



## Water stress-associated isolation barriers between two sympatric oak species

Grégoire Le Provost, Benjamin Brachi, Isabelle Lesur, Céline Lalanne, Karine Labadie, Jean-Marc Aury, Corinne da Silva, Dragos Postolache, Thibaut Leroy, Christophe Plomion

### ► To cite this version:

Grégoire Le Provost, Benjamin Brachi, Isabelle Lesur, Céline Lalanne, Karine Labadie, et al.. Water stress-associated isolation barriers between two sympatric oak species. 2021. hal-03351973

**HAL Id: hal-03351973**

**<https://hal.inrae.fr/hal-03351973>**

Preprint submitted on 22 Sep 2021

**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.

## Water stress-associated isolation barriers between two sympatric oak species

Grégoire Le Provost<sup>1,†</sup>, Benjamin Brachi<sup>1</sup>, Isabelle Lesur<sup>1,2</sup>, Céline Lalanne<sup>1</sup>, Karine Labadie<sup>3</sup>, Jean-Marc Aury<sup>3</sup>, Corinne Da Silva<sup>3</sup>, Dragos Postolache<sup>4</sup>, Thibault Leroy<sup>1,5</sup>, Christophe Plomion<sup>1</sup>.

1: INRAE, Univ. Bordeaux, BIOGECO, F-33610 Cestas, France.

2: Helix Venture, F-33700 Mérignac, France

3: Commissariat à l'Energie Atomique (CEA), Institut de Génomique (IG), Génoscope, F- 91057 Evry, France.

4: Forest Research and Management Institute, Cluj Napoca, str.Horea 65, Cluj, 400202, Romania.

5: IRHS-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071 Beaucouzé, France

† Corresponding author: Grégoire Le Provost

BIOGECO, INRA, Univ. Bordeaux, F-33610 Cestas, France.

Tel: +33(0)535385334,

Fax: +33(0)557122881

E-mail: [gregoire.le-provost@inra.fr](mailto:gregoire.le-provost@inra.fr)

Running Head: reproductive barrier loci in oaks

# Abstract

Drought and waterlogging impede tree growth and may even lead to tree death. With climate change, these environmental factors are a growing source of concern, particularly for temperate forests. Oaks, an emblematic group of tree species, have evolved a range of adaptations to cope with these constraints. The two most widely distributed European species — pedunculate oak (PO) and sessile oaks (SO) — have overlapping ranges, but are highly constrained locally by soil water content variation. These differences in local ecological requirements provide a powerful biological model for studying the role of ecological barriers in speciation. We used an experimental set-up mimicking the ecological preferences of these species, in which seedlings were subjected to waterlogging and drought. We studied gene expression in roots by RNA-seq and identified genes differentially expressed between treatments with different outcomes depending on species. These “species x environment”-responsive genes revealed adaptive molecular strategies involving adventitious and lateral root formation, aerenchyma formation in PO, and osmoregulation and ABA regulation in SO. With this experimental design, we also identified genes with expression profiles presenting a “species” effect regardless of imposed constraints with important roles in intrinsic reproductive barriers. Finally, we compared our findings with those for a genome scan of species divergence and found that the candidate genes were enriched in highly differentiated SNPs. This suggests that many of the genes involved in the contrasting transcriptomic responses are subject to natural selection and that gene regulation helps to maintain these two different oak species in sympatry.

## Introduction

One of the major goals in evolutionary biology is understanding the genetic architecture of speciation, the number of genes involved in the evolution of reproductive barriers, their effect sizes and functions (Campbell et al, 2018)). There is now a clear consensus among evolutionary biologists about the continuity of the speciation process, including the gradual accumulation of barriers to gene flow in geographically isolated populations (Coyne & Orr, 2004). However, the long-standing debate about the contribution of local adaptation to the initiation and development of reproductive isolation between non-geographically isolated populations remains unresolved. This question is particularly difficult to answer given that most species are known to have experienced successive periods of isolation and contact in the past (Bierne et al., 2011). It remains possible to identify loci that have contributed to reproductive isolation (referred to hereafter as barriers to gene flow) in pairs of closely related species that are still exchanging genes, and to formulate functional hypotheses about the ecological or non-ecological (intrinsic) nature of reproductive barriers, by considering gene annotations. It is, indeed, possible to identify ecological and non-ecological barriers to gene flow by scanning genomes for regions with high levels of relative divergence. Many scientific articles published over the last decade have identified candidate reproductive barrier genes through analyses of whole-genome sequence data from closely related species (reviewed by Ravinet et al., 2017; Wolf et al., 2010).

As a complementary approach to classical studies aiming to analyze the level of sequence divergence between sister species (reviewed by Storz, 2005), studies of the relationships between divergences in gene expression and species persistence have also yielded interesting results. Fay and Wittkopp (2008) suggested that variations in transcription levels have a tremendous potential

for generating evolutionary novelty that can then be harnessed by natural selection. This approach is based on the detection of genes displaying significant differential expression (significantly differentially expressed genes or DEGs) between two or more focal species and searching for (i) genes displaying “species-specific” differential expression, regardless of the environmental conditions, or (ii) “species-specific” DEGs expressed in particular environments. The first group would be expected to contain intrinsic reproductive isolation barriers between sister species, whereas the second, containing genes displaying significant species-by-environment interactions, would be expected to include important genes underlying the strategies of the species concerned for coping with the challenging environmental conditions encountered in its ecological niche. The contribution of these ecological barriers to overall reproductive isolation (sometimes referred to as ecological speciation) remains a matter of debate. However, these specific contributions can be deciphered by identifying genes in an experimental set-up mimicking the different environmental requirements contributing to the ecological preferences of adapted species and, consequently, to overall reproductive isolation.

Most of the transcriptomic research studies on speciation performed to date involved the use of animal models, including *Littorina saxatilis* marine snails (Martínez-Fernández et al., 2010), the lake whitefish (*Coregonus clupeaformis*; Renaut et al. 2010, Jeukens et al., 2010) and Darwin's finches (Abzhanov et al., 2006). In plants, the role of diverging gene expression in speciation has been studied mostly in hybrid systems. For example, in *Senecio* species, some studies have reported profound changes in gene expression in new hybrids, favoring the survival of these plants in new environments not accessible to their parents (Chapman et al.; 2013, Hegarty et al., 2008). In sunflower, Lai et al., (2006) found that environmental pre-adaptation was driven mostly by expression divergence (reviewed by Mack & Nachman, 2017). However, in forest trees, very few such studies have been performed (reviewed by (Lexer et al., (2004)).

European white oaks are an excellent model for investigating such divergence in expression. These sister species diverged recently (<10 million years ago (mya), Hubert et al., 2014; Manos & Hipp, 2021) and have probably experienced a series of periods of allopatric isolation and contact associated with glacial retreats and postglacial recolonizations (Leroy et al., 2017; 2020). Within this species complex, the speciation of sessile oak (*Quercus petraea* Matt. Liebl., hereafter SO) and pedunculate oak (*Quercus robur* L. hereafter PO) has been studied in detail (reviewed in (Kremer & Hipp, 2020). In “On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life”, Darwin himself — in an attempt to define species and speciation introduced pedunculate and sessile oaks as an example illustrating the difficulties of clearly defining different species on the basis of very closely related but different life forms. Indeed, PO and SO are known to hybridize (Lepais & Gerber, 2011) and their distinction on the basis of morphological criteria is not straightforward, to the extent that some evolutionary biologists have described them as “botanical horrors” (Rieseberg, Wood, & Baack, 2006). That said, they do have known different ecological preferences imposed largely by environmental factors, such as light, soil composition and humidity (Epron & Dreyer, 1990), although there is some overlap between their habitats (Leroy et al., 2020). In brief, SO is more tolerant of water shortage (Bréda, et al., 1993) and has a higher water-use efficiency than PO (Ponton et al, 2002). It is therefore better able to survive both chronic and extreme droughts (Bréda & Badeau, 2008). The severe heat waves and prolonged rainfall deficit observed in mainland France in 1976 and 2003 provide a clear illustration of the long-term consequences of such events in terms of oak tree decline in France and, more broadly, in Europe (Breda et al., 2006; van der Werf et al., 2007). Remarkably, these events had different effects on PO and SO (Thomas & Hartmann, 1996), with SO displaying higher growth rates during these dry years (Becker et al., 1994; Bréda & Badeau, 2008; Friedrichs et al., 2009; Lebourgeois, 2006) and PO populations having a higher mortality

rate (Durand et al., 1983; Lebourgeois et al., 2015). Similarly, Truffaut et al. (2017) provided evidence for changes in species occupancy at a fine spatiotemporal scale in a mixed stand of these two species, with an expansion of the SO population attributed to differences in survival after the repeated drought events observed in the study area (west of France) over the last three decades. Conversely, PO can tolerate long periods of waterlogging and is generally found in moist habitats.

Based on these observations and their consequences in terms of the adaptation of oak forest management to a warming climate, research has focused on characterizing the differences between SO and PO, with the aim of incorporating the knowledge obtained into management decisions. Differences between SO and PO have mostly been addressed at the phenotypic level (Parelle et al. (2007), for traits relating to hypoxia tolerance (Dupouey & Badeau, 1993; Kremer et al., 2002), leaf and fruit morphology and water use efficiency (Ponton et al., 2002). Leroy *et al.* (2020) recently addressed this question from a population genomics perspective, making use of the the oak genome sequence (Plomion et al. 2018) to perform genome scans for divergence between the four main European white oak species, including PO and SO. They identified candidate genes for reproductive isolation between species, including a large number of transcription factor genes. This study was particularly informative about the reproductive barriers between temperate (*Q. robur* & *Q. petraea*) and Mediterranean species (*Q. pyrenaica* & *Q. pubescens*), but shed little light on the genes contributing to the specific reproductive barriers between *Q. robur* and *Q. petraea*.

Only a few studies have investigated the differences in gene expression patterns between SO and PO. Porth et al. (2005) provided a first insight into species differences in mRNA levels between these two species in conditions of water deficit. They detected an upregulation of osmotic adjustment-related genes in SO. In a study based on suppressive subtractive hybridization, Le Provost et al. (2012) provided further insight into the genes potentially involved in the adaptation



of PO to waterlogging. In particular, they identified genes with expression levels displaying strong species-by-environment interactions. These observations suggested that PO had evolved specific strategies to cope with waterlogging. They also showed that tolerance to waterlogging was driven by a rapid switch of PO metabolism to the fermentative pathway in hypoxic conditions, a strategy that could alleviate energy loss caused by oxygen deprivation.

Our objective here was to investigate in more detail the contribution of gene expression to the maintenance of these two species in sympatry. We subjected PO and SO seedlings to environmental conditions mimicking the ecological constraints (i.e. an excess or deficit of water) to which they are adapted in natural conditions. We then harvested root tips, quantified gene expression by RNA-seq and identified the main factor (species, environment, their interaction) accounting for differential gene expression. We hypothesized that genes displaying a significant “species x environment” interaction effect would shed light on the different molecular strategies that had evolved in each species for adaptation to the most frequently encountered type of water stress, whereas genes displaying a significant “species” effect regardless of environmental conditions (i.e. always upregulated in either PO or SO) would highlight important functions underlying intrinsic barriers between these two species. We also combined our expression data with data from a whole-genome scan. We showed that the two categories of candidate genes described above were enriched in SNPs displaying a high level of genetic differentiation between the two species, confirming that these candidate genes play an important role in building extrinsic and intrinsic reproductive barriers between these two co-occurring species.

Overall, this study suggests that genes involved in molecular plasticity to the environmental conditions faced by these two species in their respective ecological niches are targets of divergent selection between the two species. This finding improves our understanding of the genetic basis of ecological speciation in European white oaks, and the identification of these candidate barrier loci

is of prime importance for the design of sustainable oak forest management strategies to guide the trajectory of these species more effectively in the current context of climate change.

## Materials and Methods

### *Plant material*

The plant material used here was described in a previous study by Le Provost *et al.* (2016). Briefly, we sampled half-sib progenies from three unrelated mother trees for both SO (located in “Foret Domaniale” of Laveyron, latitude 43°45’49”N, longitude 0°13’11”W) and PO (from “Ychoux forest”, latitude 44°33’33”N, longitude –0°96’66”W). Before germination, we confirmed the species status of the mother trees with the diagnostic SNP markers described by Guichoux *et al.* (2011). We harvested the acorns in the fall of 2013 and sowed them in a 1:1 mixture of peat and sand in 0.2 L pots. After germination, we transferred the seedlings to a greenhouse (16-hour photoperiod, 25°C during the day and 20°C at night) for five weeks, until they had three fully developed leaves.

### *Experimental design and measurement of physiological traits*

Sixty homogeneous seedlings (i.e. with three fully developed leaves) were selected for each mother tree and species, corresponding to a total of 360 seedlings (60 half-sibs \* 3 mother trees used as biological replicates \* 2 species) and exposed to a waterlogging/drought stress cycle. An overview of the experimental design is presented in Supplementary Figure 1.

(i) Waterlogging treatment: The 360 seedlings were placed in five 100 L plastic containers corresponding to five blocks. The plastic containers were filled with water that was deoxygenated directly by bubbling with N<sub>2</sub>. The seedlings were immersed in the water such that the water level reached 2 cm above the collar. The O<sub>2</sub> concentration in each container was monitored daily with a portable O<sub>2</sub> electrode (Cellox 325, WRW, Weilheim, Germany). Each block comprised 12 siblings

from the same species and mother tree. We sampled white roots from 10 seedlings (i.e. two in each of the five blocks) for each species and each mother tree after 0 (control), 6 hours and 24 hours (short-term stress) and 9 days (long-term stress) of hypoxia. The roots were immediately frozen in liquid nitrogen and stored at -80°C until RNA extraction. The 10 seedlings were pooled for further RNA-seq analysis.

(ii) Drought stress treatment: We removed the remaining 120 seedlings from the deoxygenated solution on day 10. Once the excess water had percolated through the substrate by gravity, the seedlings were maintained at field capacity for three weeks and were then subjected to drought stress (Supplementary Figure 1). Drought stress was achieved by keeping water levels in the substrate at 15% of field capacity for nine days. Before the start of the experiment, we determined the water retention capacity of the substrate by weighing the pots at field capacity and then again after drying the substrate (at 65°C for 24 hours). The amount of substrate per pot was also determined before sowing the acorns, to make it possible to measure the maximum water content per pot (~170 g) independently of the biomass produced. Drought stress was then applied in the greenhouse, as previously described (Marguerit *et al.*, 2012). We monitored the amount of water in the substrate daily, by weighing the pots with a precision of +/- 1 g (Sartorius, Aubagne, France), until the water content in the pots reached 15% field capacity (corresponding to a weight loss of 144 g). Pots that dried faster were maintained at 15% water content for nine days before the harvesting of white roots. In parallel, we maintained a subset of seedlings at field capacity to serve as a control. We harvested white roots from these seedlings at the same time as from the stressed sample to prevent confounding ontogenetic effects (Supplementary Figure 1). As for the waterlogging treatment, we also generated three biological replicates for gene expression analysis by pooling the seedlings from the same mother tree (i.e. 10 seedlings per mother tree, Supplementary Figure 1). We also measured leaf predawn and midday water potential both for the

controls and treated seedlings before each sampling, using a Scholander-type pressure chamber as described by Hsiao (1990). Measurements were made on five seedlings each for the control and the drought-stressed treatment.

### *RNA extraction and sequencing*

RNA was extracted and purified as described by Le Provost *et al.* (2007). Residual genomic DNA was removed with an RNase-free DNase, RQ1 (Promega, Madison, WI, USA), according to the manufacturer's instructions, before purification. We assessed the quantity and quality of the extracted RNA on an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA).

For the waterlogging treatment, we constructed nine libraries for each species, corresponding to three biological replicates x three treatments (control, short-term response assessed by pooling equimolar amounts of total RNA extracted after six and 24 hours, and long-term response assessed on RNA extracted after nine days of waterlogging).

For the drought stress treatment, we generated six libraries for each species, corresponding to three biological replicates x two treatments (control and long-term response, i.e. after nine days of water levels at 15% field capacity in the substrate).

We generated these 30 libraries as described by Le Provost *et al.* (2016) and according to the Illumina protocol (TruSeq Stranded mRNA Sample Prep Kit). Briefly, we selected mRNA from 2 µg of total RNA. The mRNA was then chemically fragmented and reverse transcribed with a random hexamer primer. The second strand of the cDNA was generated, 3'-adenylated, and Illumina adapters were added. We amplified DNA fragments (with adapters) by PCR with Illumina adapter-specific primers. We quantified the libraries with a Qubit Fluorometer (Life Technologies,

NY, USA). We estimated their size with Agilent 2100 Bioanalyzer technology (Agilent). We then sequenced each library by 101 base- read chemistry, in a paired-end flow cell, on an Illumina HiSeq2000 (Illumina, San Diego, CA, USA). More than 24 million useable reads were generated for each library (Supplementary Table 1).

### *Cleaning, mapping and identification of differentially expressed genes*

We applied the following two-step procedure to identify differentially expressed genes. We first removed low-quality reads (quality value<20). The high-quality reads were then aligned with the 25,808 oak gene models previously published with the reference oak genome (Plomion et al., 2018), with the BWA mapper V0.6.1 aligner (Li & Durbin, 2009) and a maximum insert size of 600 bp, with four mismatches allowed. We then selected gene models with at least 90 reads (i.e. over all biological replicates) for the differential analysis. In the second step, we used the DESeq2 package (Love et al., 2014) to identify differentially expressed genes with a  $p$ -value<0.01 after adjustment for multiple testing with a false discovery rate (FDR) of 5%. The expression level of each gene was quantified by the TMM method. We considered only differentially expressed genes for which at least a two-fold change in expression was observed. For each experiment, we assessed the effects of treatment, species and their interaction in likelihood ratio tests implemented in the DESeq2 package.

The treatment and species effects were assessed by comparing a model without interaction (M1) with two simplified models for the treatment (M2) and species (M3) effects.

M1:  $Y_{ijk} = \mu + T_i + S_j + \varepsilon_{ijk}$  where  $T_i$  is the treatment effect ( $i$ =“control”, “short-term”, “long-term” for waterlogging and  $i$ =“control” or “long-term” for drought stress), and  $S_j$  is the species effect ( $j$ = “PO” or “SO”).

$$M2: Y_{jk} = \mu + S_j + \varepsilon_{jk}$$

$$M3: Y_{ik} = \mu + T_i + \varepsilon_{ik}$$

For interaction effects, we compared a complete model:  $Y_{ijk} = \mu + T_i + S_j + (S*T)_{ij} + \varepsilon_{ijk}$  (M4) to M1.

Differential gene expression analyses therefore yielded six genes sets: (#1) genes differentially expressed between waterlogged and control conditions (across species), (#2) genes differentially expressed between drought and control conditions (across species), (#3) genes differentially expressed between species throughout the waterlogging experiment, (#4) genes differentially expressed between species in the drought stress experiment, (#5) genes displaying significant treatment-by-species interaction in the waterlogging experiment, and (#6) gene displaying significant treatment-by-species interaction in the drought stress experiment. An additional gene set was also generated from the overlap between genes displaying a “species” effect regardless the treatment considered (gene set #7).

Annotations for the differentially expressed genes were recovered from the reference oak genome (Plomion et al., 2018).

### *Geneset and subnetwork enrichment analysis*

We performed GO-term enrichment analysis for the six gene sets described above with the topGO R package (Alexa, 2010). We corrected  $p$ -value for false discovery rate and considered ontology terms with a corrected  $p$ -value below 0.05 to be significantly enriched.

We then performed enrichment analysis with Pathway Studio software (Pathway Studio®, Elsevier 2017), as described by Le Provost et al., (2016). Our main goal was to identify genes

potentially involved in ecological preferences. Therefore, in each type of experiment, we focused on the sets of genes differentially expressed between species (gene sets #3, #4 and #7) and the set of genes presenting significant treatment-by-species interactions (gene sets #5 and #6). Indeed, we expected (i) genes displaying significant treatment-by-species interactions to capture the different molecular strategies to the two water regimes evolved by each of the species, and (ii) genes displaying a species effect to have higher basal levels of expression in the tolerant species (i.e. PO for excess water and SO for water deficit). An additional network was generated with genes displaying a species effect whatever the stress considered. The previous set of genes may reveal important mechanisms underlying extrinsic barriers, whereas this last gene set should reveal genes involved in reproductive isolation (intrinsic barriers) between PO and SO.

#### *Reverse transcription and quantitative real-time PCR (qPCR)*

For each experiment, we selected three genes displaying a significant treatment, species or interaction effect for qPCR validation. All the genes analyzed and their associated effects are listed in supplementary Table 2. We performed qPCR quantification as described by Le Provost *et al.* (2012) on 1 µg of RNA, on a Chromo4™ Multicolor Real-Time PCR detection system (Bio-Rad Laboratories, Hercules, CA, USA) with standard PCR parameters. The fluorescence data obtained were analyzed with the Genex macro. This macro uses an algorithm derived from that described by Vandesompele *et al.* (2002) for qPCR normalization and quantification. We normalized fluorescence data against two control genes (Qrob\_P0517760.2 encoding a hypothetical protein and Qrob\_P0217660.2 encoding a ubiquitin-protein ligase RHF1A). All the primer pairs used for qPCR were designed with Primer3 software (Rozen & Skaletsky, 2000) (Supplementary Table 2). We selected control genes as follows. We first mined our digital expression analysis to identify genes that were not differentially expressed. We then selected six genes from this gene set for



which coverage was similar in each biological replicate. We then evaluated the stable expression of these six genes in our biological samples with Genorm software (Vandesompele et al., 2002). We identified two genes with very stable expression levels of expression (Qrob\_P0517760.2: “Unknown protein” and Qrob\_P0217660.2: “Ubiquitin protein ligase RHF1A”). This validation was performed with the same trees as for RNAseq analysis, with independent RNA extractions. This analysis therefore served as a technical verification of the RNA-seq approach.

### *Relative genetic divergence between differentially expressed genes*

We investigated whether differentially expressed genes presented evidence of potentially adaptive genetic divergence, by assessing the overlap between the six gene sets and loci with high  $F_{ST}$  fixation indices between PO and SO. Divergence estimates were obtained with the pool-seq data recently described by Leroy et al., (2020).  $F_{ST}$  was calculated with the popoolation2 software suite (Kofler et al., 2011) for each SNP. We then assessed enrichment in high  $F_{ST}$  SNPs among the genes of each of the six gene sets. We defined high  $F_{ST}$  SNPs empirically by considering the 1, 0.5, 0.1, 0.01 and 0.001% right tails of the genome-wide  $F_{ST}$  distribution. We first calculated the proportion of genes overlapping with high- $F_{ST}$  SNPs for each gene set and each threshold. For each gene set of size  $N$ , we then built a null distribution by randomly sampling  $N$  genes 1,000 times, each time calculating the proportion of randomly drawn genes overlapping with high- $F_{ST}$  SNPs. Enrichment was calculated by dividing the proportion of genes from each gene set overlapping with high- $F_{ST}$  SNPs by the mean proportion of random genes overlapping with high- $F_{ST}$  SNPs in the corresponding null distribution. We determined the significance of enrichment by comparing the proportion of genes from each gene set overlapping with high- $F_{ST}$  SNPs from the corresponding null distribution.

## Results

### *Characterization of the treatments applied*

For waterlogging, we measured the O<sub>2</sub> concentration in each container daily (Supplementary Figure 2A). The O<sub>2</sub> concentration in the water in the container decreased sharply just after the start of N<sub>2</sub> bubbling and became stable (close to 1 mg l<sup>-1</sup>) after 5 minutes. N<sub>2</sub> bubbling was then stopped. The low level of O<sub>2</sub> in the water was immediately consumed by the root system, maintaining O<sub>2</sub> concentration in the water close to 1 mg l<sup>-1</sup> throughout the experiment. The samples harvested in this experiment may therefore be considered to have been under hypoxic stress (i.e. only residual O<sub>2</sub> in the rhizosphere). Two thirds of the seedlings ( $N=240$ ) were harvested during the waterlogging experiment. We found visible morphological differences between the species after nine days of waterlogging stress, with hypertrophied lenticels observed only in the immersed portion of the stem of the tolerant species (Le Provost et al., 2016).

The remaining seedlings ( $N=120$ ) were used in the drought stress experiment. We considered seedlings to be drought-stressed only if the combined weight of seedling, 0.5 L pot and soil fell to below 144 g (i.e. approx. substrate at about 15% of field capacity). We maintained all the seedlings at this threshold for nine days by weighing the pots daily and adjusting their weights with an appropriate volume of water. An analysis of the dehydration kinetics (Supplementary Figure 2B) revealed that all seedlings reached the desiccation threshold at about the same time. We measured leaf predawn and midday water potentials for each species just before sampling (Supplementary Figure 2C). We performed the same measurements on the control seedlings harvested on the same date, to prevent confounding ontogenetic effects. For both PO and SO, an increase (in absolute

value) in leaf predawn and midday water potential was observed in response to stress (comparison of control and stressed samples). The midday water potentials of stressed plants were 1.5 MPa for PO and 1.7 MPa for SO. Predawn water potentials were close to 1 MPa. These results suggest that the water stress applied in this experiment was moderate, probably due to the medium used, which contained 50% peat. Bréda *et al.* (1993) reported a leaf predawn water potential value of 2 MPa for SO seedlings experiencing severe water stress. Here, we observed significant differences in leaf predawn and midday water potentials between the two treatments (i.e. control vs. stress), for both species. The difference between PO and SO under conditions of drought stress was not significant although the difference in midday water potential between the control and stressed samples was smaller for SO, suggesting a better tolerance to water stress.

#### *Global characterization of transcriptomic data*

PO and SO seedlings were subjected to a cycle of waterlogging and drought stresses. We monitored the intensity of the constraints applied through observations and measurements throughout the experiment (see Supplementary Figure 1). Root tips were then sampled and 30 cDNA libraries were constructed from mRNA extracted from these tissues: 18 libraries for waterlogging and 12 for drought conditions. For each library, 23 to 43 million reads were generated. An overview of the sequenced libraries is shown in Supplementary Table 1, along with their SRA accession numbers. For the waterlogging experiment, three sampling points were analyzed (control C, short term ST, and long term LT responses) for both species. For the drought stress experiment, we analyzed two sampling points in both species (control and long-term response, after 9 days of water deficit). We obtained a mean Phred score of 35 for the libraries analyzed in this study. Finally, we successfully mapped 82% to 88% of the sequences from each library onto oak gene models.

We investigated the reproducibility of biological replicates and the correlation between the various samples by performing principal component analysis (PCA) on the waterlogging (18 samples, Figure 1A) and drought stress (12 samples, Figure 1B) datasets. This descriptive analysis showed close clustering within biological replicates, attesting to the quality of the RNAseq data. It also showed that the samples (species x treatment) clustered into different groups corresponding to the different combinations analyzed in this study. However, whereas the PCA for the drought stress samples displayed clear differences in expression between species and between treatments (Figure 1B), that for the waterlogged samples revealed clear differences only between the three treatments (Figure 1A). Subtle differences in terms of the number of genes and/or changes in expression between species therefore probably remain to be discovered by other approaches.

The adjusted data were also evaluated with a clustering approach in Expander software (Shamir et al., 2005). This analysis confirmed the high degree of reproducibility for the RNAseq data, with biological replicates grouped together in the same cluster (Figure 1).

### *Number and effect size of differentially expressed genes*

DEGs (i.e. with an absolute fold-change in expression  $\geq 2$  and adjusted  $p$ -values  $\leq 0.01$ ) were identified separately for the waterlogging and drought stress experiments with the DESeq2 package. We used the likelihood ratio test to identify genes displaying significant “species”, “treatment” and/or “interaction” effects.

For the waterlogging experiment, 18,155 genes covered by more than 90 reads remained for abundance analysis. Most of these genes (18,146, 99.9%) were covered by sequences from both species. We identified five genes specific to PO (i.e. covered by PO reads only): Qrob\_P0335670.2

(germacrene-D-synthase), Qrob\_P0495260.2 (glucose-induced degradation protein), Qrob\_P0281130.2 (kinase with adenine nucleotide alpha hydrolase-like domain), Qrob\_P0336470.2 (PPPDE thiol peptidase family protein), Qrob\_P0125380.2 (SABATH methyltransferase). We also identified four genes specific to SO: Qrob\_P0362510.2 (myb-like HTH transcriptional regulator family protein), Qrob\_P0119020.2 (cysteine-rich repeat secretory protein 38), Qrob\_P0598470.2 (subtilase family protein T6P5), and Qrob\_P0605080.2 (14-3-3-like protein GF14 omega protein).

We also identified 820 genes with a significant treatment effect, 1,167 genes with a significant species effect, and 319 genes with a significant interaction (Table 1, Supplementary Files 1, 2, and 3 and Figure 2A and 2B). Only 25 genes had all three types of effect (Figure 2B), but 183 genes had a significant treatment effect only, 522 had a significant species effect only and 278 genes had an interaction effect.

For the drought stress experiment, 17,688 genes were considered according to the criteria defined above. We identified 11 genes specific to either PO or SO: two genes for PO (Qrob\_P0374520.2 (unknown protein) and Qrob\_P0084850.2 (cyclic nucleotide gated channel 1) and nine genes for SO (Qrob\_P0711010.2 (unknown protein), Qrob\_P0675500.2 (aspartyl protease family protein), Qrob\_P0092630.2 (TIR-NBS-LRR class disease resistance protein), Qrob\_P0627090.2 (HXXXD-type acyl-transferase family protein), Qrob\_P0446600.2 (NDR1/HIN1-like protein), Qrob\_P0169890.2 (unknown protein), Qrob\_P0684400.2 (cycloartenol synthase), Qrob\_P0362510.2 (myb-like HTH transcriptional regulator family protein) and Qrob\_P0593920.2 (DUF946 family protein).

Application of the same statistical criteria as for the waterlogging experiment led to the identification of 122 genes with a significant treatment effect, 395 genes with a significant species effect, and 17 genes with a significant interaction (Table 1, Supplementary File 4, 5 and 6, Figure

2A and 2B). The overlap between these three types of effects is shown in the Venn diagram of Figure 2B. No gene displayed all three types of effect simultaneously, and we found 68 genes with a significant treatment effect only, 339 genes with a significant species effect only and 15 genes with a significant interaction.

We found that 184 of the genes displaying a “species” effect (i.e. 13.4% of the “species” effect genes) were differentially expressed in both experiments (Supplementary File 7). This overlap is highly significant overlap (chi<sup>2</sup> test on a 2x2 contingency table: chi<sup>2</sup> = 1154, *p*-value < 2.2e-16). These 184 genes displaying a species effect whatever the stress considered are potentially relevant as genes involved in reproductive isolation (intrinsic barriers) between PO and SO. These genes are listed in supplementary file 7.

#### *Validation of RNA-seq results by qPCR*

We performed reverse transcription-quantitative PCR analysis for 10 genes (six identified in the waterlogging experiment and four identified in the drought stress experiment, respectively). The expression patterns of the tested genes revealed by qPCR were similar to those obtained by RNAseq (see supplemental information 1), confirming the reliability of the RNA-seq data.

#### *Gene set and subnetwork enrichment analyses*

The main goal of this study was to highlight the molecular mechanisms involved in adaptation to waterlogging or drought stress, in PO and SO, respectively, rather than to identify genes showing a plastic response to these constraints. We therefore chose to perform gene set and subnetwork

enrichment analyses exclusively on the genes displaying a significant interaction (in either experiment, i.e. gene sets #5 and #6) or species (across experiments, i.e. gene set #7) effect. Indeed, we hypothesized that the interaction term would reveal genes defining species-specific adaptive molecular strategies underlying reproductive barriers, whereas genes displaying a “species” effect, whatever the treatment considered, would reveal intrinsic functions driving the reproductive isolation between PO and SO, regardless of environmental conditions. Genes displaying a “species” effect in specific environmental conditions (gene sets #3 and #4) may also constitute an important source of information because higher basal levels of expression in the tolerant species may explain its better adaptation to the environmental constraints with which it is usually confronted in natural conditions. Below, we focus exclusively on the set of genes displaying a “species” effect across experiments. The species-specific genes displaying a response to either waterlogging or drought stress separately are presented in Supplemental Information 2.

(i) *Gene set enrichment analysis.*

The results are shown only for the first 100 gene ontologies in Supplementary files 3, 6 and 7. A graphical representation of the first 20 ontologies is also available in the same supplementary Files for a clearer view of the ontologies regulated.

For interaction-responsive genes, the highest levels of enrichment during waterlogging were observed for the biological processes (BPs) “positive regulation of flavonoid biosynthetic process”, “regulation of meristem structural organization”, “secondary shoot formation”, “sphingolipid biosynthetic process” and “activation of MAPK activity involved in osmosensor”, and for the molecular functions (MFs) “phosphorelay response regulator activity”, “ubiquitin-protein

transferase activity”, “peptide deformylase activity”, “sphingolipid delta-8 desaturase activity”. Very different BPs (“regulation of membrane potential”, “calcium ion transport”, “potassium ion transport” and “amino acid import”) and MFs (intracellular cyclic nucleotide activated cation channel”, “inward rectifier potassium channel activity”, “cAMP binding”, “cGMP binding” and “calmodulin binding”) were found to be enriched during drought conditions (Supplementary File 3 and 6 and Figure 2C).

The results of the gene set enrichment analysis for genes displaying a “species” effect whatever the treatment considered are shown in supplementary file 7 and Figure 2D. The BPs “oxidation-reduction process”, “response to stimulus”, “defense response”, “triterpenoid biosynthetic process” and “triterpenoid metabolic process” were found to be enriched, and enrichment was observed for the MFs “oxygen binding”, “ADP binding”, “oxidoreductase activity”, “iron ion binding” and “drug binding”.

(ii) Subnetwork enrichment analysis was performed to highlight key molecular players involved in adaptation to waterlogging in PO or drought stress in SO. A network based on interaction-responsive genes was generated for each treatment. The functional networks are shown in Figure 3 for waterlogging and Figure 4 for drought stress. The genes with an interaction effect in waterlogged conditions formed several important hubs related to “root growth”, “response to auxin stimulus”, “jasmonate response”, “meristem function”, “cell expansion”, “plant defense” and “cell death”. Different hubs were identified for drought stress, with molecular functions related to “defense response”, “response to auxin stimulus”, “growth yield” and “seed yield”.

An additional subnetwork was also generated from the genes displaying a “species” effect whatever the treatment analyzed. This network is shown in Figure 5. We identified hubs related to:



“flowering time”, “cell differentiation”, “cell proliferation”, “developmental process” and “seed germination”. Meristematic cells generally display an enrichment in these molecular functions. This result adds weight to the hypothesis that this last gene set contains genes potentially involved in the formation of intrinsic barriers driving the reproductive isolation between PO and SO, regardless of environmental conditions.

*Overlap between loci showing a high degree of genetic differentiation between species and differentially expressed genes*

For this comparison, we used a published dataset from a study investigating adaptive differentiation between PO and SO based on pool sequencing (Leroy et al., 2020). The mean  $F_{ST}$  value across the 39,644,639 SNPs with  $F_{ST} > 0$  was 0.0545, suggesting weak overall genetic differentiation. However,  $F_{ST}$  estimates ranged from  $1.17e^{-6}$  to 1. We assessed enrichment in gene sets identified by differential expression analysis for SNPs in the 1, 0.5, 0.1, 0.01, and 0.001% right-hand tails of the genome-wide  $F_{ST}$  distribution, corresponding to thresholds of 0.14, 0.217, 0.454, 0.803, and 1, respectively (Figure 5). In each of the six gene sets (#1 to #6), we found significant enrichment in SNPs displaying a high degree of differentiation between the two oak species (Figure 6). Enrichment ratios exceeded six-fold for the species and treatment effects in the drought experiment, but all six gene sets displayed significant enrichment, suggesting that the sets of genes differentially expressed in our experiments included genes related to adaptive differentiation between the two species. The lists of genes including highly differentiated markers for the various thresholds are provided in Supplementary File 8.

We also assessed the enrichment in high- $F_{ST}$  SNPs among genes displaying differential expression between species and across treatments (Figure 6, rightmost barplot). Again, we found a significant overlap, suggesting potential involvement in intrinsic barriers to reproduction between PO and SO.

## Discussion

The few studies published to date on the molecular plasticity of European white oaks during waterlogging or drought stress were all performed in single species (Folzer et al., 2006; Parent et al., 2008; Spieß et al., 2012), limiting the significance of the results produced with respect to the role of gene expression in species ecological preferences. Moreover, these studies were generally performed with targeted genomic approaches and therefore focused on a limited number of genes, again reducing the scope of the findings concerning the molecular mechanisms driving oak adaptation. We investigated the contribution of gene expression to the divergence of PO and SO in a factorial design in which each species was subjected to two sets of water stress (excess or deficit) corresponding to the environmental constraints to which these two species are better adapted. This made it possible to focus on the “species x treatment” interaction term and to reveal species-specific adaptive molecular features presumably involved in the ecological preferences of these species (319 DEG for waterlogging and 17 for drought stress). This experiment also made it possible to identify genes expressed differentially between species regardless of the treatment imposed (184 genes), suggesting intrinsic roles in driving the reproductive isolation between PO and SO. In addition to these two sets of DEGs (the focus of the discussion), genes differentially expressed between the species in each treatment may also provide an important source of information, because genes with higher basal levels of expression in tolerant than in intolerant species may

contribute to adaptation to a specific environmental condition. A dedicated discussion on these genes is provided in Supplemental Information 2.

Molecular mechanisms underlying the ecological preferences of PO and SO

We focus here on the interaction-responsive genes identified in the waterlogging and drought stress experiments. We hypothesized that genes with expression levels significantly affected by the species x treatment interaction would include genes involved in the respective ecological preferences of the two species that could potentially be involved in ecological barriers.

### **Waterlogging experiment**

Waterlogging had a stronger effect than drought stress in terms of the number of DEGs (319) displaying “species x treatment” interactions. The subnetwork generated (Figure 3) revealed important hubs related to root growth (“root growth”), hormone signaling (“response to auxin stimulus” and “jasmonate response”), cell differentiation and expansion (“meristem function” and “cell expansion”), plant defense and cell death, suggesting a key role for molecular mechanisms triggering the formation of aerenchyma and adventitious root formation. In PO, Parelle et al., (2006) reported that morphological and anatomical changes (such as the formation of adventitious roots and aerenchyma) were essential to cope with hypoxia. They also reported the formation of larger amounts of these adaptive structures in PO than in SO. This formation of aerenchyma or adventitious roots in waterlogging-tolerant plants may be triggered by genes belonging to the main hubs identified here, suggesting that the molecular mechanisms identified here are of particular

importance for the adaptation of PO to waterlogging. We review some of the genes identified below. In the following section, we focus our discussion on some genes taking into account their  $F_{ST}$  value.

### (1) Programmed cell death leading to aerenchyma formation

Programmed cell death (PCD) is a major mechanism underlying plant development and adaptation to abiotic or biotic stress (Smith et al., 2015). In particular, it is involved in the modification of root architecture under drought stress and the formation of aerenchyma in hypoxic conditions. Tolerant plants typically produce aerenchyma in response to hypoxic conditions through the lysis and death of cells in the root cortex to produce additional gas space (Yamauchi et al., 2013). The tolerance to waterlogging observed in PO may be partially explained by its ability to produce more aerenchyma than SO (Parelle et al., 2006), a finding confirmed by our network analysis, which identified several genes related to cell death, a process essential to aerenchyma formation (Figure 4). The genes upregulated in PO included (i) *LCB1* (*Qrob\_P0060010.2*,  $F_{ST}:0.43$ ) encoding a long-chain base 1 (LCB1) subunit of serine palmitoyltransferase. In *Arabidopsis thaliana*, this gene has been shown to regulate PCD by inducing reactive oxygen intermediates (Shi et al., 2007), (ii) *ATG6* (*Qrob\_P0479390.2*,  $F_{ST}:0.19$ ) which encodes a core autophagy-related protein involved in the regulation of autophagy (Xu et al., 2017), (iii) *WRKY 9* (*Qrob\_P0436950.2*,  $F_{ST}: 0.24$ ), which is known to be involved in cell death regulation in tobacco (Liu and He, 2017) and (iv) *SKP1* (*Qrob\_P0747290.2*,  $F_{ST}:0.2$ ) encoding S-phase kinase-associated protein 1 and including multiple differentiated SNPs with a maximum  $F_{ST}$  value of 0.73. To our knowledge, this is the first report linking potentially *SKP1* protein to aerenchyma formation.

## (2) Root growth and phytohormone biosynthesis

The formation of adventitious roots is another major morphological response to waterlogging in tolerant species. These roots contain aerenchyma (i.e. air channels), which favors the diffusion of gases in conditions of submersion (Evans, 2003). We identified important hubs potentially involved in the establishment of adventitious roots (root growth, cell expansion, meristem function) (Figure 3). We also detected hubs related to phytohormones (response to auxin stimulus and jasmonate response) known to play a key role in the formation of adventitious roots. Our analysis thus highlighted molecular networks related to this morphological adaptation in PO.

Our analysis highlighted genes (upregulated in PO) involved in either primary root growth inhibition or lateral root development. A first subset of these genes is involved in auxin signaling (*HDA6*, *WRKY23*, *AFB3*) or jasmonate responses (*PHYB*, *SKP1*, *4CL2*, *MYC2*) consistent with the modification of the root system during hypoxic responses in PO being driven by phytohormones. For instance, *MYC2* (*Qrob\_P0056680.2*,  $F_{ST}$ : 0.24) encodes a basic helix-loop-helix (bHLH) DNA-binding protein. Kazan and Manners (2013) identified this gene as a master regulator of many aspects of jasmonate signaling playing a strong role in the formation of lateral and adventitious roots. *EXO* (*Qrob\_P0764920.2*,  $F_{ST}$ : 0.14) is under the control of brassinosteroid. Its overexpression in *Arabidopsis* has been reported to promote shoot and root growth (Schröder et al., (2009)).

*RBOHB* (*Qrob\_P0493470.2*,  $F_{ST}$ : 0.28), which encodes a respiratory burst oxidase protein involved in the production of reactive oxygen species, is also upregulated in PO. The reactive oxygen species (ROS) produced by the product of *RBOHB* genes are involved in several developmental processes. Montiel et al., (2013) reported that the overexpression of this gene promotes lateral root growth,

suggesting a potential role in adventitious root formation in PO. We also identified *NINJA* (*Qrob\_P0009970.2*,  $F_{ST}$ : 0.72) as being more strongly expressed in PO. This gene encodes an interactor of the jasmonate Zim domain 10. In a study of *Arabidopsis* mutants, Acosta et al. (2013) reported a downregulation of jasmonate signaling by *NINJA* that was associated with a large decrease in the size of the root system. We hypothesize that the overexpression of *NINJA* in PO is associated with its ability to maintain adventitious root formation.

Two other genes were found to be downregulated in PO: *HDA6* (*Qrob\_P0451700.2*,  $F_{ST}$ : 0.82), which encodes an RPD3-type histone deacetylase, a key enzyme involved in the jasmonate response (Wu et al., (2008), and *CYP94B3* (*Qrob\_P0369790.2*,  $F_{ST}$ : 0.49), which encodes a cytochrome P450 protein involved in jasmonate catabolism. Heitz et al. (2012) reported a correlation between *CYP94B3* overexpression and insensitivity to jasmonate and a smaller lateral root system. These results highlight a stronger inhibition of the jasmonate response in PO, which is essential for lateral root formation, potentially accounting for the increase in lateral root system size in PO during waterlogging.

### (3) Meristem function and response to auxin stimulation

The *WRKY23* gene (*Qrob\_P0697030.2*,  $F_{ST}$ : 0.59), which is involved in the auxin response, was more strongly expressed in PO. Grunewald et al., (2012) reported an important role for this gene in the establishment of an auxin gradient in the root system, allowing the maintenance of root meristematic activity and lateral root formation.

The *AFB3* gene (*Qrob\_P0036220.2*,  $F_{ST}$ : 0.75), which encodes an auxin receptor F box protein, was also upregulated in PO. *AFB* genes belong to a multigene family regulated by auxin and

involved in several plant developmental processes. In *Arabidopsis thaliana*, Vidal et al. (2010) showed that the *AFB3* mutant was characterized by altered lateral root growth, whereas the primary root system was unaffected, suggesting a key role for AFB3 in modulating the architecture of the root system.

Overall, our network analysis highlighted several molecular mechanisms potentially involved in the formation of adaptive structures known to favor oxygen diffusion in the root system (aerenchyma and adventitious roots). This suggests that molecular mechanisms analogous to those observed in other tolerant plant species are upregulated in PO relative to SO.

### ***Drought stress***

The functional network obtained from the 17 genes found to be differentially expressed in response to drought is presented in Figure 4. Three genes upregulated in SO were considered to be of particular importance (*PP2-A1*, *CT-BMY*, and *WAT1*). The *WAT1* gene (*Qrob\_P0331600.2*,  $F_{ST}$ : 0.23) encodes a vacuolar auxin transporter. Auxin plays a crucial role in abiotic stress responses controlling vascular development (Ranocha et al., 2013). Zhang et al. (2017) reported that this gene was upregulated in drought-tolerant maize lines during drought stress. We also identified a *CT-BMY* gene (*Qrob\_P0567500.2*,  $F_{ST}$ : 0.25) encoding a beta-amylase known to be upregulated during abiotic stress. The upregulation of this gene is generally correlated with maltose accumulation, which protects proteins, membranes and the photosynthetic electron transport chain during water stress (Kaplan & Guy, 2004). Finally, we identified *PP2-A1* (*Qrob\_P0654080.2*,  $F_{ST}$ : 0.36), a gene involved in the plant defense response. Zhang et al. (2011) reported a role for PP2-A1 in the activation of plant defense responses during biotic stress.

In conclusion, our functional network analysis identified drought-responsive genes in SO related to three major pathways: ABA response, osmoregulation and plant defense, all of which are known to be upregulated in drought-tolerant genotypes. These results suggest that SO has a better capacity to maintain both its primary root growth and osmoregulation, facilitating the maintenance of water metabolism during water shortage and greater tolerance to drought stress. In addition, most of the genes highlighted by our functional network presented at least one SNP above the 90% quantile of the genome-wide  $F_{ST}$  distribution, suggesting that our functional analysis yielded genes not only important for the response to stress in SO, but also probably subject to strong positive selection.

### ***Molecular mechanisms underlying intrinsic reproductive isolation barriers***

We focus here on the genes displaying a “species” effect regardless of the treatment imposed. We hypothesize that this gene set contains key molecular players potentially involved in intrinsic barriers to reproductive isolation. Indeed, our functional analysis (Figure 5) highlighted several processes (i.e. “flowering time”, “cell proliferation”, “cell differentiation”) related to this biological function. Our transcriptome data were obtained from roots rather than floral meristems, but the molecular mechanisms identified here probably reflect important functions of plant meristems generally. This hypothesis is also supported by the high level of differentiation of the genes identified in our biological network, ranging from 0.15 (F9L11.8) to 1 (BAS1). Moreover, 164 of the 184 genes identified in the whole dataset (89%) were also found in the 0.1% tail of the genome-wide  $F_{ST}$  distribution, with max  $F_{ST}$  values ranging from 0.14 to 1 (Supplementary File 8). Some of these genes may, therefore, correspond to intrinsic reproductive barrier loci.



Eight genes in the biological network shown in Figure 4 were considered to be of particular interest. Four of these genes relate to the “flowering time” process: *BAS1* (Qrob\_P0551020.2, Fst: 1), *AP3* (Qrob\_P0454040.2, Fst: 0.64), *CYP94B3* (Qrob\_P0369770.2, Fst: 0.47) and *F12M16.14* (Qrob\_P0657620.2, Fst: 0.31). Sandhu et al. (2012) showed that the *BAS1* gene encodes a cytochrome P450 protein involved in brassinosteroid catabolism and reported a key role for brassinosteroid inactivation during the floral induction of *AP3*, a key component of meristem identity in plants. *AP3* interacts with *AP1* to specify floral meristem identity during floral transition in *Arabidopsis thaliana* (Liu et al., 2007). *CYP94B3* also encodes a cytochrome P450 protein. Bruckhoff et al., (2016) suggested that this gene might play an important role in controlling flowering time by inactivating the phytohormone jasmonyl isoleucine. We also identified two other CYP proteins with high Fst values in our gene list (but not in our functional network): *CYP72A9* (Qrob\_P0084340.2, Fst: 1) and *CYP76G1* (Qrob\_P0088450.2, Fst: 0.75). These findings suggest that some members of this group may be major targets of natural selection. Finally, we identified an *F12M16.14* gene encoding a  $\text{Na}^+/\text{Ca}^{2+}$  exchanger-like protein involved in maintaining  $\text{Ca}^{2+}$  homeostasis. Li et al., (2016) reported that transgenic *Arabidopsis* lines displayed alterations to flowering time due to changes in the expression of two major flowering genes (the *Constans* and *Flowering* loci).

Three of the genes identified in our network analysis (*AREB3* (Qrob\_P0187720.2, Fst: 0.75) *TT2* (Qrob\_P0304680.2, Fst: 0.66) and *F4P9.36* (Qrob\_P0554070.2, Fst: 0.57)) belong to the “seed germination process” (Figure 5): i) *AREB3* encodes an ABA-responsive element binding protein with a bZIP domain. Wang et al. (2015) reported a key role for this gene in germination, seed development and embryo maturation. Moreover, Hoth et al. (2010) showed that *AREB3* is co-

expressed with a sugar transporter gene (*AtSUC1*) in pollen, ii) *TT2*, encoding an R2R3 Myb transcription factor involved in anthocyanin/proanthocyanin accumulation was also identified. Zhao et al. (2019) reported that *Arabidopsis* mutants with a downregulated *TT2* gene had a shorter dormancy period, suggesting an important role for this gene in meristem functioning, and iii) *F4P9.36*, encoding a protein similar to dihydroflavonol-4-reductase. Østergaard et al. (2001) showed that this protein was involved in embryogenesis and seed maturation in *Arabidopsis thaliana*.

Finally, we identified a *SOT17* gene (Qrob\_P0168490.2, Fst: 0.72) associated with cell proliferation. *SOT17* is similar to desulfoglucosinolate sulfotransferase. The *SOT17* gene has been reported to be expressed in the shoot apical meristem and is potentially involved in cell proliferation in the root meristem via the glucose TOR signaling pathway (Kim et al., (2017); Xiong & Sheen (2013)).

However, most of these 184 genes were not found in our subnetwork enrichment analysis (166 of 185). Below, we focus exclusively on those with high Fst values.

The genes with high Fst values included three genes (two *CAS1* genes (Qrob\_P0684390.2, Fst:1) and a *NOT2* gene (Qrob\_P0689750.2, Fst 0.77)) involved in male gametophyte initiation. Wang et al., (2013) reported that the downregulation of *NOT2* in *Arabidopsis* caused severe defects in the male gametophyte. We also identified two *CAS1* gene encoding a cycloartenol synthase 1 involved in sterol biosynthesis. Babiychuk et al., (2008) showed that the *CAS1* gene have a key role in the male gametophyte function and in the meristematic activity. We identified a *F14M4.3* gene (Fst: 0.92) similar to a short-chain dehydrogenase reductase 5 (SDR). Cheng et al., (2002) reported that the SDR super family is involved in both hormone biosynthesis in mammals and sex determination

in maize. Three BGLU17 genes with Fst values ranging from 0.97 to 1 (Qrob\_P0491340.2, Qrob\_P0491360.2 and Qrob\_P0597130.2) were also found in our analysis. BGLU17 encodes beta glucosidase proteins potentially involved in meristem functioning through the initiation of cell division. Brzobohaty et al. (1993) reported that beta glucosidase cleaved the cytokinin-O-glucoside conjugates, which are biologically inactive, to release active cytokinin, which is essential to initiate cell division in the meristem. A PAP2 gene (Qrob\_P0010280.2, Fst: 0.79) encoding a protein highly similar to phytochrome-associated protein 2 was also found in our gene list. Kobayashi et al. (2010) showed that the pattern of meristem initiation was disorganized in rice mutants for this gene. In the mutant transgenic lines, new meristems were unable to develop into spikelet meristems, highlighting a key role for PAP2 in the initiation of floral meristems. Finally, an F13K3.9 (Qrob\_P0012240.2, Fst: 0.89) gene encoding a 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase protein was identified. Ciannamea et al. (2006) showed that this gene was downregulated during vernalization, suggesting a possible role in floral meristem initiation.

In conclusion, our analysis highlighted several genes potentially involved in flowering time regulation, meristem functioning, embryo maturation, cell proliferation, male gametophyte initiation and floral meristem identity. All the genes discussed in this section had high Fst values, suggesting that they are potential targets of natural selection. Overall, these results provide support for a potential role of these genes in maintaining intrinsic barriers to reproduction between oak species.

## Acknowledgments

This work was supported by INRAE, the Genoscope: the *Commissariat à l’Energie Atomique et aux Energies Alternatives* (CEA) and ANR (GENOAK project 2011 BSV6 009 01). We thank the Genotoul Bioinformatics Platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing computing resources, and the Genome Transcriptome Facility of Bordeaux (grant from *Investissements d’Avenir, Convention attributive d’aide EquipEx Xyloforest* ANR-10-EQPX-16-01) for providing the infrastructure for RT-qPCR. IL received funding from INRAE and TB received funding from the ANR and the European Union's *ERC program* (TREEPEACE # FP7-339728).

## References

- Abzhanov, A., Kuo, W. P., Hartmann, C., Grant, B. R., Grant, P. R., & Tabin, C. J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin’s finches. *Nature*, 442(7102), 563–567. doi: 10.1038/nature04843
- Acosta, I. F., Gasperini, D., Chételat, A., Stolz, S., Santuari, L., & Farmer, E. E. (2013). Role of NINJA in root jasmonate signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 110(38), 15473–15478. doi: 10.1073/pnas.1307910110
- Alexa, A. (2010). Gene set enrichment analysis with topGO. *October*, 27–27.
- Babiychuk, E., Bouvier-Nave, P., Compagnon, V., Suzuki, M., Muranaka, T., Van Montagu, M., ... Schaller, H. (2008). Allelic mutant series reveal distinct functions for Arabidopsis cycloartenol synthase 1 in cell viability and plastid biogenesis. *Proceedings of the National Academy of Sciences*, 105(8), 3163–3168. doi: 10.1073/pnas.0712190105

- Becker, M., Nieminen, T. M., & G  r  mia, F. (1994). Short-term variations and long-term changes in oak productivity in northeastern France. The role of climate and atmospheric CO 2. *Annales Des Sciences Foresti  res*, 51(5), 477–492. doi: 10.1051/forest:19940504
- Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20(10), 2044–2072. doi: 10.1111/j.1365-294X.2011.05080.x
- Br  da, N., Cochard, H., Dreyer, E., & Granier, A. (1993). Field comparison of transpiration, stomatal conductance and vulnerability to cavitation of *Quercus petraea* and *Quercus robur* under water stress. *Annales Des Sciences Foresti  res*, 50(6), 571–582. doi: 10.1051/forest:19930606
- Br  da, N., Cochard, H., Dreyer, E., & Granier, A. (1993). Water transfer in a mature oak stand (*Quercus petraea*): Seasonal evolution and effects of a severe drought. *Canadian Journal of Forest Research*, 23(6), 1136–1143. doi: 10.1139/x93-144
- Br  da, Nathalie, & Badeau, V. (2008). Forest tree responses to extreme drought and some biotic events: Towards a selection according to hazard tolerance? *Comptes Rendus Geoscience*, 340(9), 651–662. doi: 10.1016/j.crte.2008.08.003
- Br  da, Nathalie, Huc, R., Granier, A., & Dreyer, E. (2006). Temperate forest trees and stands under severe drought: A review of ecophysiological responses, adaptation processes and long-term consequences. *Annals of Forest Science*, 63(6), 625–644. doi: 10.1051/forest:2006042
- Bruckhoff, V., Haroth, S., Feussner, K., K  nig, S., Brodhun, F., & Feussner, I. (2016). Functional Characterization of CYP94-Genes and Identification of a Novel Jasmonate Catabolite in Flowers. *PLOS ONE*, 11(7), e0159875. doi: 10.1371/journal.pone.0159875

- Brzobohaty, B., Moore, I., Kristoffersen, P., Bako, L., Campos, N., Schell, J., & Palme, K. (1993). Release of active cytokinin by a beta-glucosidase localized to the maize root meristem. *Science*, 262(5136), 1051–1054. doi: 10.1126/science.8235622
- Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, 124(4), 561–583. doi: 10.1093/biolinnean/bly063
- Chapman, M. A., Hiscock, S. J., & Filatov, D. A. (2013). Genomic Divergence during Speciation Driven by Adaptation to Altitude. *Molecular Biology and Evolution*, 30(12), 2553–2567. doi: 10.1093/molbev/mst168
- Cheng, W.-H., Endo, A., Zhou, L., Penney, J., Chen, H.-C., Arroyo, A., ... Sheen, J. (2002). A Unique Short-Chain Dehydrogenase/Reductase in Arabidopsis Glucose Signaling and Absciscic Acid Biosynthesis and Functions. *The Plant Cell*, 14(11), 2723–2743. doi: 10.1105/tpc.006494
- Ciannamea, S., Busscher-Lange, J., de Folter, S., Angenent, G. C., & Immink, R. G. H. (2006). Characterization of the Vernalization Response in *Lolium perenne* by a cDNA Microarray Approach. *Plant and Cell Physiology*, 47(4), 481–492. doi: 10.1093/pcp/pcj015
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Oxford, New York: Oxford University Press.
- Dupouey, J. L., & Badeau, V. (1993). Morphological variability of oaks (*Quercus robur* L, *Quercus petraea* (Matt) Liebl, *Quercus pubescens* Willd) in northeastern France: Preliminary results. *Annales Des Sciences Forestières*, 50(Supplement), 35s–40s. doi: 10.1051/forest:19930702

- Durand, P., Gelpe, J., Lemoine, B., Riom, J., & Timbal, J. (1983). Le dépérissement du C hêne pédonculé dans les Pyrénées-Atlantiques. *Revue Forestière Française*, (5), 357. doi: 10.4267/2042/21670
- Epron, D., & Dreyer, E. (1990). Stomatal and non stomatal limitation of photosynthesis by leaf water deficits in three oak species: A comparison of gas exchange and chlorophyll a fluorescence data. *Annales Des Sciences Forestières*, 47(5), 435–450. doi: 10.1051/forest:19900503
- Evans, D. E. (2003). Aerenchyma formation. *New Phytologist*, 161(1), 35–49. doi: DOI: 10.1046/j.1469-8137.2003.00907.x
- Fay, J. C., & Wittkopp, P. J. (2008). Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*, 100(2), 191–199. doi: 10.1038/sj.hdy.6801000
- Folzer, H., Dat, J. F., Capelli, N., Rieffel, D., & Badot, P.-M. (2006). Response of sessile oak seedlings (*Quercus petraea*) to flooding: An integrated study. *Tree Physiology*, 26(6), 759–766.
- Friedrichs, D. A., Büntgen, U., Frank, D. C., Esper, J., Neuwirth, B., & Löffler, J. (2009). Complex climate controls on 20th century oak growth in Central-West Germany. *Tree Physiology*, 29(1), 39–51. doi: 10.1093/treephys/tpn003
- Grunewald, W., Smet, I. D., Lewis, D. R., Löfke, C., Jansen, L., Goeminne, G., ... Beeckman, T. (2012). Transcription factor WRKY23 assists auxin distribution patterns during Arabidopsis root development through local control on flavonol biosynthesis. *Proceedings of the National Academy of Sciences*, 109(5), 1554–1559. doi: 10.1073/pnas.1121134109

- Guichoux, E., Lagache, L., Wagner, S., Léger, P., & Petit, R. J. (2011). *Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (Quercus spp.).* 11(3), 578–585. doi: DOI: 10.1111/j.1755-0998.2011.02983.x
- Hegarty, M. J., Barker, G. L., Brennan, A. C., Edwards, K. J., Abbott, R. J., & Hiscock, S. J. (2008). Changes to gene expression associated with hybrid speciation in plants: Further insights from transcriptomic studies in Senecio. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1506), 3055–3069. doi: 10.1098/rstb.2008.0080
- Heitz, T., Widemann, E., Lugan, R., Miesch, L., Ullmann, P., Désaubry, L., ... Pinot, F. (2012). Cytochromes P450 CYP94C1 and CYP94B3 Catalyze Two Successive Oxidation Steps of Plant Hormone Jasmonoyl-isoleucine for Catabolic Turnover. *Journal of Biological Chemistry*, 287(9), 6296–6306. doi: 10.1074/jbc.M111.316364
- Hoth, S., Niedermeier, M., Feuerstein, A., Hornig, J., & Sauer, N. (2010). An ABA-responsive element in the AtSUC1 promoter is involved in the regulation of AtSUC1 expression. *Planta*, 232(4), 911–923. doi: 10.1007/s00425-010-1228-4
- Hsiao, T. (n.d.). Measurement of plant water status. In *Irrigation of Agricultural Crops* (Vol. 30, pp. 243–249). Steward, B.A., Nielsen, D.R.
- Hubert, F., Grimm, G. W., Jousset, E., Berry, V., Franc, A., & Kremer, A. (2014). Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*. *Systematics and Biodiversity*, 12(4), 405–423. doi: 10.1080/14772000.2014.941037
- Jeukens, J., Renaut, S., St-Cyr, J., Nolte, A. W., & Bernatchez, L. (2010). The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing: TRANSCRIPTOMIC DIVERGENCE IN LAKE WHITEFISH. *Molecular Ecology*, 19(24), 5389–5403. doi: 10.1111/j.1365-294X.2010.04934.x



- Kaplan, F., & Guy, C. L. (2004).  $\beta$ -Amylase Induction and the Protective Role of Maltose during Temperature Shock. *Plant Physiology*, 135(3), 1674–1684. doi: 10.1104/pp.104.040808
- Kazan, K., & Manners, J. M. (2013). MYC2: The Master in Action. *Molecular Plant*, 6(3), 686–703. doi: 10.1093/mp/sss128
- Kim, H.-J., Wu, C.-Y., Yu, H.-M., Sheen, J., & Lee, H. (2017). Dual CLAVATA3 peptides in Arabidopsis shoot stem cell signaling. *Journal of Plant Biology*, 60(5), 506–512. doi: 10.1007/s12374-017-0083-2
- Kobayashi, K., Maekawa, M., Miyao, A., Hirochika, H., & Kyoizuka, J. (2010). PANICLE PHYTOMER2 (PAP2), encoding a SEPALLATA subfamily MADS-box protein, positively controls spikelet meristem identity in rice. *Plant and Cell Physiology*, 51(1), 47–57. doi: 10.1093/pcp/pcp166
- Kofler, R., Orozco-terWengel, P., Maio, N. D., Pandey, R. V., Nolte, V., Futschik, A., ... Schlötterer, C. (2011). PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLOS ONE*, 6(1), e15925. doi: 10.1371/journal.pone.0015925
- Kremer, A., Dupouey, J. L., Deans, J. D., Cottrell, J., Csaikl, U., Finkeldey, R., ... Badeau, V. (2002). Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science*, 59(7), 777–787. doi: 10.1051/forest:2002065
- Kremer, A., & Hipp, A. L. (2020). Oaks: An evolutionary success story. *New Phytologist*, 226(4), 987–1011. doi: 10.1111/nph.16274
- Lai, Z., Gross, B. L., Zou, Y., Andrews, J., & Rieseberg, L. H. (2006). Microarray analysis reveals differential gene expression in hybrid sunflower species. *Molecular Ecology*, 15(5), 1213–1227. doi: 10.1111/j.1365-294X.2006.02775.x

- Le Provost, G., Herrera, R., Paiva, J. A. P., Chaumeil, P., Salin, F., & Plomion, C. (2007). A micromethod for high throughput RNA extraction in forest trees. *Biological Research*, 40(3), 291–297.
- Le Provost, G., Lesur, I., Lalanne, C., Da Silva, C., Labadie, K., Aury, J. M., ... Plomion, C. (2016). Implication of the suberin pathway in adaptation to waterlogging and hypertrophied lenticels formation in pedunculate oak ( *Quercus robur* L.). *Tree Physiology*, tpw056. doi: 10.1093/treephys/tpw056
- Le Provost, G., Sulmon, C., Frigerio, J. M., Bodenes, C., Kremer, A., & Plomion, C. (2012). Role of waterlogging-responsive genes in shaping interspecific differentiation between two sympatric oak species. *Tree Physiology*, 32(2), 119–134. doi: 10.1093/treephys/tp123
- Lebourgeois, F., Drénou, C., Bouvier, M., & Lemaire, J. (2015). Caractérisation de la croissance des chênaies pédonculées atlantiques dépérissantes: Effets des sécheresses et relation avec l'architecture des houppiers. *Revue Forestière Française*, (4). doi: 10.4267/2042/59289
- LEBOURGEOIS (François). (2006). Sensibilité au climat des Chênes sessile et pédonculé dans le réseau RENECOFOR. Comparaison avec les hêtraies. *Revue Forestière Française*, (1). doi: 10.4267/2042/5720
- Lepais, O., & Gerber, S. (2011). Reproductive Patterns Shape Introgression Dynamics and Species Succession Within the European White Oak Species Complex. *Evolution*, 65(1), 156–170. doi: 10.1111/j.1558-5646.2010.01101.x
- Leroy, T., Louvet, J.-M., Lalanne, C., Provost, G. L., Labadie, K., Aury, J.-M., ... Kremer, A. (2020). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytologist*, 226(4), 1171–1182. doi: <https://doi.org/10.1111/nph.16095>

- Leroy, T., Rougemont, Q., Dupouey, J.-L., Bodenes, C., Lalanne, C., Belser, C., ... Plomion, C. (2018). Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *BioRxiv*, 246637. doi: 10.1101/246637
- Leroy Thibault, Roux Camille, Villate Laure, Bodénès Catherine, Romiguier Jonathan, Paiva Jorge A. P., ... Kremer Antoine. (2017). Extensive recent secondary contacts between four European white oak species. *New Phytologist*, 214(2), 865–878. doi: 10.1111/nph.14413
- Lexer, C., Heinze, B., Alia, R., & Rieseberg, L. H. (2004). Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: Lessons from wild annual sunflowers (*Helianthus* spp.). *Forest Ecology and Management*, 197(1), 49–64. doi: 10.1016/j.foreco.2004.05.004
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, P., Zhang, G., Gonzales, N., Guo, Y., Hu, H., Park, S., & Zhao, J. (2016). Ca<sup>2+</sup>-regulated and diurnal rhythm-regulated Na<sup>+</sup>/Ca<sup>2+</sup> exchanger AtNCL affects flowering time and auxin signalling in *Arabidopsis*. *Plant, Cell & Environment*, 39(2), 377–392. doi: 10.1111/pce.12620
- Liu, C., Zhou, J., Bracha-Drori, K., Yalovsky, S., Ito, T., & Yu, H. (2007). Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes. *Development*, 134(10), 1901–1910. doi: 10.1242/dev.003103
- Liu, Y., & He, C. (2017). A review of redox signaling and the control of MAP kinase pathway in plants. *Redox Biology*, 11, 192–204. doi: 10.1016/j.redox.2016.12.009

- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12). doi: 10.1186/s13059-014-0550-8
- Manos, P., & Hipp, A. (2021). An Updated Infrageneric Classification of the North American Oaks (*Quercus* Subgenus *Quercus*): Review of the Contribution of Phylogenomic Data to Biogeography and Species Diversity. *Forests*, 12, 786. doi: 10.3390/f12060786
- Marguerit, E., Brendel, O., Lebon, E., Van Leeuwen, C., & Ollat, N. (2012). Rootstock control of scion transpiration and its acclimation to water deficit are controlled by different genes. *New Phytologist*, 194(2), 416–429. doi: 10.1111/j.1469-8137.2012.04059.x
- Martínez-Fernández, M., Bernatchez, L., Rolán-Alvarez, E., & Quesada, H. (2010). Insights into the role of differential gene expression on the ecological adaptation of the snail *Littorina saxatilis*. *BMC Evolutionary Biology*, 10(1), 356. doi: 10.1186/1471-2148-10-356
- Montiel, J., Arthikala, M.-K., & Quinto, C. (2013). Phaseolus vulgaris RbohB functions in lateral root development. *Plant Signaling & Behavior*, 8(1), e22694. doi: 10.4161/psb.22694
- Østergaard, L., Lauvergeat, V., Næsted, H., Mattsson, O., & Mundy, J. (2001). Two differentially regulated Arabidopsis genes define a new branch of the DFR superfamily. *Plant Science*, 160(3), 463–472. doi: 10.1016/S0168-9452(00)00407-6
- Parelle, J., Brendel, O., Bodénès, C., Berveiller, D., Dizengremel, P., Jolivet, Y., & Dreyer, E. (2006). Differences in morphological and physiological responses to water-logging between two sympatric oak species (*Quercus petraea* [Matt.] Liebl., *Quercus robur* L.). *Annals of Forest Science*, 63(8), 11. doi: 10.1051/forest:2006068
- Parelle, J., Brendel, O., Jolivet, Y., & Dreyer, E. (2007). Intra- and interspecific diversity in the response to waterlogging of two co-occurring white oak species (*Quercus robur* and *Q. petraea*). *Tree Physiology*, 27(7), 1027–1034. doi: 10.1093/treephys/27.7.1027

- Parent, C., Berger, A., Folzer, H., Dat, J., Crevècoeur, M., Badot, P.-M., & Capelli, N. (2008). A novel nonsymbiotic hemoglobin from oak: Cellular and tissue specificity of gene expression. *New Phytologist*, 177(0), 142–154. doi: 10.1111/j.1469-8137.2007.02250.x
- Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., ... Salse, J. (2018). Oak genome reveals facets of long lifespan. *Nature Plants*, 1. doi: 10.1038/s41477-018-0172-3
- Ponton, S., Dupouey, J.-L., Bréda, N., & Dreyer, E. (2002). Comparison of water-use efficiency of seedlings from two sympatric oak species: Genotype x environment interactions. *Tree Physiology*, 22(6), 413–422.
- Porth, I., Koch, M., Berenyi, M., Burg, A., & Burg, K. (2005). Identification of adaptation-specific differences in mRNA expression of sessile and pedunculate oak based on osmotic-stress-induced genes. *Tree Physiology*, 25(10), 1317–1329. doi: 10.1093/treephys/25.10.1317
- Ranocha, P., Dima, O., Nagy, R., Felten, J., Corratgé-Faillie, C., Novák, O., ... Goffner, D. (2013). *Arabidopsis* WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis. *Nature Communications*, 4, 2625. doi: 10.1038/ncomms3625
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. doi: 10.1111/jeb.13047
- Rieseberg, L. H., Wood, T. E., & Baack, E. J. (2006). The nature of plant species. *Nature*, 440(7083), 524–527. doi: 10.1038/nature04402
- Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology (Clifton, N.J.)*, 132, 365–386.

- Sandhu, K. S., Hagely, K., & Neff, M. M. (2012). Genetic Interactions Between Brassinosteroid-Inactivating P450s and Photomorphogenic Photoreceptors in *Arabidopsis thaliana*. *G3 &#58; Genes|Genomes|Genetics*, 2(12), 1585–1593. doi: 10.1534/g3.112.004580
- Schröder, F., Lisso, J., Lange, P., & Müssig, C. (2009). The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. *BMC Plant Biology*, 9, 20. doi: 10.1186/1471-2229-9-20
- Shamir, R., Maron-Katz, A., Tanay, A., Linhart, C., Steinfeld, I., Sharan, R., ... Elkon, R. (2005). EXPANDER--an integrative program suite for microarray data analysis. *BMC Bioinformatics*, 6, 232. doi: 10.1186/1471-2105-6-232
- Shi, L., Bielawski, J., Mu, J., Dong, H., Teng, C., Zhang, J., ... Zuo, J. (2007). Involvement of sphingoid bases in mediating reactive oxygen intermediate production and programmed cell death in *Arabidopsis*. *Cell Research*, 17(12), 1030–1040. doi: 10.1038/cr.2007.100
- Smith, S. J., Kroon, J. T. M., Simon, W. J., Slabas, A. R., & Chivasa, S. (2015). A Novel Function for Arabidopsis CYCLASE1 in Programmed Cell Death Revealed by Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) Analysis of Extracellular Matrix Proteins. *Molecular & Cellular Proteomics*, 14(6), 1556–1568. doi: 10.1074/mcp.M114.045054
- Spieß, N., Oufir, M., Matušíková, I., Stierschneider, M., Kopecky, D., Homolka, A., ... Wilhelm, E. (2012). Ecophysiological and transcriptomic responses of oak (*Quercus robur*) to long-term drought exposure and rewatering. *Environmental and Experimental Botany*, 77, 117–126. doi: 10.1016/j.envexpbot.2011.11.010
- Storz, J. F. (n.d.). INVITED REVIEW: Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, 14(3), 671–688. doi: 10.1111/j.1365-294X.2005.02437.x

- Thomas, F. M., & Hartmann, G. (1996). Soil and tree water relations in mature oak stands of northern Germany differing in the degree of decline. *Annales Des Sciences Forestières*, 53(2–3), 697–720. doi: 10.1051/forest:19960247
- Truffaut, L., Chancerel, E., Ducousso, A., Dupouey, J. L., Badeau, V., Ehrenmann, F., & Kremer, A. (2017). Fine-scale species distribution changes in a mixed oak stand over two successive generations. *New Phytologist*, 215(1), 126–139. doi: 10.1111/nph.14561
- van der Werf, G. W., Sass-Klaassen, U. G. W., & Mohren, G. M. J. (2007). The impact of the 2003 summer drought on the intra-annual growth pattern of beech (*Fagus sylvatica* L.) and oak (*Quercus robur* L.) on a dry site in the Netherlands. *Dendrochronologia*, 25(2), 103–112. doi: 10.1016/j.dendro.2007.03.004
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3(7), RESEARCH0034.
- Vidal, E. A., Araus, V., Lu, C., Parry, G., Green, P. J., Coruzzi, G. M., & Gutiérrez, R. A. (2010). Nitrate-responsive miR393/AFB3 regulatory module controls root system architecture in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 107(9), 4477–4482. doi: 10.1073/pnas.0909571107
- Wang, L., Song, X., Gu, L., Li, X., Cao, S., Chu, C., ... Cao, X. (2013). NOT2 Proteins Promote Polymerase II–Dependent Transcription and Interact with Multiple MicroRNA Biogenesis Factors in *Arabidopsis*. *The Plant Cell*, 25(2), 715–727. doi: 10.1105/tpc.112.105882
- Wang, Z., Cheng, K., Wan, L., Yan, L., Jiang, H., Liu, S., ... Liao, B. (2015). Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes. *BMC Genomics*, 16(1), 1053. doi: 10.1186/s12864-015-2258-x

- Wolf, J. B. W., Bayer, T., Haubold, B., Schilhabel, M., Rosenstiel, P., & Tautz, D. (2010). Nucleotide divergence vs. gene expression differentiation: Comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology*, *19*, 162–175. doi: 10.1111/j.1365-294X.2009.04471.x
- Wu, K., Zhang, L., Zhou, C., Yu, C.-W., & Chaikam, V. (2008). HDA6 is required for jasmonate response, senescence and flowering in Arabidopsis. *Journal of Experimental Botany*, *59*(2), 225–234. doi: 10.1093/jxb/erm300
- Xiong, Y., & Sheen, J. (2013). Moving beyond translation: Glucose-TOR signaling in the transcriptional control of cell cycle. *Cell Cycle*, *12*(13), 1989–1990. doi: 10.4161/cc.25308
- Xu, G., Wang, S., Han, S., Xie, K., Wang, Y., Li, J., & Liu, Y. (2017). Plant Bax Inhibitor-1 interacts with ATG6 to regulate autophagy and programmed cell death. *Autophagy*, *13*(7), 1161–1175. doi: 10.1080/15548627.2017.1320633
- Yamauchi, T., Shimamura, S., Nakazono, M., & Mochizuki, T. (2013). Aerenchyma formation in crop species: A review. *Field Crops Research*, *152*, 8–16. doi: 10.1016/j.fcr.2012.12.008
- Zhang, C., Shi, H., Chen, L., Wang, X., Lü, B., Zhang, S., ... Dong, H. (2011). Harpin-induced expression and transgenic overexpression of the phloem protein gene AtPP2-A1 in Arabidopsis repress phloem feeding of the green peach aphid *Myzus persicae*. *BMC Plant Biology*, *11*, 11. doi: 10.1186/1471-2229-11-11
- Zhang, X., Liu, X., Zhang, D., Tang, H., Sun, B., Li, C., ... Li, Y. (2017). Genome-wide identification of gene expression in contrasting maize inbred lines under field drought conditions reveals the significance of transcription factors in drought tolerance. *PLOS ONE*, *12*(7), e0179477. doi: 10.1371/journal.pone.0179477



Zhao, P., Li, X., Jia, J., Yuan, G., Chen, S., Qi, D., ... Liu, G. (2019). *BHLH92* from sheepgrass acts as a negative regulator of anthocyanin/proanthocyanidin accumulation and influences seed dormancy. *Journal of Experimental Botany*, 70(1), 269–284. doi: 10.1093/jxb/ery335

**Data accessibility:** All sequences generated in this study were deposited in the Short Read archive of NCBI under PRJEB17875 (Waterlogging experiment) and PRJED19536 (Drought stress experiment) accession numbers.

## Author contributions

GLP and CP designed the study. GLP, IL, BB and CP wrote the manuscript. BB and TL were involved in Fst enrichment analysis. IL performed bioinformatics analysis with GLP. GLP and DP were involved the RT-qPCR experiments. cDNA library construction and sequencing were performed by KL and JMA. All the authors read and approved the manuscript.

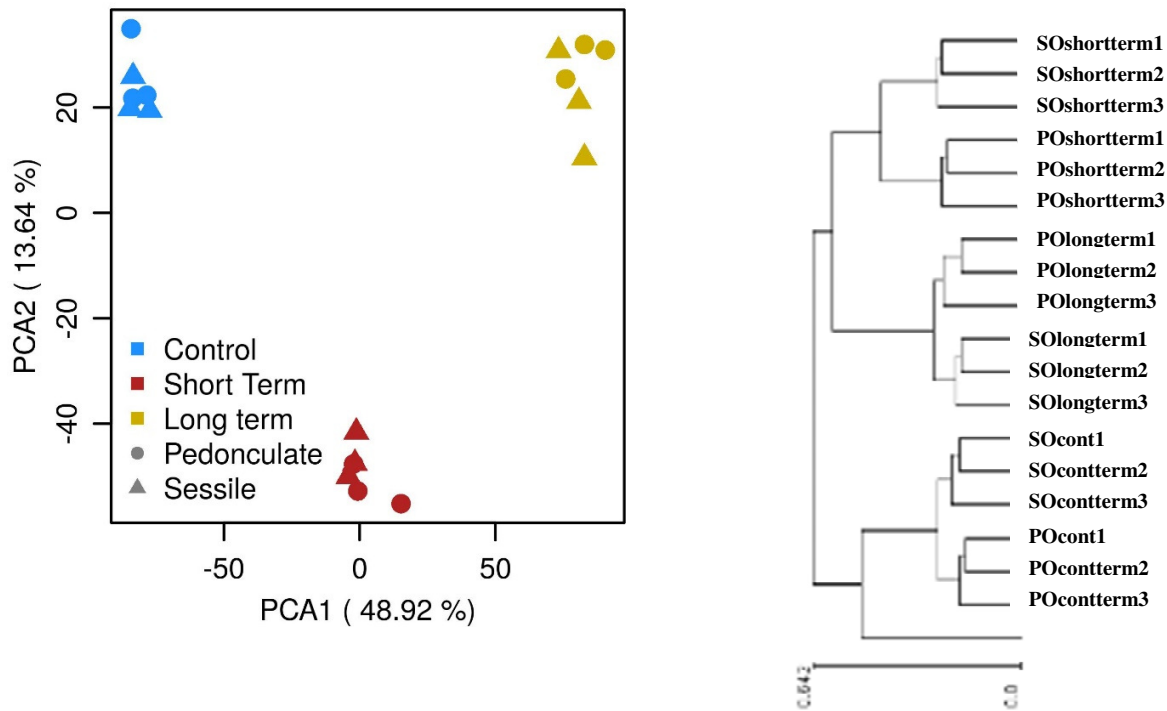
## Tables

Table 1: Summary of the DESeq2 analysis for the two experiments (waterlogging and drought). Abbreviations: PO: pedunculate oak, SO: sessile oak. The number of genes in each cell was determined with the following parameters: adjusted  $P$  value  $< 0.01$  and fold change  $\geq 2$ . For each effect, the compared models  $M_i$  (see materials and method section) are indicated in parentheses.

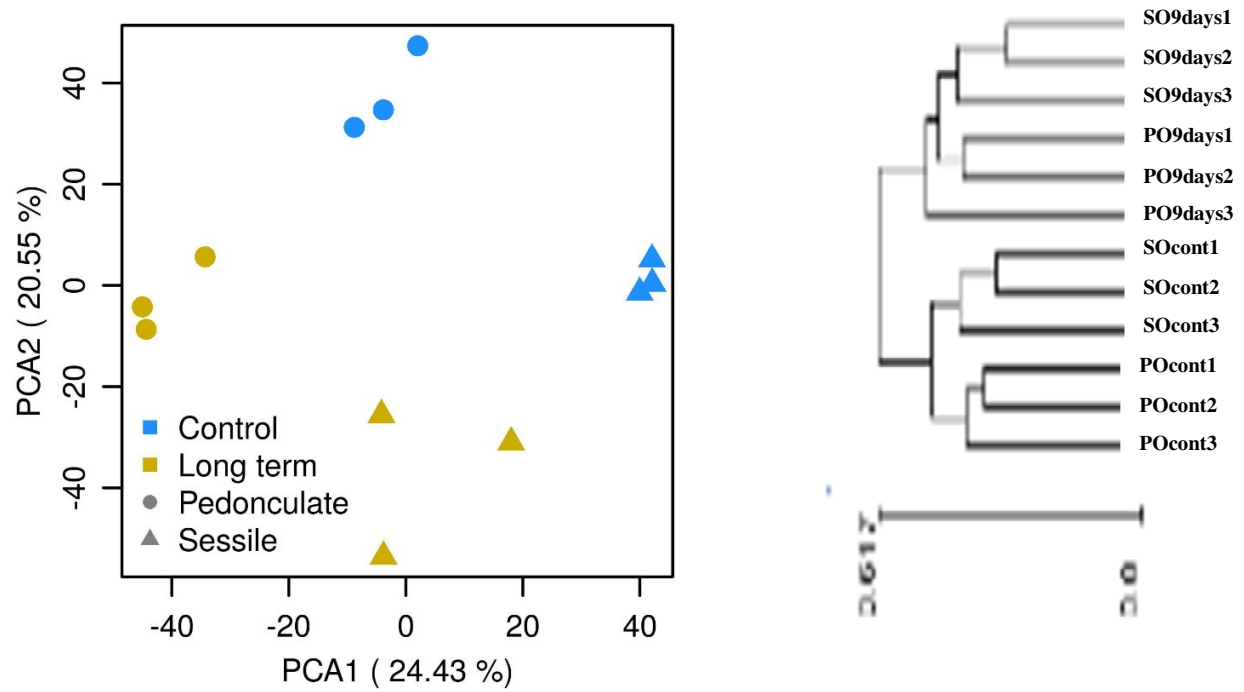
	WATERLOGGING		
	Treatment (M1-M2)	Species (M1-M3)	Interaction (M4)
PO	466	682	NA
SO	354	485	NA
<b>TOTAL</b>	<b>820</b>	<b>1,167</b>	<b>319</b>
	Drought		
	Treatment (M1-M2)	Species (M1-M3)	Interaction (M4)
PO	<b>62</b>	<b>179</b>	<b>NA</b>
SO	<b>60</b>	<b>216</b>	<b>NA</b>
<b>TOTAL</b>	<b>122</b>	<b>395</b>	<b>17</b>

## Figures

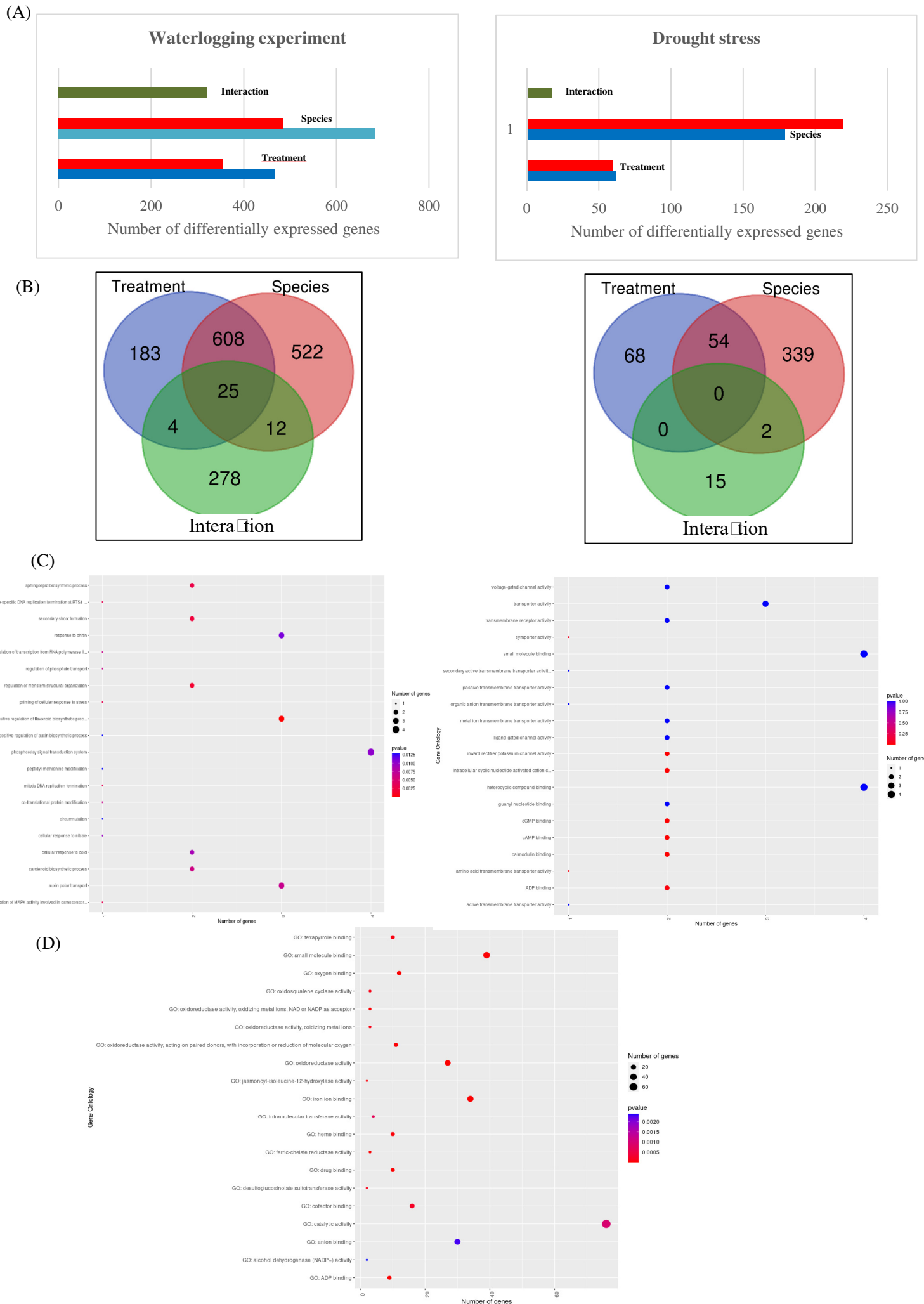
(A)



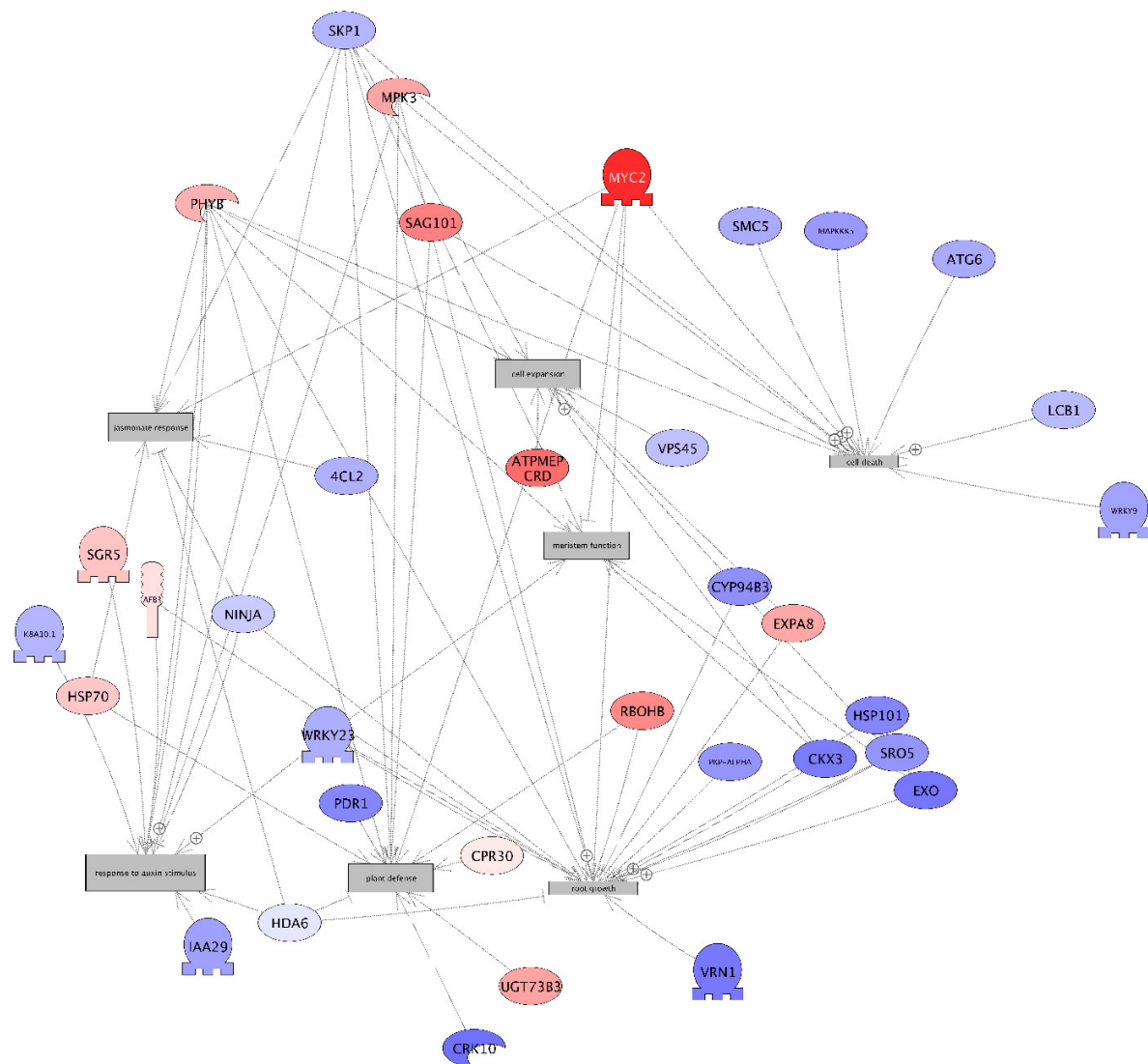
(B)



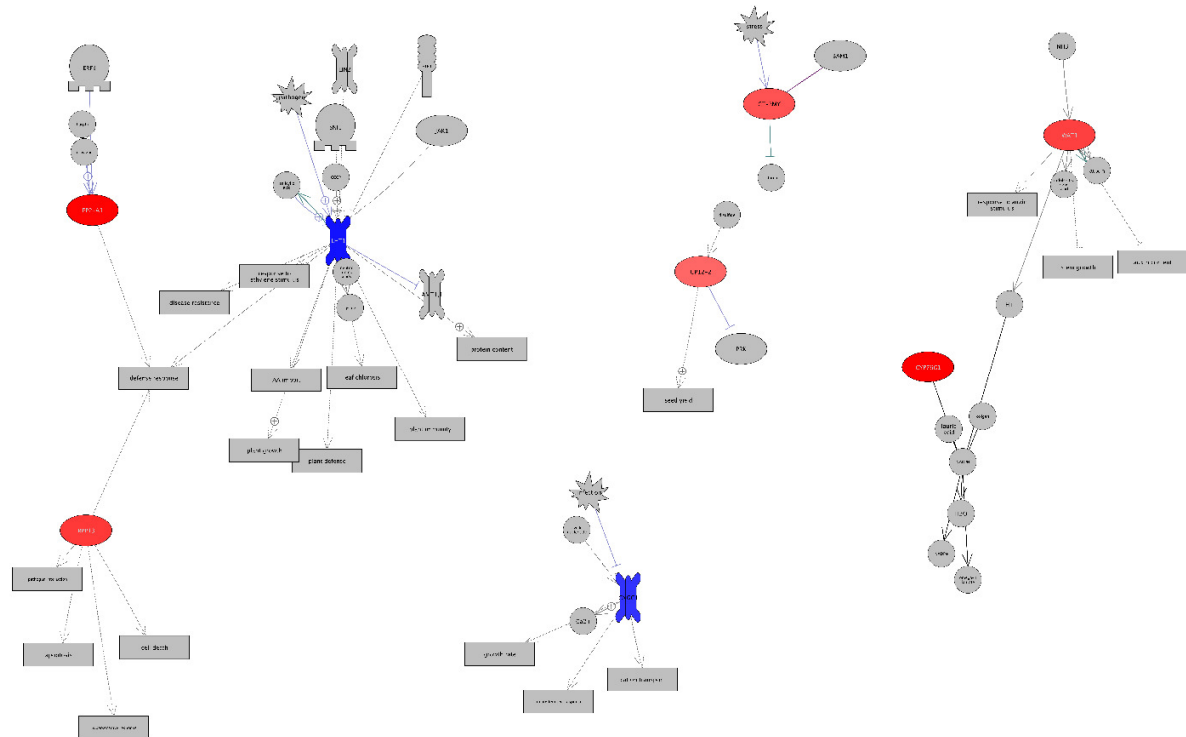
**Figure 1:** PCA and hierarchical clustering analysis.



**Figure 2:** Illustration of the main results obtained from the DEGs analysis. Panel A: Bar charts illustrating the number of DEGS obtained for each effect both for waterlogging (left section) and drought stress (right section). Panel B: Venn diagram of the genes showing a significant treatment, species or interaction effect. Left panel: waterlogging experiment, Right panel: Drought stress experiment. Panel C: Most enriched ontologies (BP) obtained for genes displaying a significant treatment-by-species interaction both for waterlogging (left part) and drought stress (right part). Panel (D) Most enriched ontologies (BP) obtained for genes displaying a significant species whatever the stress analyzed.

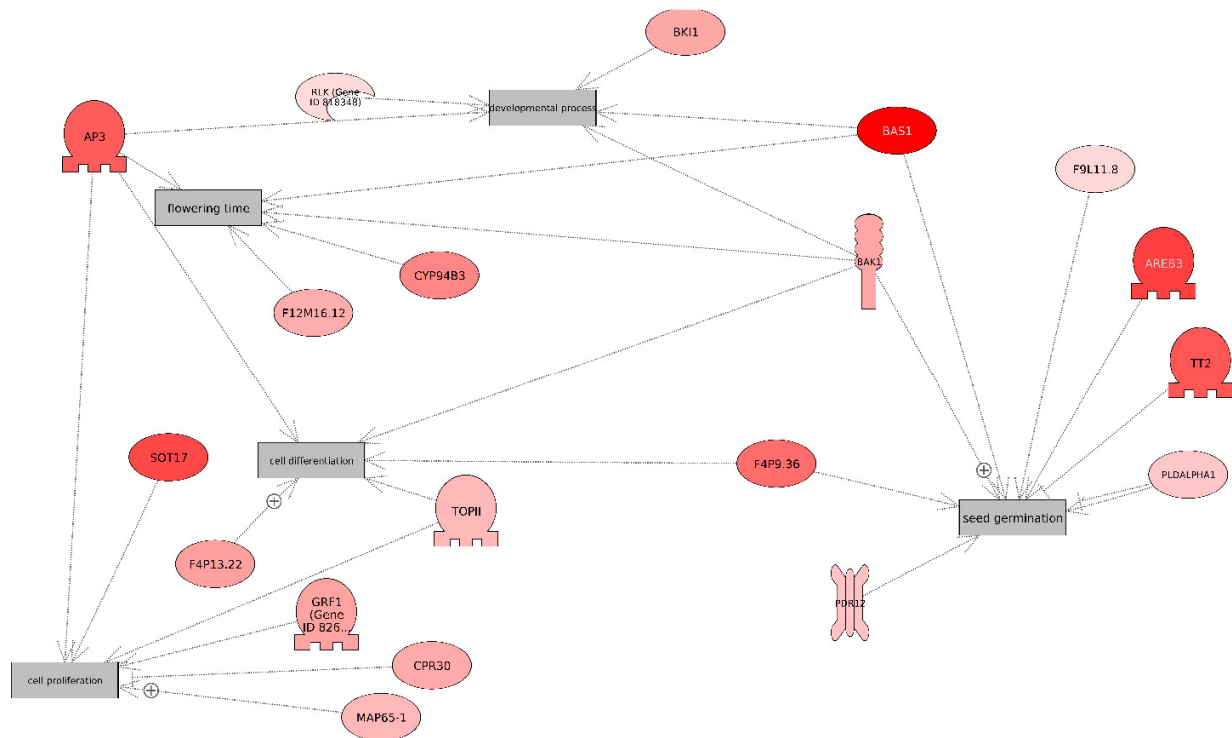


**Figure 3:** Functional network for genes displayed a significant treatment-by-interaction effect in the waterlogging experiment.

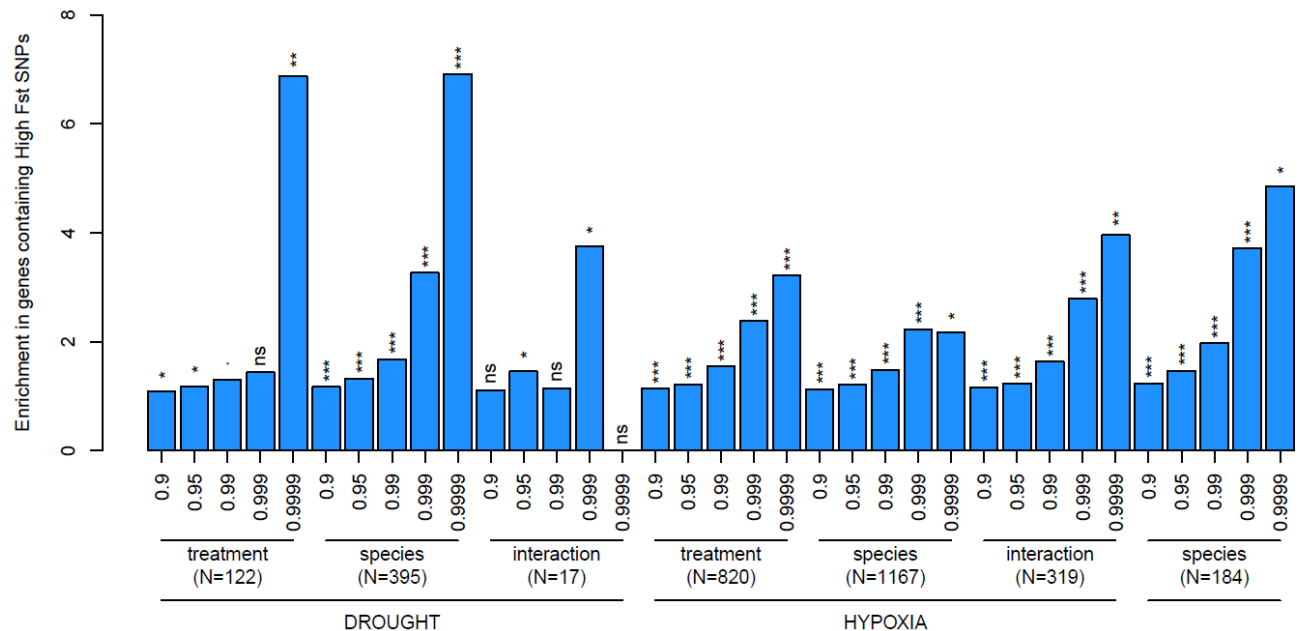


**Figure 4:** Functional network for genes displayed a significant treatment-by-interaction effect in the Drought stress experiment.





**Figure 5:** Functional network for genes displayed a significant species effect whatever the stress applied.



**Figure 6:** Genes differentially expressed between treatments and species in the drought and water-logging experiments are enriched in highly differentiated single nucleotide polymorphisms (SNPs) between *Q. robur* and *Q. petraea*. SNPs differentiation between the two species *Q. robur* and *Q. petraea* were obtained from Leroy et al., (2019). The y-axis represents the enrichment of the seven differentially expressed genes sets investigated in this study (x-axis). For example, an enrichment of above six for genes differentially expressed between treatments in the Drought experiments indicates that this gene set included over six times more genes with highly differentiated SNPs than expected by chance. For each genes set, the five bars correspond to enrichment for genes including SNPs with FST values above the 0.9, 0.95, 0.99, 0.999, 0.9999 quantile of the genome-wide FST distribution, respectively. Significance of enrichment estimates were assessed using 1000 permutations and are indicated above each bars: ns: non-significant, .: p-value < 0.1, \*: p-value < 0.05, \*\*: p-value < 0.01, \*\*\*: p-value < 0.001.

## Supporting Information

**Supplemental Information 1:** Validation of RNA-seq results by qPCR.

**Supplemental Information 2:** “Species”-responsive genes in the waterlogging and drought stress experiments.

**Supplementary Tables 1:** Overview of the cDNA libraries constructed in this study.

**Supplementary Table 2:** List of the primer pairs used for qPCR analysis.

**Supplementary Figures 1:** Schematic representation of the experimental design used in this study.

**Supplementary Figure 2** Ecophysiological characterization of the experimental design. Evolution of the O<sub>2</sub> concentration, water content in the substrate and water potential during waterlogging or drought stress.

**Supplementary Files 1, 2, 3, 4, 5, 6, 7, 8** are available online Inrae dataverse: G. Le Provost, B. Brachi, I. Lesur, C. Lalanne, K. Labadie, JM Aury, C Da Silva, D Postolache, T Leroy, C Plomion. Water stress-associated isolation barriers between two sympatric oak species , <https://doi.org/10.15454/SONFVF>, Portail Data INRAE, V3.0. These files includes normalized values for RNAseq data, Fold change Ratio and Gene set enrichment analysis. Supplementary file 8 include Fst value for the differentially expressed genes and their associated effect.