De novo transcriptome assembly and analysis of differential gene expression in response to drought in European beech

Markus Müller, Sarah Seifert, Torben Lübbe, Christoph Leuschner, Reiner Finkeldey

1 Forest Genetics and Forest Tree Breeding, Faculty for Forest Sciences and Forest Ecology, University of Goettingen, Goettingen, Lower-Saxony, Germany, 2 Plant Ecology and Ecosystems Research, Albrecht von Haller Institute for Plant Sciences, University of Goettingen, Goettingen, Lower-Saxony, Germany, 3 University of Kassel, Kassel, Hesse, Germany

* mmuellef@gwdg.de

Abstract

Despite the ecological and economic importance of European beech (Fagus sylvatica L.) genomic resources of this species are still limited. This hampers an understanding of the molecular basis of adaptation to stress. Since beech will most likely be threatened by the consequences of climate change, an understanding of adaptive processes to climate change-related drought stress is of major importance. Here, we used RNA-seq to provide the first drought stress-related transcriptome of beech. In a drought stress trial with beech saplings, 50 samples were taken for RNA extraction at five points in time during a soil desiccation experiment. De novo transcriptome assembly and analysis of differential gene expression revealed 44,335 contigs, and 662 differentially expressed genes between the stress and normally watered control group. Gene expression was specific to the different time points, and only five genes were significantly differentially expressed between the stress and control group on all five sampling days. GO term enrichment showed that mostly genes involved in lipid- and homeostasis-related processes were upregulated, whereas genes involved in oxidative stress response were downregulated in the stressed seedlings. This study gives first insights into the genomic drought stress response of European beech, and provides new genetic resources for adaptation research in this species.

Introduction

The climate of Central Europe has significantly warmed during the past 40 years and is expected to continue to do so in the decades to come [1]. Severe and recurrent droughts have been identified as a major threat to the vitality and productivity of European forests, including Central European beech forests, which mainly occur in a humid sub-oceanic climate [2, 3]. European beech (Fagus sylvatica L.) is considered to be more drought sensitive than most other deciduous tree species of the region [4–7]. Thus, this species may face growth reductions and perhaps increased mortality under a warmer and summer-drier climate in various regions.
of Central Europe [8, 9], and the investigation of its adaptation potential to changing environmental conditions and the mechanism of drought tolerance are of great importance. Several studies revealed physiological and morphological differences among beech populations of different origin for drought related traits [10–17]. This suggests that beech trees may differ with respect to drought adaptation, which offers a potential for selecting genotypes with better performance under climate warming. Nevertheless, knowledge about the molecular basis of drought stress tolerance is still scarce for this species. Existing studies about the genetic basis of drought stress tolerance in beech studied on only a few potential drought stress genes. Despite its relatively small genome size (ca. 540 Mbp; [18]) and its economic and ecological importance, genomic resources of beech are in fact still very limited [19]. The *F. sylvatica* genome has not been sequenced yet, and to date, there is only one transcriptome (related to dormancy regulation; [19]) available for this species. Therefore, more genomic resources are needed to understand the molecular basis of adaptation.

Here, we report the first drought stress-related transcriptome of beech. A drought stress experiment with saplings under controlled conditions was conducted, and samples from five stressed plants and five well-watered control plants were taken at five different points in time over the course of the experiment for sequencing. This study gives first insights into the genomic drought stress response of beech. Additionally, new genomic resources for beech including new candidate genes for drought stress tolerance are reported, which can be used in further studies.

### Materials and methods

**Drought stress experiment, sample collection, and genotyping with microsatellite markers**

The beech plants used in this study were part of a larger drought stress experiment as described in [20]. Briefly, 1- to 2-year old saplings were obtained from a nursery close to Göttingen (Germany), and cultivated under uniform conditions for 16 months. Five saplings each were grown in pots with a diameter of 0.58 m (0.05 m³ volume) with equal distances among plants. The pots were placed outdoors under a rain shelter made of transparent plexiglass allowing to control water supply, while the microclimate was close to natural conditions. The experiment consisted of a moist and a dry treatment. Drought was applied in the period July to September 2011, and May to August 2012. By regular irrigation (every 3 to 5 days), the volumetric soil water content (SWC) of the pots was kept more or less constant during the experiment, i.e. fluctuated moderately below target values of maximal SWC of ca. 21% (95% of field capacity in the soil) in the moist treatment, and ca. 12% (57% of field capacity) in the drought treatment. In total, 10 plants (5 plants of the drought stress group, and 5 plants of the control group) were used for the present study. Stomatal conductance (*Gₛ*, mmol m⁻² s⁻¹) was measured in 2012 on June 28, July 5, July 12, July 19, and July 26 around noon (ca. 11 a.m. to 2 p.m.) to infer the intensity of drought stress. The measurements were conducted with a porometer (Delta-T Devices, Cambridge, UK) on each two fully developed leaves per plant. The leaves were tagged and repeatedly measured on the five selected sampling days in the year 2012 (June 28, July 5, July 12, July 19, and July 26), always on the last day before the next irrigation event. Statistically significant differences between the drought and control group were identified with the non-parametric Mann-Whitney U-test implemented in STATISTICA 12.5 (StatSoft, Tulsa, USA). Two leaves per tree were sampled at every sampling day, immediately frozen in liquid nitrogen, and stored at -60°C until RNA extraction.

Since the saplings of the drought stress experiment were obtained from a nursery, SSR genotyping was used to infer the neutral genetic structure among the ten selected individuals for
RNA-seq. Total DNA was extracted from leaves not used for RNA extraction with a DNeasy 96 Plant Kit (QIAGEN, Hilden, Germany). The amount and quality of DNA were analyzed by 1% agarose gel electrophoreses with 1 X TAE as running buffer. SSR genotyping was conducted as described in [21]. Briefly, nine highly polymorphic SSR markers including three EST microsatellite markers were used [22–25]. Primers were labeled with fluorescent dyes and pooled into three different sets for multiplexing. After PCR, the SSR fragments were separated on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, USA), and scored using GeneMapper 4.0 (Applied Biosystems, Foster City, USA). Genetic diversity indices observed heterozygosity ($H_0$), expected heterozygosity ($H_e$), and fixation index ($F$) were calculated with GenAlEx 6.5 [26, 27]. Statistic differences between the drought and control group for the genetic diversity indices were tested using the Mann-Whitney U-test implemented in STATISTICA 12.5 (StatSoft, Tulsa, USA).

Sample preparation and RNA sequencing

In total, 50 samples (10 plants on five sampling days) were used for RNA extraction. Total RNA was extracted using the RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). Extracted RNA was sent to Chronix Biomedical GmbH (Göttingen, Germany) for library preparation and sequencing. RNA quality and integrity were evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). cDNA library preparations were conducted using the NEBNext Ultra RNA Library Prep Kit for Illumina (New England BioLabs, Frankfurt am Main, Germany). Additionally, a normalized composite sample was created to enhance the de novo transcriptome assembly. For that, extracted RNA of all 50 samples was pooled to a single composite sample. The cDNA library was prepared using the Mint-2 cDNA synthesis kit (Evrogen, Moscow, Russia), and normalized using the Trimmer-2 cDNA normalization kit (Evrogen, Moscow, Russia). The 51 libraries (50 single samples from the five time points plus one composite sample) were paired-end sequenced on five lanes on an Illumina HiSeq2000 platform (Illumina, San Diego, CA, USA). Each library was uniquely tagged with a barcode to allow library pooling for sequencing, and control and treatment samples were always sequenced together on one lane.

De novo transcriptome assembly, read mapping and sequence annotation

The de novo assembly was conducted with the CLC Genomics workbench 7.04 (CLC bio, Aarhus, Denmark) based on a de Bruijn graphs approach. Sequencing adapters were removed and the sequence reads were further trimmed for quality and ambiguity. The de novo assembly was conducted based on the composite sample using default parameters in the CLC Genomics workbench. The program cd-hit-est [28] with a sequence identity threshold of 0.95 was used to reduce the redundancy of the assembly. Finally, reads of the 50 samples of the different sampling days were mapped to the newly created assembly. Sequence annotation was carried out with the software BLAST2GO [29]. For that, contigs were searched against the NCBI non-redundant (nr) protein database using BLASTX [30] with an E-value cut-off of $1 \times 10^{-3}$. Based on these results, gene ontology (GO) terms [31] were assigned to the sequences. GO-slim mapping against the plant slim file (The Arabidopsis Information Resource (TAIR), http://www.arabidopsis.org) using BLAST2GO [32] was conducted to give an overview of the GO term distribution over the entire transcriptome.
Identification of differential gene expression

For the identification of differentially expressed genes (DEGs) between the stress and control group, two different methods were used: edgeR 3.4.0 [33] implemented in the CLC Genomics Workbench, and the R/Bioconductor package DESeq2 1.12.4. [34]. For both methods, genes with a FDR (false discovery rate) < 0.1 [35] were considered to be differentially expressed. Analyzes were carried out for each of the five time points separately. A GO term enrichment analysis was performed to identify functional categories enriched in DEGs between the stress and control group. For that, the software BLAST2GO [29] with Fisher’s exact test was used. A FDR [35] threshold of 0.05 was applied.

Quantitative real-time PCR

For the confirmation of differential gene expression revealed by RNA-seq, quantitative real-time PCR was used. In total, 12 samples (six stressed plants and six control plants) from the fifth sampling day were used for qPCR validation. The samples included eight plants, which were also used for RNA-seq and four additional plants of the drought stress experiment (two plants of the stress group and two plants of the control group), which were not included in the RNA-seq analysis. Total RNA was extracted as described for the RNA-seq experiment, and 500 ng RNA was used for cDNA synthesis using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA, USA) using Oligo(dT)_{20} primer. Genes for validation were selected based on their fold change and biological function. Gene specific primers were designed using Primer-BLAST [36] (S1 File). Actin was used as a reference gene and primers were obtained from [37]. Identity of the target sequences was confirmed by sequencing of PCR-products. RT-PCR was performed in a TOptical 96 Thermocycler (Biometra, Göttingen, Germany) with three technical replicates for each sample. Each well included 4 μL HPCL-grade H₂O, 10 μL innuMIX qPCR MasterMix SyGreen (Analytik Jena, Jena, Germany), 2.5 μL of forward and reverse primers (5 pmol), and 1 μL diluted cDNA (1:10). The PCR program comprised the following steps: pre-incubation for 3 min at 95°C, 45 cycles of amplification for 5 s at 95°C, 5 s at 58°C, and 15 s at 72°C. Relative gene expression was calculated with the software GenEx 6.1 (MultiD Analyses AB, Göteborg, Sweden). Primer efficiencies were determined by dilution series for the analyzed genes.

Results

Drought stress experiment and SSR genotyping

Stomatal conductance ($G_s$) measured around noon of the plants selected for RNA-seq was significantly different between the drought and control group over all sampling days ($p<0.0001$). Mean $G_s$ was lower in the drought stress plants than the control plants on each sampling day (Fig 1; difference significant except for July 5).

Genotyping of the ten selected samples for RNA-seq with microsatellite markers revealed a mean observed heterozygosity of 0.608, a mean expected heterozygosity of 0.507, and a mean fixation index of -0.076. There were no statistically significant differences for the genetic diversity indices between the drought stress and control group. All individuals had a unique multi-locus genotype, hence no clones were selected.

Sequencing output and de novo transcriptome assembly

Sequencing of the composite sample revealed 43,309,878 raw reads, which resulted in 41,186,808 reads after quality trimming (S2 File). The reads were assembled into 44,868
contigs with an average length of 764 bp, and a N50 contig length of 1,252 bp. After applying cd-hit-est to reduce the redundancy of the assembly, the number of contigs decreased to 44,335. Sequencing of the 50 samples of the drought stress experiment revealed a total of 2,324,791,074 reads after quality trimming. Individual libraries revealed 34,922,516 to 58,414,100 reads (mean 46,495,821.48 reads) (S2 File).

Sequence annotation

BLAST results were obtained for 64.6% of all contigs. Although, an E-value cut-off of 1e^-3 was chosen, most BLAST results were supported by much lower E-values. The complete annotation file can be found in the S3 File. The five species, which gave the highest number of BLAST hits were *Vitis vinifera*, *Citrus sinensis*, *Malus domestica*, *Theobroma cacao*, and *Populus trichocarpa*. GO terms were successfully assigned to 75.3% of the sequences with BLAST hits. GO slim terms with the highest abundance for cellular component, molecular function, and biological process terms were “cell”, “catalytic activity”, and “metabolic process” (Fig 2).

Fig 1. Box plots of stomatal conductance (Gs) measured around noon of the drought stress and control group for the different sampling days. Asterisks indicate significant differences between the groups on each of the sampling days (*p<0.05, **p<0.01).
Identification of differential gene expression

Significantly DEGs between the drought stress and control group were detected for all analyzed sampling days with both applied programs (edgeR and DESeq2) (S4 File), whereby DESeq2 detected more DEGs than edgeR on three of the five days (Fig 3). Nevertheless, many genes were overlapping between the two programs. Over all five sampling days 662 different genes were found to be differentially expressed among the stress and control group (only genes, which were revealed by both, edgeR and DESeq2) (S5 File). Thereby, the number of DEGs varied among sampling days ranging from 65 on June 28 to 364 on July 19 (Fig 4). More genes were downregulated than upregulated in the stress group on each sampling day. Gene expression was relatively specific to the respective sampling date (Fig 5) with 41.5% (June 28) to 66.5% (July 19) of genes exclusively expressed on the considered day. Only five genes (protein yls9-like (contig_2897), UDP-glycosyltransferase74b1-like (contig_1957), receptor-like protein 12 (contig_21713), probable lrr receptor-like serine threonine-protein kinase at4g36180-like (contig_11937), and protein p21-like (contig_6745)) were significantly differentially expressed between the stress and control group on all five sampling days. All of these genes were downregulated in the stressed plants. GO term enrichment was statistically significant for the 662 unique DEGs, whereby GO terms were overrepresented in the set of DEGs compared to the
Fig 3. Venn diagrams of the number of DEGs between drought and control group for the different sampling days revealed by edgeR and DESeq2. Venn diagrams were prepared using the online tool provided by VIB and Ghent University (http://bioinformatics.psb.ugent.be/webtools/Venn/).

https://doi.org/10.1371/journal.pone.0184167.g003

Fig 4. Number of DEGs on the different sampling days.

https://doi.org/10.1371/journal.pone.0184167.g004
total set of transcripts. Enriched GO terms with highest significance were “phospholipid catabolic process” (GO: 0009395), “glycerophospholipid catabolic process” (GO: 0046475), “cellular phosphate ion homeostasis” (GO: 0030643), “cellular anion homeostasis” (GO: 0030002), and “cellular trivalent inorganic anion homeostasis” (GO: 0072502) in the upregulated DEGs (Fig 6), and the GO terms “oxidoreductase activity” (GO: 0016705), “secondary metabolite biosynthetic process” (GO: 0044550), and “secondary metabolic process” (GO: 0019748) were most significantly enriched in the downregulated DEGs (Fig 7).
Eleven of the twelve selected genes showed expression profiles similar to those observed in the RNA-seq experiment (Fig 8). The genes *Galactinol synthase family protein* and *Low-temperature-induced 65 kda* were upregulated in the stress group, while the genes *Nitrate transporter-like*, *Octicosapeptide phox bem1p family isoform 1*, *Protein p21-like*, *CCT motif family protein isoform partial*, *Probable lrr receptor-like serine threonine-protein kinase at4g36180-like*, *Receptor-like protein 12*, *UDP-glycosyltransferase74b1-like*, *Protein yls9-like* and *Serine-threonine protein plant* were downregulated in this group. Only the expression level of the gene *Cytochrome p450* was not statistically significant between the stress and control group contrary to the
Fig 7. GO terms significantly enriched in downregulated DEGs compared to the reference gene set (total set of sequences with assigned GO terms) over all sampling days.

https://doi.org/10.1371/journal.pone.0184167.g007
RNA-seq data. The expression levels, however, became significantly different between the two groups for this gene, after exclusion of the samples not used in the RNA-seq experiment.

**Discussion**

**Drought stress experiment and SSR genotyping**

Lower means of stomatal conductance measured at noon ($G_s$) between the stress and control group on four of the five measuring days indicate that the drought treatment negatively affected leaf water status and gas exchange regulation. However, $G_s$ was also dependent on weather conditions, notably vapor pressure deficit (VPD), which varied over the course of the experiment. The drought-exposed plants also developed a number of morphological, 

---

Fig 8. Comparison of expression patterns between RNA-seq and qPCR for selected transcripts. Bars indicate mean log$_2$ fold change and whiskers indicate standard errors.

https://doi.org/10.1371/journal.pone.0184167.g008
anatomical and other physiological modifications to water shortage (reduced aboveground productivity and root length, smaller leaf areas, reduced xylem hydraulic conductivity, smaller vessel diameters, higher embolism resistance), which distinguished them from the control plants [20, 38, 39].

SSR genotyping was used to characterize the neutral genetic structure of the ten selected individuals for RNA-seq. The analysis revealed observed and expected heterozygosities similar to those revealed by other studies in European beech (e.g., [25, 40, 41]. The genetic diversity indices were not different between the control and stress group. All individuals showed unique multilocus genotypes, hence no clones were selected for the analysis and the ten individuals represent real biological replicates.

**De novo transcriptome assembly and sequence annotation**

The *de novo* transcriptome assembly was based on a normalized composite sample comprising all samples of the experiment to maximize gene discovery. The normalization step reduces high abundance transcripts and equalizes transcript concentrations. Hence, the redundancy of the cDNA library is reduced, which increases the efficiency of sequencing and rare gene discovery. In total, 41,186,808 high quality sequencing reads were assembled into 44,868 contigs. The average length of 764 bp and a N50 contig length of 1,252 bp are comparable to the results of *de novo* transcriptome assemblies in other forest tree species [19, 42, 43]. The usage of the program cd-hit-est [28] to reduce the redundancy of the assembly only slightly decreased the number of contigs (to 44,335). This shows that the assembly algorithm implemented in the CLC Genomics workbench and the normalization of the cDNA library resulted in a low redundancy *de novo* transcriptome assembly. This is in line with studies, which identified CLC as one of the leading assemblers producing low redundant assemblies [44, 45].

BLAST results were obtained for 64.6% of all contigs, whereby the five woody taxa *Vitis vinifera*, *Citrus sinensis*, *Malus domestica*, *Theobroma cacao*, and *Populus trichocarpa* were the species, which gave the highest number of BLAST hits. Although many transcripts were not functionally annotated, this study provides more than 28,500 annotated transcripts, which can directly be used for further research in European beech. The majority of the unannotated transcripts may be due to the lack of a reference genome of *F. sylvatica*, but the data set may also include some new beech-specific transcripts, since pooling and normalization of samples for *de novo* assembly should have enhanced the power of gene detection. GO-slim mapping was used to get an overview of the GO-term distribution over the whole transcriptome (based on the composite sample). The GO-terms with the highest abundance in the three categories “cellular component”, “molecular function”, and “biological process” were similar to the results of [19], who investigated the beech transcriptome related to dormancy regulation.

**Identification of differential gene expression**

For the identification of differentially expressed genes the two widely used tools edgeR and DESeq2 [33, 34] were used. Both tools were found to perform better than other tools when the number of biological replicates was lower than 12 (as in the present study) [46]. Since different studies showed that the number of significantly DEGs can differ between edgeR and DESeq2 (or DESeq the previous version of DESeq2) [46–48], we decided to apply both of them. Indeed, both programs revealed a different number of DEGs for the different sampling days, whereby DESeq2 detected more DEGs than edgeR2 on three of the five days. Nevertheless, many DEGs were revealed by both programs. Only these overlapping genes were regarded as differentially expressed between the drought stress and control group in this study. Hence, the reported number of DEGs might be a rather conservative estimate.
Interestingly, the number of DEGs differed among sampling days. It increased from the first to the fourth sampling day and decreased on the last sampling day. The same pattern was observed for GO term distribution (S6 File), whereas the GO terms “ion binding” (GO: 0043167), “organic cyclic compound binding” (GO: 0097159), “heterocyclic compound binding” (GO: 1901363), and “transferase activity” (GO: 0016740) were among the most frequent ones on all sampling days. Zhang et al. [49] found a correlation of the number of drought-regulated genes with drought stress intensity and duration. This might also partially explain the observed pattern in the present study. While drought stress intensity (expressed by the SWC of the pots) was kept relatively constant during the experiment, the duration of drought stress increased from the first to the last sampling day. This, however, cannot explain the decrease of the number of DEGs on the last sampling day. The difference in gene expression in course of the experiment may rather be explained by a combination of both, the duration of drought stress treatment, and some variation in drought stress levels due to weather conditions resulting in different VPDs as indicated by the stomatal conductance. Furthermore, samples were not taken at exactly the same time on the different sampling days, which may additionally have induced some variation in drought-induced gene expression.

Only five genes (protein yls9-like (contig_2897), UDP-glycosyltransferase74b1-like (contig_1957), receptor-like protein 12 (contig_21713), probable lrr receptor-like serine threonine-protein kinase at4g36180-like (contig_11937), and protein p21-like (contig_6745)) were significantly differentially expressed between the stress and control group on all five sampling days. All of these genes are assumed to be involved in some kind of stress resistance. For instance, a UDP-glycosyltransferase74b1-like gene was shown to be upregulated under drought stress in tobacco roots [50]. UDP-glycosyltransferases are the most common enzymes catalyzing glycosylation, a process required in a number of biological processes during plant growth and development, and can also be involved in abiotic stress adaptation [51]. The YLS9 gene is related to leaf senescence in Arabidopsis thaliana, an important process for plant survival and adaptation to unfavorable environmental conditions [52]. Furthermore, the sequence of protein YLS9-like contains a LEA domain. LEA proteins play crucial roles in cellular dehydration tolerance [53]. Receptor-like proteins (RLPs) are cell surface receptors that typically obtain a leucine-rich repeat domain [54]. This kind of domain was also found in the sequences of receptor-like protein 12 and probable lrr receptor-like serine threonine-protein kinase at4g36180-like in the present study. RLPs and receptor-like protein kinases are involved in different biological processes such as plant development, disease resistance, and stress tolerance [55, 56]. Protein P21 is a member of the PR-5 family, the members of which are also known as thaumatin-like proteins (TLPs), and are responsive to biotic and abiotic stress [57, 58]. Interestingly, all of these five genes in the present study were downregulated in drought stress plants compared to the control group. Thus, their expression might regulate some downstream stress response genes/pathways. Li et al. [51] for example, showed that the expression of a UDP-glycosyltransferase influenced the expression of other stress-inducible genes. Downregulation of the gene enhanced drought tolerance in Arabidopsis seedlings, whereas on over-expression reduced drought tolerance. The opposite trend was detected for mature Arabidopsis plants [51]. Nevertheless, the function of the five genes in drought stress response in beech remains open in our study, but they are interesting candidates for further studies. The different expression of these genes between the stress and control group on each of the sampling days indicate an important role in stress tolerance in beech. Using the developed primers for qPCR, these genes can immediately be investigated in other beech populations or desiccation experiments. Furthermore, the genes are candidate genes for drought stress tolerance, which can be used in association studies.
Over all five sampling days, 662 different genes were found to be differentially expressed between the stress and control group. We performed an enrichment analysis for GO terms in all up- and downregulated DEGs separately. GO terms with highest significance in the upregulated DEGs were “phospholipid catabolic process” (GO: 0009395), “glycerophospholipid catabolic process” (GO: 0046475), “cellular phosphate ion homeostasis” (GO: 0030643), “cellular anion homeostasis” (GO: 0030002), and “cellular trivalent inorganic anion homeostasis” (GO: 0072502). Thus, mainly genes involved in lipid- and homeostasis-related processes were upregulated. Homeostasis signaling leads to stress tolerance in plants [59], and lipids are important membrane components. Changes in their composition may help to maintain membrane integrity under water stress [60]. In the downregulated DEGs most significantly enriched GO terms were “oxidoreductase activity” (GO: 0016705), “secondary metabolite biosynthetic process” (GO: 0044550), and “secondary metabolic process” (GO: 0019748). All these GO terms can be connected with oxidative stress. Drought stress can cause oxidative stress through an enhanced production of reactive oxygen species (ROS) [61]. At low levels, however, ROS may also function as components of the stress-signaling pathway, triggering defense and/or acclimation responses [61, 62]. Antioxidant enzymes have often been called “first line of defense” against oxidative stress, whereby the activity of these enzymes can be enhanced or repressed depending on species, genotype, and stress duration/severity [63–66].

Shinozaki and Yamaguchi-Shinozaki (2007) [67] classified products of drought stress-inducible genes in Arabidopsis into functional ones, involved in abiotic stress tolerance (e.g., water channels, LEA proteins, chaperones, detoxification enzymes), and regulatory ones, involved in stress-responsive gene expression (e.g., transcription factors, protein kinases, enzymes involved in phospholipid metabolism). The 662 DEGs between the stress and normally watered plants include many genes, which can be assigned to these groups and/or are common drought stress inducible genes (e.g., late embryogenesis abundant proteins, aquaporins, heat shock proteins, and protein kinases). Similar results were revealed by a recent study of the water stress-related transcriptome of Quercus lobata [68], a species from the same family as European beech (Fagaceae).

In general, the results of RNA-seq may not be representative for the whole exome, since genes with low expression levels may be missed, or some genes are expressed in unsampled tissue [69]. In the present study, we only sampled leaf material, hence, genes involved in drought stress response may be different in other plant tissues such as roots or xylem. Furthermore, one weakness of the study design is that we did not take samples at the very beginning of the drought stress treatment, and the plants had been subjected to drought before the start of the investigated drought experiment. Thus, genes involved in the first response to drought stress may not have been identified. Some acclimations to drought stress from the first drought treatment can also not be ruled out, since plants can develop a stress memory after repeated drought periods [70].

Quantitative real-time PCR

Quantitative real-time PCR was used to confirm differential gene expression revealed by RNA-seq. For qPCR validation, samples from individuals used for RNA-seq and additional samples from the drought stress experiment not used for RNA-seq were investigated. Eleven of the twelve selected genes showed similar expression as in the RNA-seq experiment. This is in line with other studies, which revealed a high correlation between qPCR and RNA-seq data [71–73]. Only for one tested gene (Cytochrome p450), the expression pattern revealed by RNA-seq could not be confirmed by qPCR. Nevertheless, after exclusion of the samples not used in
the RNA-seq experiment, a validation of the RNA-seq experiment was possible. Thus, the expression of this gene might be genotype-specific.

Conclusions

In this study, we report more than 28,500 annotated transcripts, which can directly be used for adaptation research in *F. sylvatica*. In total, 662 DEGs were identified between the drought stress and control group. These genes are candidate genes for drought stress adaptation, and can, for instance, be further investigated in association studies. This study shows that gene expression can be specific for different time points during drought stress treatment. GO term enrichment revealed that mostly genes involved in lipid- and homeostasis-related processes were upregulated in the stress plants, whereas genes involved in oxidative stress response were downregulated. This is a first insight into the genomic drought stress response of European beech. Further research may unravel the mechanism of genomic drought stress adaptation in greater detail by investigating important mechanisms of gene regulation such as alternative splicing or epigenetic effects. Nevertheless, the establishment of a reference genome for *F. sylvatica* would be an important prerequisite for a deeper understanding of adaptation in this species.

Supporting information

S1 File. Genes with regarding primer sequences used for qPCR. All genes were amplified with an annealing temperature of 58˚C. The contig numbers refer to the sequence description in the transcriptome assembly. (PDF)

S2 File. Overview over the quality and quantity of sequencing reads for each sample before and after trimming. (PDF)

S3 File. Annotation for the complete set of contigs. (XLSX)

S4 File. Lists of differentially expressed genes between the stress and control group for the different sampling days. (XLSX)

S5 File. List of all 662 differentially expressed genes between the stress and control group over all sampling days. (XLSX)

S6 File. GO term distribution for the different sampling days. (PDF)

Acknowledgments

The drought experiment with the beech saplings was conducted in the Experimental Botanical Garden of Goettingen University; we gratefully acknowledge the experimental support and plant care provided by the team of gardeners. We further thank Julia Beck for support during sequencing.
Author Contributions

Conceptualization: Markus Müller, Sarah Seifert, Torben Lübbe, Christoph Leuschner, Reiner Finkeldey.

Data curation: Markus Müller.

Formal analysis: Markus Müller.

Funding acquisition: Christoph Leuschner, Reiner Finkeldey.

Investigation: Markus Müller, Sarah Seifert, Torben Lübbe.

Methodology: Markus Müller, Sarah Seifert, Torben Lübbe, Christoph Leuschner, Reiner Finkeldey.

Project administration: Christoph Leuschner, Reiner Finkeldey.

Resources: Christoph Leuschner, Reiner Finkeldey.

Supervision: Christoph Leuschner, Reiner Finkeldey.

Validation: Markus Müller.

Visualization: Markus Müller.

Writing – original draft: Markus Müller.

Writing – review & editing: Markus Müller, Torben Lübbe, Christoph Leuschner, Reiner Finkeldey.

References


38. Lübke T, Schuld B, Leuschner C. Acclimation of leaf water status and stem hydraulics to drought and tree neighbourhood: alternative strategies among the saplings of five temperate deciduous tree species. Tree Physiol. 2016.


