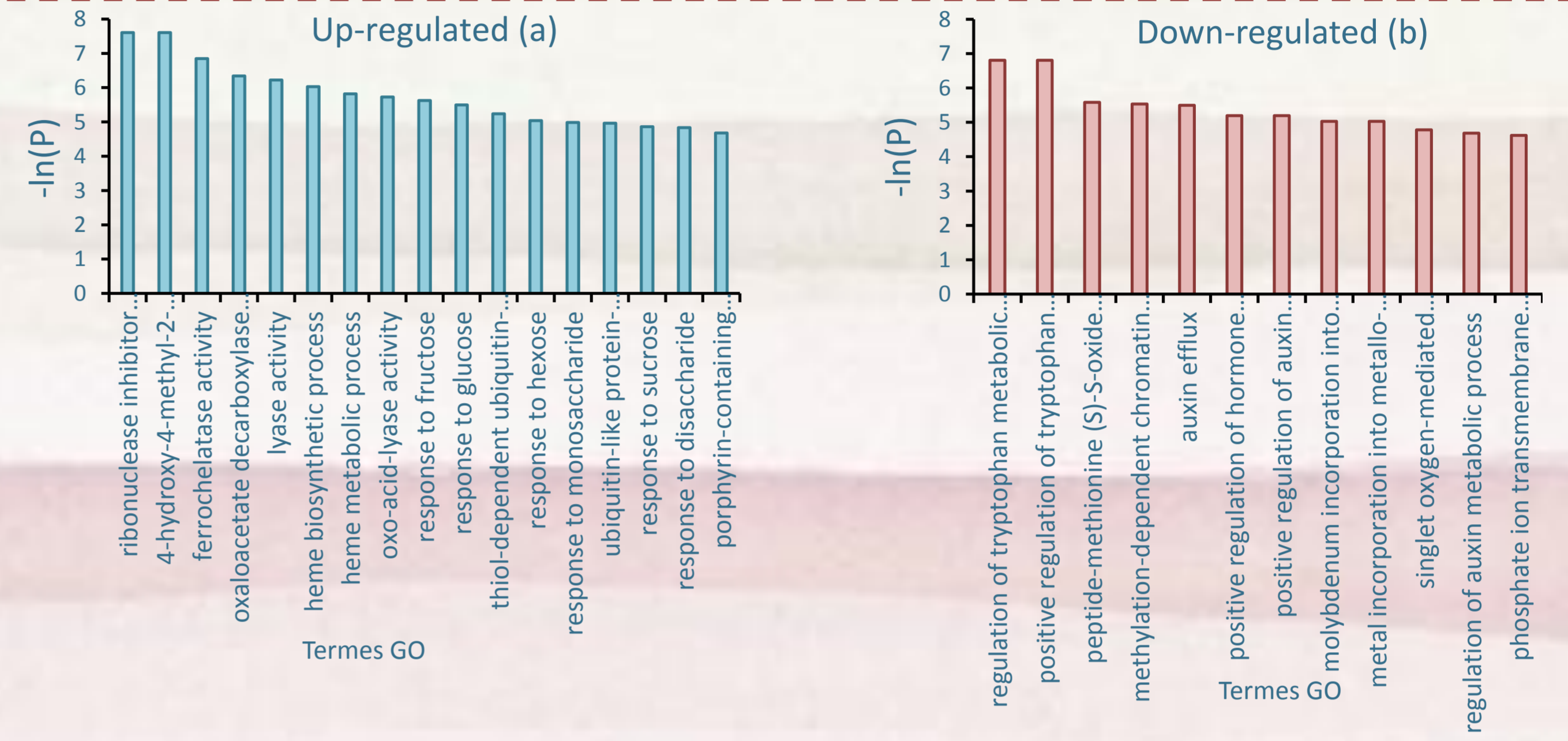


Figure 3 | Enrichment analysis based on GO functional categories for transcripts that were significantly at p-value < 0.01 (a) up-regulated and (b) down-regulated in response to *Q. pubescens* drought treatment during summer.

6. Plasticity of photosynthetic machinery: light reactions

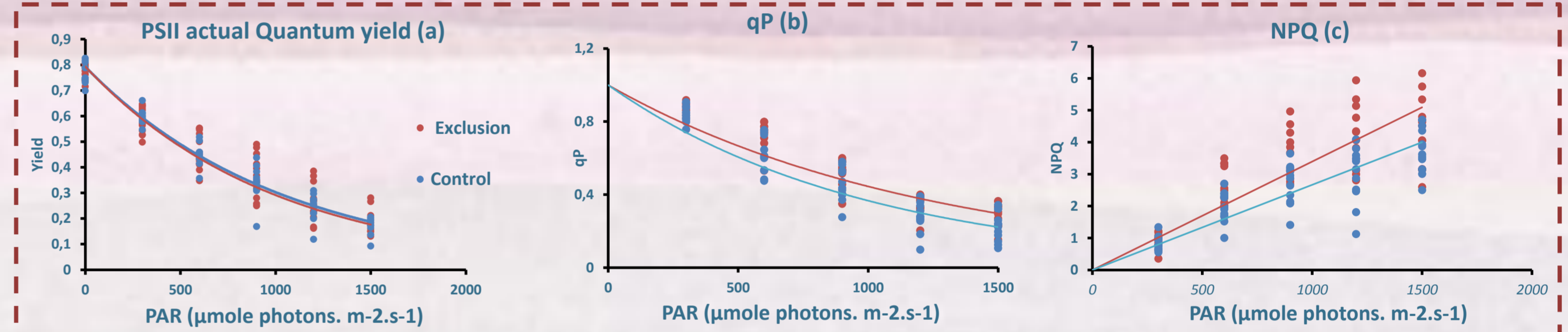


Figure 4 | Chlorophyll fluorescence parameter measurements during summer: Kinetics of Photosystem II quantum yield (a), photochemical quenching (b) and Non-photochemical quenching coefficients (c).

During drought, light energy absorbed in excess is mainly dissipated as heat, as shown by the increase of the Non-photochemical quenching (NPQ) coefficient.

7. What relationship with pigments?

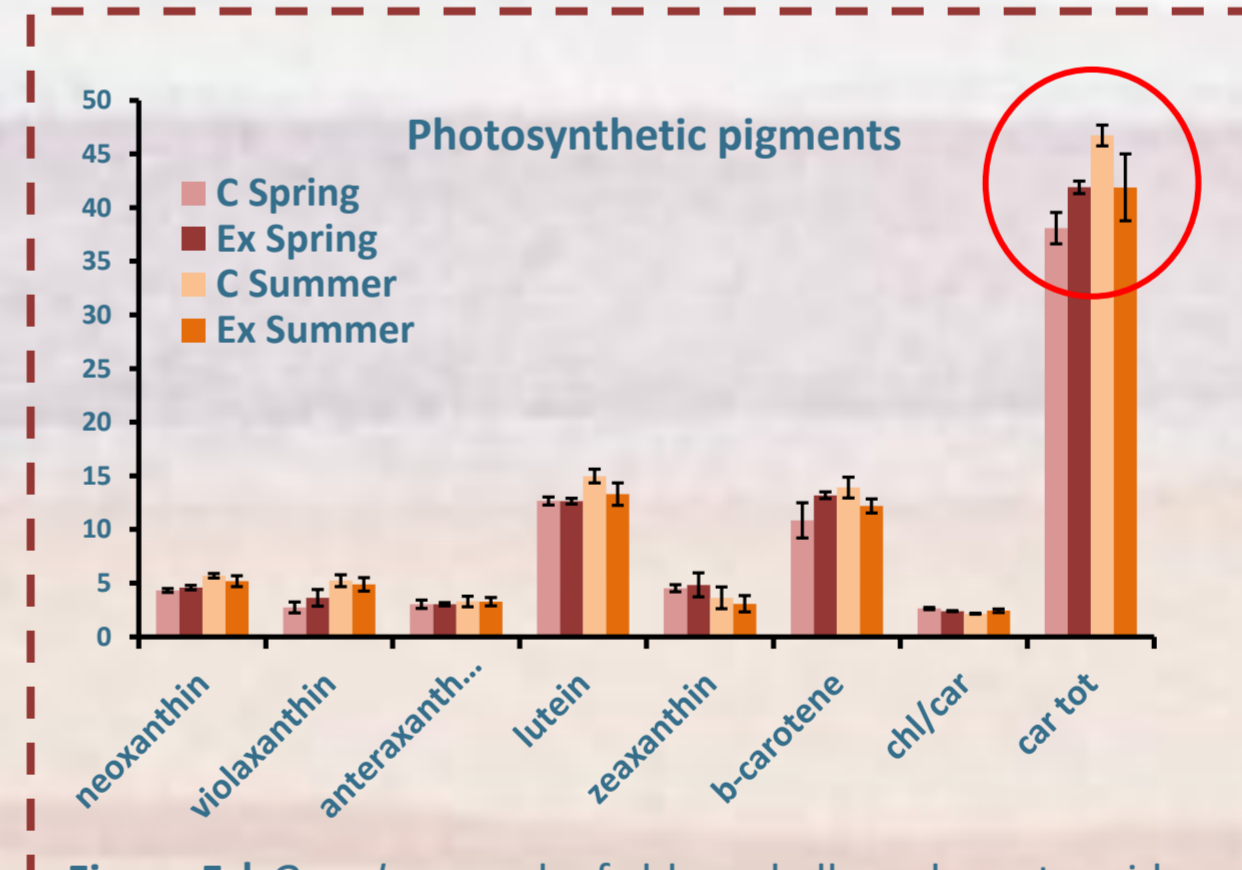


Figure 5 | *Q. pubescens* leaf chlorophylls and carotenoids relative contents by HPLC analysis. Only total carotenoids change with rain exclusion and seasons

8. The proteins of photosystems I and II?

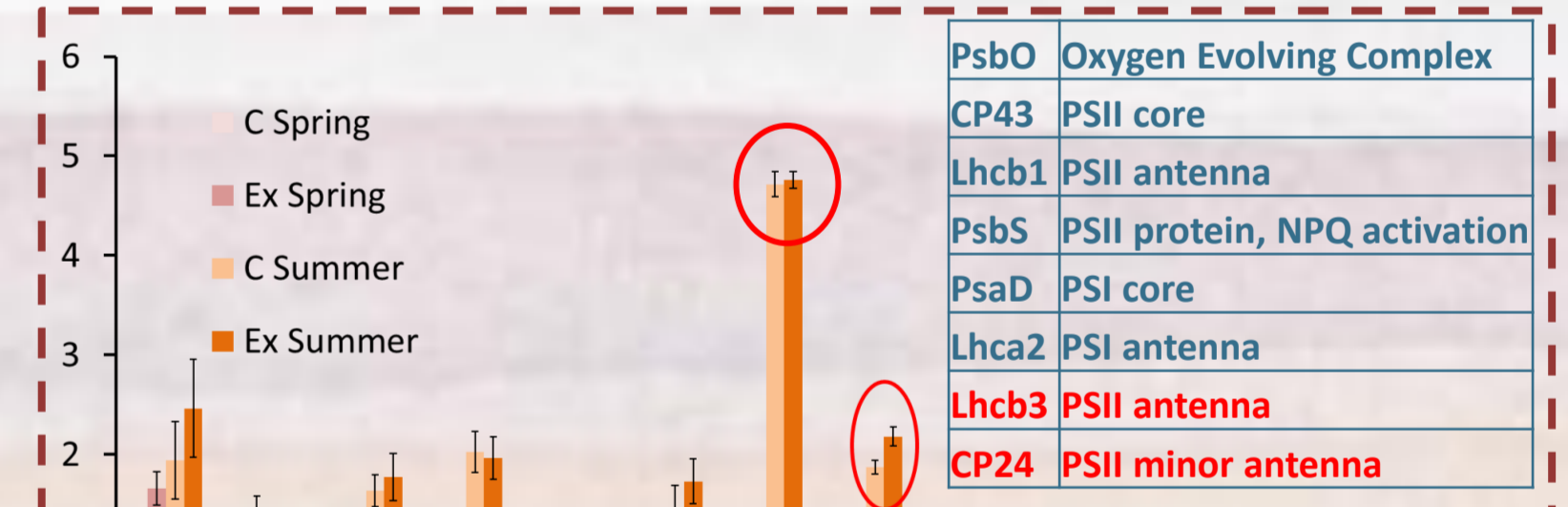


Figure 6 | Relative abundance of photosystems II and I proteins after electrophoresis separation
• CP24 increases with the exclusion rain during the summer
• Very clear effect of an increased synthesis of Lhcb3 in summer and spring

9. Targeted Metabolomics: Differential Analysis

Compound name	Fold change			
	Summer	Spring	Spring/Summer	
Lactic acid	1.0	1.6	0.7	0.4
Oxalic acid	1.8	1.2	0.4	0.6
Succinic acid	1.0	1.2	0.4	0.4
Pyroglutamic acid	1.8	2.1	1.3	1.1
Xylulose	0.9	1.8	9.8	4.9
Sorbitol	1.4	1.1	2.1	2.5
Sucrose	1.1	1.1	0.6	0.6
Quinic acid	0.8	0.6	0.6	0.8
Malic acid	1.6	1.3	1.2	1.5
Gallic acid	1.0	1.6	0.6	0.4
Catechin	0.5	0.5	0.3	0.3

Table 3 | Metabolite markers in response to drought throughout the seasons: Exclusion (E) versus Control (C) and spring versus summer.

Pyroglutamic acid identified as a metabolic marker of drought. Sugars as Xylulose and Sorbitol are seasonal metabolite markers

10. Enrichment of the metabolic pathway

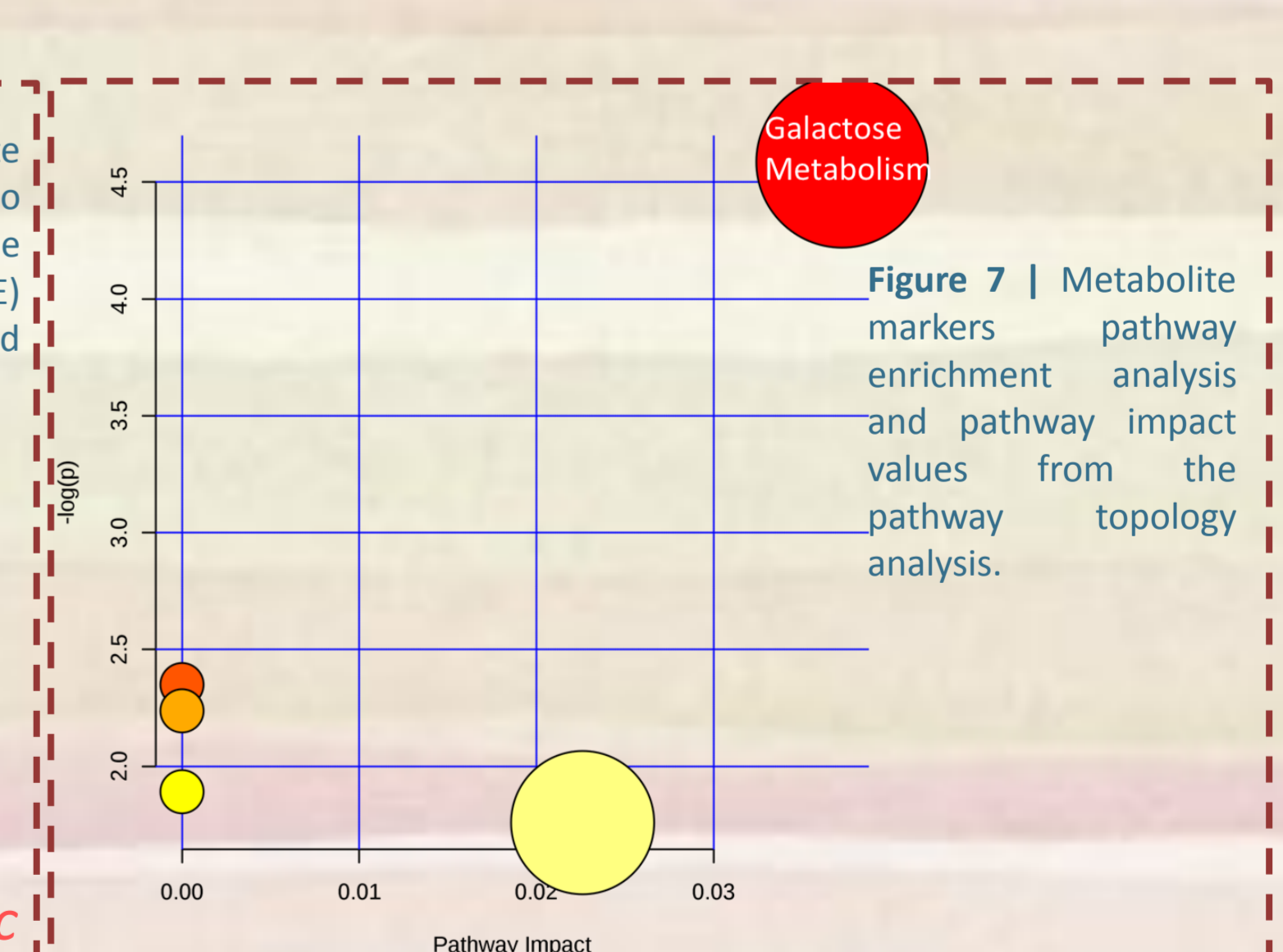


Figure 7 | Metabolite pathway enrichment analysis and pathway impact values from the pathway topology analysis. The water deficit significantly activates the galactose metabolism pathway

11. Conclusion and Perspective

This project has identified genetic, biochemical, chemical and physiological markers of plant adaptation to climate change in a cross-cutting approach. It has been carried out around several disciplinary fields: Genomics, Biochemistry, Ecophysiology, Chemistry, Bioinformatics and Modeling, opening perspectives for more ambitious projects.

[1] Bolger, A. M. et al. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170 (2014).

[2] Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* advance online publication, (2011).

[3] <https://www.qiagenbioinformatics.com/products/clc-assembly-cell/>

[4] Liu J, Li G, Chang Z, et al. BinPacker: Packing-Based De Novo Transcriptome Assembly from RNA-Seq Data. *Lengauer T, ed. PLoS Computational Biology*. 2016;12(12):e1004772. doi:10.1371/journal.pcbi.1004772.

[5] Sirdén J, Välimäki N, Mäkinen V (2014) Indexing graphs for path queries with applications in genome research. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 11: 375–388. doi: 10.1109/tcbb.2013.2297101

[6] Li H*, Handsaker B*, Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/Map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078–9. PMID: 19505313

[7] Nicolas L, Bray, Harold P, Minet, P, Melsted and Lior Pachter. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34, 525–527 (2016). doi:10.1038/nbt.3519

[8] Felipe A. Simão, Robert M. Waterhouse, Panagiotis Ioannidis, Eugeny M. Zdobnov; BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, Volume 31, Issue 19, 1 October 2015, Pages 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>

[9] Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology*, 15, 550. doi: 10.1186/s13059-014-0550-8

[10] Robinson MD, McCarthy DJ, Smyth GK (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." *Bioinformatics*, 26(1), 139–140.

[11] Swinton J (2009). Venn diagrams in R with the Vennrable package

