Supplementary information

Mining differential gene expression in *Fagus crenata* seedlings in response to short-term soil drought stress

Takeshi Torimaru, Hinako Ao, Yasuaki Akaji, Shinji Akada, Yasunori Ohmiya, Yosuke Matsuda, Hiromitsu Kisanuki, Nobuhiro Tomaru

Appendix 1 *De novo* genome assembly for *Fagus crenata*

DNA extraction and sequencing

To obtain the reference genome for *Fagus crenata*, an individual tree standing in a common garden at Mie University, Tsu, Japan, was selected. The individual originated from a population in the northernmost region of the species (southern Hokkaido, northern Japan), and was estimated to be about 30 years old in 2024. In April 2019, genomic DNA was extracted from fresh leaves using NucleoBond® HMW DNA (MACHEREY-NAGEL GmbH & Co. KG, Germany), and extracts >20 kbp were selected using a Short Read Eliminator (PacBio Inc., United States). A long-read library was built using a Ligation Sequencing Kit 1D SQK-LSK109 (Oxford Nanopore Technologies, United Kingdom) and sequenced by Nanopore PromethION (Oxford Nanopore Technologies) though a commercial sequencing supplier (GeneBay Inc., Japan). The same tree was sampled again in April 2020 to obtain further fresh leaves for constructing an Omni-C library; the DNA extraction and library construction, as well as sequencing (using an Illumina Novaseq 6000), were carried out by a commercial sequencing supplier (Dovetail Inc., United States). The DNA extraction, labeling by direct labelling enzyme (DLE1), and molecular imaging on a Bionano Saphyr instrument, were also carried out by a commercial supplier (AS ONE Co., Japan).

Genome assembly

Raw reads generated from a Nanopore sequencer (DRR455010) were filtered through

NanoFilt (De Coster et al. 2018) with option '-q 7 -l 1000' (ca. 2.1 M reads, total sequence length ca. 41.2 Gb), which was then used for hybrid assembly using an MaSuRCA v3.3.3 assembler (Zimin et al. 2013; Zimin et al. 2017) combined with Illumina short reads (2×150bp) from (Tsukamoto et al. 2020) (DRR190836). The barcodes of 10X Genomics Chromium linked reads from (Tsukamoto et al. 2020) were processed using Longranger basic v2.2.2 (https://support.10xgenomics.com/genome-exome/software/pipelines/latest/installation), and these processed reads were used in Tigmint v1.1.2 (Jackman et al. 2018) to rectify the mis-assemblies present in the draft genome. The corrected draft genome was scaffolded using ARKS v1.1.1 (Coombe et al. 2018) and LINKS v1.8.7 (Warren et al. 2015), and the improved draft genome was soft-masked with WindowMasker v1.0.0 (Morgulis et al. 2006). By applying HaploMerger2 (Huang et al. 2017), we then inferred a scoring matrix specific to the beech genome, using 5% of the soft-masked diploid assembly; the haploid genome was obtained from the soft-masked assembly and the inferred scoring matrix. Using the long reads corrected (423K reads, total sequence length ca. 16Gb) by the correction module in NECAT v0.0.1 (Chen et al. 2021), additional redundancy of the haploid genome was removed using Purge Haplotigs v1.1.0 (Roach et al. 2018). The clean Omni-C reads [ca. 152M reads (2×150bp), total sequence length ca. 23Gb] (DRR539902) were mapped onto the new haploid genome using Juicer v1.6 (Durand et al. 2016), and 3D-DNA v180922 (Dudchenko et al. 2017) was used to construct the chromosome-scale genome sequences with specific parameters as follows: --input 15000 --editor-saturation-centile 10 --editor-coarse-resolution 100000 --editor-coarse-region 400000 --editor-repeat-coverage 100. Following Moll et al. (2017), in which the application of

BioNano technology following Dovetail resulted in improvements compared with Dovetail and BioNano alone, we applied an optical map to the assembled sequences. To do so, runBNG v2.0 pipelines (Yuan et al. 2017) were used to construct a consensus map [357 consensus maps (total genome map length 1015Mb, genome map N50 ca. 9.3Mb)] based on the overlapping information between the optical maps and the assembled sequences. The consensus map was applied to separate the chimeric sequences with Chimericognizer (Pan & Lonardi 2019) as well as to scaffold the contigs generated by OMGS (Pan et al. 2020). Following gap closing using LR_Gapcloser (Xu et al. 2019) with the long reads corrected as above, the haploid assembly was again scaffolded using 3D-DNA v180922 (Dudchenko et al. 2017), with the parameters mentioned above. Finally, scaffolds shorter than 1 Kb were removed from the assembly.

Genome annotation

Genome annotation was performed on the final assembly. We used RepeatModeler v2.0.5 (Flynn et al. 2020) to construct a *de novo* repeat library. The assembly was soft-masked using RepeatMasker v4.1.5 (Smit et al. 2013-2015), with the aid of the identified repeats as well as additional repeats of *Arabidopsis thaliana* obtained from msRepDB (Liao et al. 2022). Protein sequences from *A. thaliana* (https://www.arabidopsis.org/download_files/Proteins/Araport 11_protein_lists/Araport11_pep_20220914.gz) were aligned to the repeat-masked assembly using Exonerate v2.4.0 (Slater & Birney 2005). As for the RNA-seq reads (sequenced by Illumina HiseqX, 2×100bp) (DRR539898−DRR539901), which were obtained from the leaves of four beech trees in a clonal seed orchard in Tohoku

Breeding Center (Takizawa, Japan), Trimmomatic v0.39 (Bolger et al. 2014) was applied to clean the data, and a *de novo* transcriptome assembly was created using Trinity v2.9.1 (Grabherr et al. 2011). This transcriptome and the soft-masked genome assemblies were input into PASA v2.5.2 (Haas et al. 2008) to produce PASA assemblies for annotation. Gene predictions were performed on the repeat-masked assembly with two different programs: BRAKER v2.1.6 (Brůna et al. 2021) [with the aid of GeneMark-ET (Lomsadze et al. 2014) and AUGUSTUS v3.4.0 (Stanke et al. 2006)] and GeMoMa v1.7.1 (Keilwagen et al. 2018), both of which incorporated the evidence from the RNA-seq data by mapping to the soft-masked genome assembly using Hisat2 v2.1.1 (Kim et al. 2019). We used the following databases to search for homologous proteins in the GeMoMa: *A. thaliana* (NCBI RefSeq assembly GCF_000001735.4), *Populus trichocarpa* (GCF_000002775.49), *Rosa chinensis* (GCF_002994745.2), and *Quercus robur* (Qrob_PM1N.fa, Qrob_PM1N_genes_20161004.gff, https://www.oakgenome.fr/index8568.html?page_id=587). Finally, the results deduced from the transcriptome evidence and gene prediction were combined into consensus coding sequence models using EVidenceModeler v1.1.1 (Haas et al. 2008). The completeness of the chromosome assembly was assessed using BUSCO v5.4.7 (Manni et al. 2021) with 2,326 single-copy orthologs from the eudicots_odb10 database.

Functional annotation

Proteins translated from the predicted genes were queried against the non-redundant database from NCBI (downloaded from https://ftp.ncbi.nlm.nih.gov/blast/db/ on 2023-02-09) using DIAMOND v2.0.15 blastp (Buchfink et al. 2021), to find any

105 homology (with *E*-values < 10^{-5}) between the predicted proteins and sequences of
106 known function. InterProScan-5.54-87.0 (Jones et al. 2014) was used to predict protein
107 family membership as well as the presence of functional domains and sites in the
108 predicted proteins. These results files were formatted by xml and input into Blast2GO
109 Basic 6.0
110 [https://resources.biobam.com/blast2go/prod/installers/Blast2GO_windows-x64_6_0_3.
111 exe; see also Conesa et al. (2005)], to assign the final functional annotations to the
112 genes.
113
114 *The species' genome feature*
115
116 The features of the assembled and analyzed genome
117 (BKZX02000001-BKZX02000625) were comparable to those reported for *F. sylvatica*
118 (Mishra et al. 2022), which is phylogenetically closely related to *F. crenata* (i.e. within
119 the genus *Fagus*) (Cardoni et al. 2022), especially in terms of assembled genome size
120 (541 Mb in *F. sylvatica*) and gene spaces, including scaffold N50 (46.6 Mb), GC content
121 (35.65%), gene average length (3919 bases), and average number of exons per gene
122 (4.9) (Misha et al. 2022). Our reference genome corroborates the base chromosome
123 number of 12, as previously reported based for chromosome counts (Ohba 1989). The
124 longest 12 scaffolds comprised ca. 518 Mb (Table S1), which is in accordance with a
125 previous study (530 Mb) that estimated the genome size using the *k*-mer method
126 (Tsukamoto et al. 2020). The number of genes predicted in *F. crenata* (35,116) was
127 similar to that in *F. sylvatica* (34,987) as estimated from a unigene set construction
128 (Mishra et al. 2022). The numbers in the gene *Fagus* are also comparable to *Quercus*

spp., a major genus in Fagaceae, including *Q. robur* [25,808, Plomion et al. (2018)], *Q. acutissima* [29,889, Liu et al. (2022)], *Q. variabilis* [32,466, Han et al. (2022)], *Q. mongolica* [36,553, Ai et al. (2022)], and *Q. lobata* [39,373, Sork et al. (2022)]. In addition, the large proportions of complete BUSCO genes, as well as small proportions of missing and fragmented genes, could be found not only at the gene level but also the protein level (Table S2), further conforming that the gene model developed in the present study is acceptable for annotating the genome of *F. crenata* [see also Hurgobin (2022) for quality checking the annotation of a genome].

References

Ai WF, Liu YQ, Mei M, Zhang XL, Tan EG, Liu HZ, Han XY, Zhan H, Lu XJ (2022) A chromosome-scale genome assembly of the Mongolian oak (*Quercus mongolica*). Mol Ecol Resour 22:2396-2410. https://doi.org/10.1111/1755-0998.13616

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform 3:lqaa108. https://doi.org/10.1093/nargab/lqaa108

Buchfink B, Reuter K, Drost H-G (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18:366-368. https://doi.org/10.1038/s41592-021-01101-x

Cardoni S, Piredda R, Denk T, Grimm GW, Papageorgiou AC, Schulze ED, Scoppola A, Shanjani PS, Suyama Y, Tomaru N, Worth JRP, Simeone MC (2022) 5S-IGS rDNA in wind-pollinated trees (*Fagus* L.) encapsulates 55 million years of reticulate evolution and hybrid origins of modern species. Plant J 109:909-926. https://doi.org/10.1111/tpj.15601

Chen Y, Nie F, Xie SQ, Zheng YF, Dai Q, Bray T, Wang YX, Xing JF, Huang ZJ, Wang DP, He LJ, Luo F, Wang JX, Liu YZ, Xiao CAL (2021) Efficient assembly of nanopore reads via highly accurate and intact error correction. Nat Commun 12:60. https://doi.org/10.1038/s41467-020-20236-7

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676. https://doi.org/10.1093/bioinformatics/bti610

Coombe L, Zhang J, Vandervalk BP, Chu J, Jackman SD, Birol I, Warren RL (2018) ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. BMC Bioinformatics 19:234. https://doi.org/10.1186/s12859-018-2243-x

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666-2669. https://doi.org/10.1093/bioinformatics/bty149

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL (2017) *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356:92-95. https://doi.org/10.1126/science.aal3327

174 Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL
175 (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C
176 experiments. Cell Syst 3:95-98. https://doi.org/10.1016/j.cels.2016.07.002
177 Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF (2020)
178 RepeatModeler2 for automated genomic discovery of transposable element families.
179 Proc Natl Acad Sci USA 117:9451-9457. https://doi.org/10.1073/pnas.1921046117
180 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan
181 L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N,
182 di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011)
183 Full-length transcriptome assembly from RNA-seq data without a reference genome.
184 Nat Biotechnol 29:644-652. https://doi.org/10.1038/nbt.1883
185 Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR,
186 Wortman JR (2008) Automated eukaryotic gene structure annotation using
187 EVidenceModeler and the program to assemble spliced alignments. Genome Biol
188 9:R7. https://doi.org/10.1186/gb-2008-9-1-r7
189 Han B, Wang LX, Xian Y, Xie XM, Li WQ, Zhao Y, Zhang RG, Qin XC, Li DZ, Jia KH
190 (2022) A chromosome-level genome assembly of the Chinese cork oak (*Quercus*
191 *variabilis*). Front Plant Sci 13:1001583. https://doi.org/10.3389/fpls.2022.1001583
192 Huang SF, Kang MJ, Xu AL (2017) HaploMerger2: rebuilding both haploid
193 sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics
194 33:2577-2579. https://doi.org/10.1093/bioinformatics/btx220
195 Hurgobin B (2022) Annotation of protein-coding genes in plant genomes. In: Plant
196 bioinformatics: methods and protocols (ed. Edwards D), pp. 309-326. Humana Press,
197 New York

198 Jackman SD, Coombe L, Chu J, Warren RL, Vandervalk BP, Yeo S, Xue ZY, Mohamadi
199 H, Bohlmann J, Jones SJM, Birol I (2018) Tigmint: correcting assembly errors using
200 linked reads from large molecules. BMC Bioinformatics 19:393.
201 https://doi.org/10.1186/s12859-018-2425-6
202 Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J,
203 Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M,
204 Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function
205 classification. Bioinformatics 30:1236-1240.
206 https://doi.org/10.1093/bioinformatics/btu031
207 Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J (2018) Combining RNA-seq
208 data and homology-based gene prediction for plants, animals and fungi. BMC
209 Bioinformatics 19:189. https://doi.org/10.1186/s12859-018-2203-5
210 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome
211 alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol
212 37:907-915. https://doi.org/10.1038/s41587-019-0201-4
213 Liao XY, Hu K, Salhi A, Zou Y, Wang JX, Gao X (2022) msRepDB: a comprehensive
214 repetitive sequence database of over 80000 species. Nucleic Acids Res
215 50:D236-D245. https://doi.org/10.1093/nar/gkab1089
216 Liu D, Xie XM, Tong BQ, Zhou CC, Qu K, Guo HL, Zhao ZH, El-Kassaby YA, Li W,
217 Li WQ (2022) A high-quality genome assembly and annotation of *Quercus*
218 *acutissima* Carruth. Front Plant Sci 13:1068802.
219 https://doi.org/10.3389/fpls.2022.1068802

220 Lomsadze A, Burns PD, Borodovsky M (2014) Integration of mapped RNA-seq reads
221 into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res
222 42:e119. https://doi.org/10.1093/nar/gku557
223 Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM (2021) BUSCO update:
224 novel and streamlined workflows along with broader and deeper phylogenetic
225 coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol
226 38:4647-4654. https://doi.org/10.1093/molbev/msab199
227 Mishra B, Ulaszewski B, Meger J, Pfenninger M, Gupta DK, Wötzel S, Ploch S,
228 Burczyk J, Thines M (2022) A chromosome-level genome assembly of the European
229 beech (*Fagus sylvativa*) reveals anomalies for organelle DNA integration, repeat
230 content and distribution of SNPs. Front Genet 12:691058.
231 https://doi.org/10.3389/fgene.2021.691058
232 Moll KM, Zhou P, Ramaraj T, Fajardo D, Devitt NP, Sadowsky MJ, Stupar RM, Tiffin P,
233 Miller JR, Young ND, Silverstein KAT, Mudge J (2017) Strategies for optimizing
234 BioNano and Dovetail explored through a second reference quality assembly for the
235 legume model, *Medicago truncatula*. BMC Genomics 18:578.
236 https://doi.org/10.1186/s12864-017-3971-4
237 Morgulis A, Gertz EM, Schäffer AA, Agarwala R (2006) WindowMasker:
238 window-based masker for sequenced genomes. Bioinformatics 22:134-141.
239 https://doi.org/10.1093/bioinformatics/bti774
240 Ohba H (1989) Fagaceae. In: Wild flowers of Japan woody plants (eds. Satake Y, Hara
241 H, Watari S, Tominari T). Heibonsha Ltd., Tokyo, Japan (in Japanese)
242 Pan WH, Jiang T, Lonardi S (2020) OMGS: optical map-based genome scaffolding. J
243 Comput Biol 27:519-533. https://doi.org/10.1089/cmb.2019.0310

244 Pan WH, Lonardi S (2019) Accurate detection of chimeric contigs via Bionano optical
245 maps. Bioinformatics 35:1760-1762. https://doi.org/10.1093/bioinformatics/bty850
246 Plomion C, Aury JM, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillonne
247 N, Labadie K, Le Provost G, Lesur I, Bartholomé J, Faivre-Rampant P, Kohler A,
248 Leplé JC, Chantret N, Chen J, Diévart A, Alaeitabar T, Barbe V, Belser C, Bergès H,
249 Bodénès C, Bogeat-Triboulot MB, Bouffaud ML, Brachi B, Chancerel E, Cohen D,
250 Couloux A, Da Silva C, Dossat C, Ehrenmann F, Gaspin C, Grima-Pettenati J,
251 Guichoux E, Hecker A, Herrmann S, Hugueney P, Hummel I, Klopp C, Lalanne C,
252 Lascoux M, Lasserre E, Lemainque A, Desprez-Loustau ML, Luyten I, Madoui MA,
253 Mangenot S, Marchal C, Maumus F, Mercier J, Michotey C, Panaud O, Picault N,
254 Rouhier N, Rué O, Rustenholz C, Salin F, Soler M, Tarkka M, Velt A, Zanne AE,
255 Martin F, Wincker P, Quesneville H, Kremer A, Salse J (2018) Oak genome reveals
256 facets of long lifespan. Nat Plants 4:440-452.
257 https://doi.org/10.1038/s41477-018-0172-3
258 Roach MJ, Schmidt SA, Borneman AR (2018) Purge Haplotigs: allelic contig
259 reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 19:460.
260 https://doi.org/10.1186/s12859-018-2485-7
261 Slater GS, Birney E (2005) Automated generation of heuristics for biological sequence
262 comparison. BMC Bioinformatics 6:31. https://doi.org/10.1186/1471-2105-6-31
263 Smit AFA, Hubley R, Green P (2013-2015) RepeatMasker Open-4.0.
264 http://wwwrepeatmaskerorg
265 Sork VL, Cokus SJ, Fitz-Gibbon ST, Zimin AV, Puiu D, Garcia JA, Gugger PF,
266 Henriquez CL, Zhen Y, Lohmueller KE, Pellegrini M, Salzberg SL (2022)

High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. Nat Commun 13:2047. https://doi.org/10.1038/s41467-022-29584-y

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res 34:W435-W439. https://doi.org/10.1093/nar/gkl200

Tsukamoto M, Akada S, Matsuda S, Jouyu H, Kisanuki H, Tomaru N, Torimaru T (2020) Assessments of fine-scale spatial patterns of SNPs in an old-growth beech forest. Heredity 125:240-252. https://doi.org/10.1038/s41437-020-0334-8

Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, Jones SJM, Birol I (2015) LINKS: scalable, alignment-free scaffolding of draft genomes with long reads. GigaScience 4:35. https://doi.org/10.1186/s13742-015-0076-3

Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW, Li JT (2019) LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. GigaScience 8:giy157. https://doi.org/10.1093/gigascience/giy157

Yuan YX, Bayer PE, Lee HT, Edwards D (2017) runBNG: a software package for BioNano genomic analysis on the command line. Bioinformatics 33:3107-3109. https://doi.org/10.1093/bioinformatics/btx366

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. Bioinformatics 29:2669-2677. https://doi.org/10.1093/bioinformatics/btt476

Zimin AV, Puiu D, Luo MC, Zhu TT, Koren S, Marçais G, Yorke JA, Dvorák J, Salzberg SL (2017) Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res 27:787-792. https://doi.org/10.1101/gr.213405.116

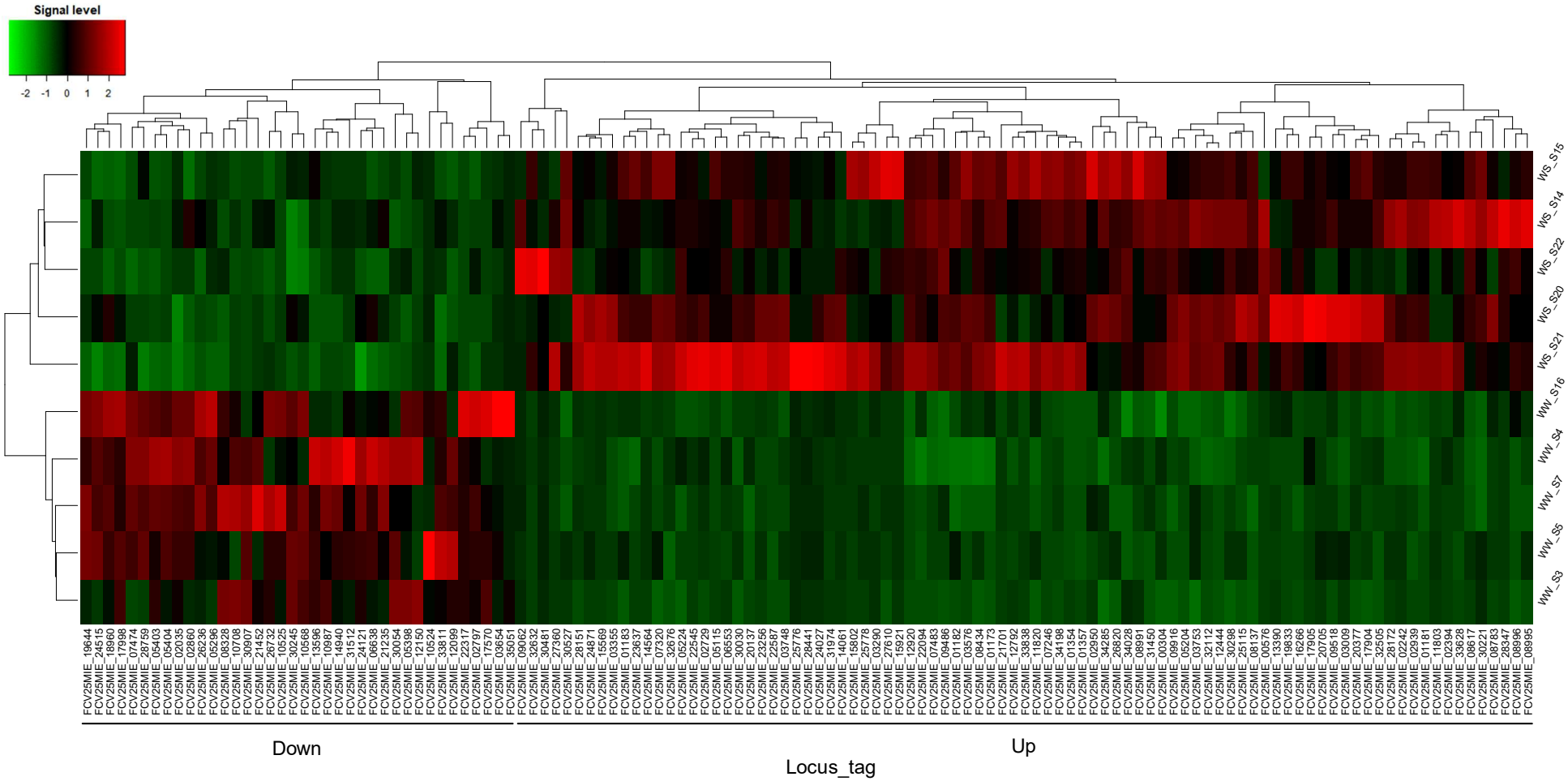

Fig. S1 A heat map of differentially expressed genes of *Fagus crenata* seedling leaves between well-watered (WW) and water-stressed (WS) treatments. The upper dendrogram shows the similarity in expression patterns of 38 down- and 89 up-regulated genes. Locus_tag represents the genes obtained from our *de novo* assembly, which has been registered with DDBJ (accession no. BKZX02000001-BKZX02000625).

Table S1 A summary of the assembly statistics for *Fagus crenata*

Assembly size (Mb)	539.1
No. of contigs	2290
Contig N50 (Kb)	851.9
No. of scaffolds	623
Scaffold N50 (Mb)	45
Longest scaffold (Mb)	80.9
Total interspersed repeats (%)	51.54
No. of genes	35116
GC content (%)	35.14
Gene average length (base)	3602.5
Average exons per gene	4.6
Pseudo-chromosome size (Mb)	518.2

Table S2 Summary statistics for the completeness of the *de novo* assembly for *Fagus crenata*

Source	Genome	Protein
Complete (%)	99.1	97.3
Single genes (%)	90.8	87.5
Duplicated gene (%)	8.3	9.8
Fragmented genes (%)	0.2	0.9
Missing genes (%)	0.7	1.8

316

317

Table S3 RNA-seq read information used for mapping to the *Fagus crenata* reference genome

Treatment	Fastq file names	No. of reads	Total length of sequences (bp)
Well-watered	S3_1.fq.gz	26,621,919	3,950,454,858
Well-watered	S3_2.fq.gz	26,621,919	3,959,942,370
Well-watered	S4_1.fq.gz	33,168,177	4,930,336,650
Well-watered	S4_2.fq.gz	33,168,177	4,931,516,062
Well-watered	S5_1.fq.gz	32,005,980	4,759,385,012
Well-watered	S5_2.fq.gz	32,005,980	4,758,457,387
Well-watered	S7_1.fq.gz	30,944,480	4,598,682,785
Well-watered	S7_2.fq.gz	30,944,480	4,597,127,136
Well-watered	S16_1.fq.gz	30,782,793	4,554,796,465
Well-watered	S16_2.fq.gz	30,782,793	4,553,103,260
Water-stressed	S14_1.fq.gz	41,228,841	6,129,039,215
Water-stressed	S14_2.fq.gz	41,228,841	6,124,266,027
Water-stressed	S15_1.fq.gz	39,085,314	5,817,496,187
Water-stressed	S15_2.fq.gz	39,085,314	5,812,171,227
Water-stressed	S20_1.fq.gz	41,336,584	6,144,497,878
Water-stressed	S20_2.fq.gz	41,336,584	6,136,429,900
Water-stressed	S21_1.fq.gz	38,532,098	5,730,026,565
Water-stressed	S21_2.fq.gz	38,532,098	5,732,982,538
Water-stressed	S22_1.fq.gz	33,665,383	5,009,465,125
Water-stressed	S22_2.fq.gz	33,665,383	5,012,197,690

Table S4 Locus_tag list for up-regulated differentially expressed genes (DEGs) in *Fagus crenata* seedlings

Rank	Locus_tag	Accession[a]	Product inferred from coding sequences[b]	Log_2(fold change)	FDR rate
1	FCV25MIE_12444	AT1G36370.1	Serine hydroxymethyltransferase 7	1.493	6.09E-15
2	FCV25MIE_07246	AT4G37850.1	Transcription factor bHLH25-like	7.316	1.04E-11
3	FCV25MIE_16266	AT3G04880.1	DNA damage-repair/toleration protein DRT102	2.113	2.10E-10
4	FCV25MIE_05115	Not assigned	5′-adenylylsulfate reductase 1, chloroplastic-like	4.250	1.25E-09
5	FCV25MIE_08137	AT3G29575.1	Ninja-family protein AFP3-like isoform X1	1.464	1.50E-09
6	FCV25MIE_01357	AT5G48850.1	Protein sulfur deficiency-induced 1-like	4.327	4.36E-09
7	FCV25MIE_22587	AT3G22740.1	Homocysteine S-methyltransferase 3	2.187	1.02E-08
8	FCV25MIE_25115	AT4G33950.1	Serine/threonine-protein kinase SAPK3-like	1.060	1.02E-08
9	FCV25MIE_12920	AT5G26220.1	Gamma-glutamylcyclotransferase 2-1-like	2.068	1.02E-08
10	FCV25MIE_30298	AT5G35810.1	Ankyrin repeat-containing protein At5g02620-like isoform X1	1.163	1.83E-08
11	FCV25MIE_28347	AT1G54330.1	NAC domain-containing protein 86-like	3.636	4.14E-08
12	FCV25MIE_05204	AT4G03210.1	Xyloglucan endotransglucosylase/hydrolase protein 9	1.956	2.65E-07
13	FCV25MIE_02939	AT5G24660.1	Protein response to low sulfur 3-like	2.733	4.21E-07
14	FCV25MIE_32112	AT4G04610.1	5'-adenylylsulfate reductase 3, chloroplastic-like	2.107	8.52E-07
15	FCV25MIE_26820	AT4G27300.1	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290 isoform X1	1.475	1.85E-05
16	FCV25MIE_07483	AT4G10770.1	Oligopeptide transporter 7-like	1.815	4.90E-05
17	FCV25MIE_09916	AT3G09600.1	Protein reveille 8-like isoform X1	1.080	5.67E-05
18	FCV25MIE_03753	AT2G39650.1	Hypothetical protein	1.017	6.17E-05
19	FCV25MIE_32632	AT4G32551.1	Transcriptional corepressor leunig-like isoform X3	6.724	1.43E-04

Table S4 *Continued*

Rank	Locus_tag	Accession[a]	Product inferred from coding sequences[b]	Log_2(fold change)	FDR rate
20	FCV25MIE_01181	AT3G19000.1	Protein DMR6-like oxygenase 2-like isoform X2	2.548	1.60E-04
21	FCV25MIE_32505	AT1G61215.1	Hypothetical protein	1.409	1.60E-04
22	FCV25MIE_23256	AT1G02500.1	S-adenosylmethionine synthase	2.455	1.60E-04
23	FCV25MIE_31450	AT4G29210.1	Glutathione hydrolase 3	1.245	3.51E-04
24	FCV25MIE_00576	Not assigned	Auxin-responsive protein SAUR68-like	5.912	4.34E-04
25	FCV25MIE_01354	AT5G48850.1	Protein sulfur deficiency-induced 1-like	4.186	4.51E-04
26	FCV25MIE_28172	AT5G16990.1	2-alkenal reductase (NADP(+)-dependent)-like	1.712	5.56E-04
27	FCV25MIE_08617	AT1G23740.1	2-methylene-furan-3-one reductase-like	2.363	7.01E-04
28	FCV25MIE_20377	AT2G36430.1	UPF0481 protein At3g47200-like	1.898	7.22E-04
29	FCV25MIE_17904	AT2G38760.1	Annexin D3	1.228	7.22E-04
30	FCV25MIE_15569	AT5G10530.1	L-type lectin-domain containing receptor kinase IX.1-like	3.259	7.83E-04
31	FCV25MIE_14061	AT1G79680.1	Wall-associated receptor kinase-like 10 isoform X1	3.766	1.33E-03
32	FCV25MIE_22545	AT3G22890.1	ATP sulfurylase 1, chloroplastic-like isoform X1	1.730	1.53E-03
33	FCV25MIE_25776	AT4G23880.1	Hypothetical protein	4.238	1.53E-03
34	FCV25MIE_34198	AT2G23970.1	Gamma-glutamyl peptidase 5-like isoform X1	2.035	2.20E-03
35	FCV25MIE_30030	AT3G14050.1	Probable GTP diphosphokinase RSH2, chloroplastic	1.142	2.20E-03
36	FCV25MIE_01182	AT3G19000.1	Protein DMR6-like oxygenase 2-like	1.465	2.37E-03

321

322

Table S4 *Continued*

Rank	Locus_tag	Accession[a]	Product inferred from coding sequences[b]	Log_2(fold change)	FDR rate
37	FCV25MIE_08434	AT5G40670.1	Cystinosin homolog	1.344	2.41E-03
38	FCV25MIE_01173	AT3G19000.1	Protein DMR6-like oxygenase 2-like isoform X3	1.731	2.50E-03
39	FCV25MIE_15802	AT1G03670.1	Ankyrin repeat-containing protein	4.310	3.83E-03
40	FCV25MIE_09486	AT3G03900.1	Adenylyl-sulfate kinase 3	1.809	4.28E-03
41	FCV25MIE_28441	AT3G45940.1	Sulfoquinovosidase	3.290	4.29E-03
42	FCV25MIE_31974	AT1G12030.1	Hypothetical protein	3.086	5.69E-03
43	FCV25MIE_33838	AT3G19500.1	Transcription factor bHLH113	3.867	6.10E-03
44	FCV25MIE_22094	AT1G04770.1	Protein sulfur deficiency-induced 1-like	1.757	6.10E-03
45	FCV25MIE_03355	AT5G05950.1	Maternal effect embryo arrest	2.238	6.22E-03
46	FCV25MIE_21701	AT1G27880.1	ATP-dependent DNA helicase Q-like 5	2.958	7.90E-03
47	FCV25MIE_03748	AT2G39670.1	Putative dual-specificity rna methyltransferase rlmn	1.313	7.90E-03
48	FCV25MIE_19833	AT1G24100.1	UDP-glycosyltransferase 74B1-like	1.763	1.01E-02
49	FCV25MIE_28151	AT2G43850.3	Integrin-linked protein kinase 1-like isoform X1	1.770	1.15E-02
50	FCV25MIE_08995	AT3G51270.1	Serine/threonine-protein kinase rio2	1.034	1.28E-02
51	FCV25MIE_02242	AT4G24380.1	Esterase CG5412 isoform X1	1.839	1.28E-02
52	FCV25MIE_11803	AT4G28510.1	Prohibitin-1, mitochondrial-like	4.027	1.47E-02
53	FCV25MIE_02394	AT3G48990.1	Oxalate--CoA ligase	1.746	1.56E-02
54	FCV25MIE_27360	AT4G15550.1	Crocetin glucosyltransferase, chloroplastic-like	2.672	1.82E-02

Table S4 *Continued*

Rank	Locus_tag	Accession[a]	Product inferred from coding sequences[b]	Log_2(fold change)	FDR rate
55	FCV25MIE_03009	AT5G62350.1	21 kDa protein-like	1.585	1.84E-02
56	FCV25MIE_32676	AT2G39200.1	MLO-like protein 12	1.755	1.88E-02
57	FCV25MIE_34285	AT5G64210.1	Ubiquinol oxidase, mitochondrial	1.084	1.94E-02
58	FCV25MIE_23637	AT4G39230.1	Phenylcoumaran benzylic ether reductase Betv6	1.511	2.02E-02
59	FCV25MIE_15921	AT1G22360.2	7-deoxyloganetin glucosyltransferase-like	5.335	2.09E-02
60	FCV25MIE_01183	AT3G19000.1	Protein DMR6-like oxygenase 2-like	2.227	2.09E-02
61	FCV25MIE_08783	AT5G40010.1	AAA-atpase asd, mitochondrial	1.421	2.52E-02
62	FCV25MIE_13390	AT5G37490.1	U-box domain-containing protein 21-like	2.007	2.58E-02
63	FCV25MIE_30221	AT3G04710.1	Ankyrin repeat and protein kinase domain-containing protein 1	1.508	2.71E-02
64	FCV25MIE_24027	AT1G35210.1	Hypothetical protein	2.944	3.01E-02
65	FCV25MIE_27610	AT4G31130.1	Photosystem II CP43 reaction center protein	1.384	3.05E-02
66	FCV25MIE_25778	AT4G23880.1	Hypothetical protein	2.995	3.05E-02
67	FCV25MIE_20705	AT2G38100.1	Protein nrt1/ ptr family 5.5	1.975	3.07E-02
68	FCV25MIE_02950	AT5G17540.1	Benzyl alcohol O-benzoyltransferase-like	2.084	3.09E-02
69	FCV25MIE_09062	AT5G16250.1	Birch protein	3.856	3.15E-02
70	FCV25MIE_03290	AT5G17680.2	TMV resistance protein N-like	2.787	3.50E-02
71	FCV25MIE_30527	AT4G29890.1	Choline monooxygenase, chloroplastic	1.520	3.61E-02
72	FCV25MIE_34028	AT1G17870.1	Probable zinc metallopeptidase EGY3, chloroplastic	1.116	3.61E-02

Table S4 *Continued*

Rank	Locus_tag	Accession[a]	Product inferred from coding sequences[b]	Log_2(fold change)	FDR rate
73	FCV25MIE_09518	Not assigned	DNA ligase 1-like	1.068	3.69E-02
74	FCV25MIE_30481	AT1G06980.1	Hypothetical protein	3.521	3.69E-02
75	FCV25MIE_06553	AT2G22240.1	Inositol-3-phosphate synthase	1.536	3.76E-02
76	FCV25MIE_12792	AT5G25880.1	NADP-dependent malic enzyme isoform X2	1.456	3.88E-02
77	FCV25MIE_33628	AT4G15248.1	B-box zinc finger protein 32-like	1.306	3.92E-02
78	FCV25MIE_08996	Not assigned	Serine/threonine-protein kinase rio2	1.143	3.98E-02
79	FCV25MIE_17905	AT2G38750.1	Annexin D4	1.597	4.04E-02
80	FCV25MIE_20137	AT1G60190.1	U-box domain-containing protein 19-like	1.183	4.04E-02
81	FCV25MIE_07320	AT5G65530.1	Receptor-like cytosolic serine/threonine-protein kinase RBK1	1.416	4.08E-02
82	FCV25MIE_03576	AT3G56200.1	Amino acid transporter AVT6C-like	1.947	4.08E-02
83	FCV25MIE_00304	AT1G09800.1	tRNA pseudouridine synthase A-like isoform X1	1.163	4.12E-02
84	FCV25MIE_08991	AT3G51270.1	Serine/threonine-protein kinase rio2	1.639	4.13E-02
85	FCV25MIE_11820	AT2G48060.2	Piezo-type mechanosensitive ion channel homolog	1.278	4.20E-02
86	FCV25MIE_05224	AT5G54470.1	Zinc finger protein constans-like 12-like	1.300	4.25E-02
87	FCV25MIE_02729	AT5G38260.1	Rust resistance kinase Lr10-like isoform X1	1.447	4.36E-02
88	FCV25MIE_14564	AT1G16260.1	Wall-associated receptor kinase-like 8	2.423	4.60E-02
89	FCV25MIE_24871	AT3G07040.1	Disease resistance protein RPM1-like	1.942	4.67E-02

[a] Gene IDs from *Arabidopsis thaliana* that were homologous with protein sequences in the reference genome. "Not assigned" indicates that the gene ID could not be assigned because of an *E*-value $\geq 10^{-5}$ in the homology search

[b] Refers to the DDBJ annotation database, accession no. BKZX02000001-BKZX02000625

Table S5 Locus_tag list for down-regulated differentially expressed genes (DEGs) in *Fagus crenata* seedlings

Rank	Locus_tag	Accession[a]	Product inferred from coding sequences[b]	Log_2(fold change)	FDR rate
1	FCV25MIE_22317	AT5G46050.1	Protein NRT1/ PTR family 5.2-like	-8.037	1.30E-28
2	FCV25MIE_05404	AT5G36160.1	Tyrosine aminotransferase-like	-2.734	6.19E-13
3	FCV25MIE_30907	AT4G19810.1	Class V chitinase-like	-5.455	5.75E-12
4	FCV25MIE_05403	AT5G36160.1	Tyrosine aminotransferase-like	-2.757	3.60E-10
5	FCV25MIE_10525	Not assigned	PHAVU_004G153200g	-5.749	3.98E-07
6	FCV25MIE_17998	AT3G54390.1	Trihelix transcription factor ASIL2	-1.333	1.61E-04
7	FCV25MIE_17570	AT1G74170.1	Receptor-like protein 15	-3.409	1.66E-04
8	FCV25MIE_28759	AT1G78380.1	Probable glutathione S-transferase	-1.142	7.22E-04
9	FCV25MIE_26236	AT3G13110.1	Serine acetyltransferase 3, mitochondrial	-1.519	1.33E-03
10	FCV25MIE_07474	AT4G36040.1	Chaperone protein dnaJ 11, chloroplastic	-1.498	2.63E-03
11	FCV25MIE_18960	AT3G61490.1	Probable polygalacturonase	-1.907	5.02E-03
12	FCV25MIE_31512	AT5G56840.1	Transcription factor MYBS3-like	-2.080	5.24E-03
13	FCV25MIE_24121	AT4G33030.1	Udp-sulfoquinovose synthase, chloroplastic	-1.076	5.95E-03
14	FCV25MIE_26732	AT5G54590.2	Calcium/calmodulin-regulated receptor-like kinase 1	-1.538	6.10E-03
15	FCV25MIE_10987	AT3G11460.1	Putative pentatricopeptide repeat-containing protein At3g11460, mitochondrial	-2.025	6.62E-03
16	FCV25MIE_35051	AT3G55770.1	LIM domain-containing protein WLIM2b	-6.235	7.01E-03
17	FCV25MIE_05398	AT5G36160.1	Tyrosine aminotransferase-like	-2.179	7.52E-03
18	FCV25MIE_10568	AT1G73850.1	DUF1666 domain-containing protein	-1.074	1.04E-02
19	FCV25MIE_30245	AT1G05600.1	Pentatricopeptide repeat-containing protein At1g05600-like	-1.107	1.06E-02
20	FCV25MIE_12099	AT1G09070.1	Protein SRC2 homolog	-2.377	1.28E-02

Table S5 *Continued*

Rank	Locus_tag	Accession[a]	Product inferred from coding sequences[b]	Log_2(fold change)	FDR rate
21	FCV25MIE_19644	AT1G67110.1	Cytokinin hydroxylase-like	-1.824	1.28E-02
22	FCV25MIE_10524	AT3G18215.1	Hypothetical protein	-3.709	1.31E-02
23	FCV25MIE_13596	AT1G25440.1	Zinc finger protein constans-like 16-like	-1.241	1.41E-02
24	FCV25MIE_21452	AT5G50000.1	Serine/threonine-protein kinase STY13-like	-2.603	1.41E-02
25	FCV25MIE_30054	AT5G45540.1	Hypothetical protein	-2.347	1.46E-02
26	FCV25MIE_10708	AT2G21100.1	Dirigent protein 22-like	-3.832	1.56E-02
27	FCV25MIE_08328	AT3G01900.1	Cytochrome P450 94B3-like	-1.870	1.56E-02
28	FCV25MIE_21235	AT3G28960.1	Amino acid transporter AVT1I-like	-1.307	1.98E-02
29	FCV25MIE_24515	AT1G12090.1	14 kDa proline-rich protein DC2.15-like	-1.306	2.04E-02
30	FCV25MIE_02035	AT4G08300.1	WAT1-related protein At4g08300-like	-1.495	2.39E-02
31	FCV25MIE_14940	AT5G07050.1	WAT1-related protein At5g07050-like	-2.132	2.57E-02
32	FCV25MIE_06638	AT5G65280.1	LanC-like protein GCL1	-1.532	3.12E-02
33	FCV25MIE_02860	AT5G54180.1	Transcription termination factor MTERF8, chloroplastic	-1.343	3.19E-02
34	FCV25MIE_02797	AT5G53486.3	Protein PSY	-1.740	3.36E-02
35	FCV25MIE_05296	AT1G03600.1	Photosystem II repair protein PSB27-H1, chloroplastic	-1.100	3.58E-02
36	FCV25MIE_33811	AT2G15220.1	Basic secretory protease	-1.825	4.04E-02
37	FCV25MIE_12150	AT1G02860.1	Probable E3 ubiquitin-protein ligase BAH1-like 1 isoform X1	-2.543	4.56E-02
38	FCV25MIE_03654	AT5G24910.1	Cytochrome P450 714A1-like isoform X5	-2.574	4.65E-02

[a] Gene IDs for *Arabidopsis thaliana* that were homologous with protein sequences from the reference genome. "Not assigned" indicates that the gene ID could not be assigned because of an *E*-value ≥ 10^{-5} in the homology search

[b] Refers to the DDBJ annotation database, accession no. BKZX02000001-BKZX02000625

Table S6 Gene ontology (GO) terms present in the differentially expressed genes (DEGs) up-regulated by drought stress in *Fagus crenata* seedlings

ID[a]	Category	Description	Log_{10}(*p*-value[b])	Log_{10}(FDR[c])	Test group[d]	Reference group[e]	Group ID
GO:0010438	Biological Processes	Cellular response to sulfur starvation	-6.607	-2.659	3/70	3/10996	1
GO:0071496	Biological Processes	Cellular response to external stimulus	-3.590	-0.244	6/70	140/10996	1
GO:0031668	Biological Processes	Cellular response to extracellular stimulus	-2.810	0.000	5/70	133/10996	1
GO:0009267	Biological Processes	Cellular response to starvation	-2.666	0.000	4/70	86/10996	1
GO:0031669	Biological Processes	Cellular response to nutrient levels	-2.413	0.000	4/70	101/10996	1
GO:0042594	Biological Processes	Response to starvation	-2.240	0.000	4/70	113/10996	1
GO:0062012	Biological Processes	Regulation of small molecule metabolic process	-2.240	0.000	4/70	113/10996	1
GO:0009991	Biological Processes	Response to extracellular stimulus	-2.204	0.000	5/70	184/10996	1
GO:0000103	Biological Processes	Sulfate assimilation	-5.071	-1.424	3/70	7/10996	2
GO:0006790	Biological Processes	Sulfur compound metabolic process	-2.690	0.000	7/70	282/10996	2
GO:0030247	Molecular Functions	Polysaccharide binding	-3.590	-0.244	4/70	49/10996	3
GO:0030246	Molecular Functions	Carbohydrate binding	-3.273	-0.024	5/70	105/10996	3
GO:0002239	Biological Processes	Response to oomycetes	-2.750	0.000	3/70	38/10996	4
GO:0045087	Biological Processes	Innate immune response	-2.240	0.000	4/70	113/10996	4
GO:0006955	Biological Processes	Immune response	-2.099	0.000	4/70	124/10996	4
GO:0046686	Biological Processes	Response to cadmium ion	-2.624	0.000	3/70	42/10996	5
GO:0004674	Molecular Functions	Protein serine/threonine kinase activity	-2.129	0.000	7/70	357/10996	6

[a] Bold letters indicate representative GO terms with the same group ID

[b] Based on the cumulative hypergeometric distribution

[c] Based on the Benjamini-Hochberg procedure to account for multiple testing when a given *Q* number of GO terms was typically identified in the reference

genome, as follows. First, all *p*-values were sorted from small to large. Second, given a *p*-value of *p* at rank *i*, it would be expected that *pQ* GO terms could be found with the same or a better *p*-value by chance under a Bonferroni correction. As we only observed *i* such GO terms, the portion of our observations that were false (i.e. FDR) was min(*pQ*/*i*, 1).

[d] The number of DEGs annotated by the corresponding GO term/the number of all DEGs annotated by GO terms.

[e] The number of genes annotated by the corresponding GO term in the reference genome/the number of all genes annotated by GO terms in the reference genome.

333

334

Table S7 Gene ontology (GO) terms present in the differentially expressed genes (DEGs) down-regulated by drought stress in *Fagus crenata* seedlings

ID[a]	Category	Description	Log_{10}(*p*-value[b])	Log_{10}(FDR[c])	Test group[d]	Reference group[e]	Group ID
GO:0009072	Biological Processes	Aromatic amino acid metabolic process	-3.088	0.000	3/28	73/10996	1
GO:0009308	Biological Processes	Amine metabolic process	-2.609	0.000	3/28	107/10996	1
GO:1901605	Biological Processes	Alpha-amino acid metabolic process	-2.603	0.000	4/28	228/10996	1
GO:0006520	Biological Processes	Amino acid metabolic process	-2.003	0.000	4/28	337/10996	1
GO:0004497	Molecular Functions	Monooxygenase activity	-2.957	0.000	3/28	81/10996	2
GO:0005506	Molecular Functions	Iron ion binding	-2.352	0.000	3/28	132/10996	2
GO:0020037	Molecular Functions	Heme binding	-2.255	0.000	3/28	143/10996	2
GO:0016705	Molecular Functions	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	-2.113	0.000	3/28	161/10996	2
GO:0046906	Molecular Functions	Tetrapyrrole binding	-2.113	0.000	3/28	161/10996	2
GO:0009753	Biological Processes	Response to jasmonic acid	-2.824	0.000	5/28	340/10996	3
GO:0070542	Biological Processes	Response to fatty acid	-2.736	0.000	5/28	356/10996	3
GO:0008270	Molecular Functions	Zinc ion binding	-2.424	0.000	5/28	420/10996	4
GO:0031668	Biological Processes	Cellular response to extracellular stimulus	-2.343	0.000	3/28	133/10996	5
GO:0071496	Biological Processes	Cellular response to external stimulus	-2.280	0.000	3/28	140/10996	5

[a] Bold letters indicate representative GO terms with the same group ID

[b] Based on the cumulative hypergeometric distribution

[c] Based on the Benjamini-Hochberg procedure to account for multiple testing when a given *Q* number of GO terms was typically identified in the reference genome, as follows. First, all *p*-values were sorted from small to large. Second, given a *p*-value of *p* at rank *i*, it would be expected that *pQ* GO terms could be found with the same or a better *p*-value by chance under a Bonferroni correction. As we only observed *i* such GO terms, the portion of

our observations that were false (i.e. FDR) was min(pQ/i, 1).

[d] The number of DEGs annotated by the corresponding GO term/the number of all DEGs annotated by GO terms.

[e] The number of genes annotated by the corresponding GO term in the reference genome/the number of all genes annotated by GO terms in the reference genome.

335

336

Table S8 KEGG pathway present in the differentially expressed genes (DEGs) up-regulated by drought stress in *Fagus crenata* seedlings

ID	Description	Log_{10}(*p*-value[a])	Log_{10}(FDR[b])	Test group[c]	Reference group[d]	Group ID
ath00920	Sulfur metabolism	-3.289	-1.140	3/70	25/10996	1
ath00230	Purine metabolism	-2.018	-0.170	3/70	69/10996	2

[a] Based on the cumulative hypergeometric distribution

[b] Based on the Benjamini-Hochberg procedure to account for multiple testing when a given *Q* number of KEGG pathways was typically identified in the reference genome, as follows. First, all *p*-values were sorted from small to large. Second, given a *p*-value of *p* at rank *i*, it would be expected that *pQ* KEGG pathways could be found with the same or a better *p*-value by chance under a Bonferroni correction. As we only observed *i* such KEGG pathways, the portion of our observations that were false (i.e. FDR) was min(*pQ*/*i*, 1).

[c] The number of DEGs annotated by the corresponding KEGG pathway/the number of all DEGs annotated by KEGG pathways.

[d] The number of genes annotated by the corresponding KEGG pathway in the reference genome/the number of all genes annotated by KEGG pathways in the reference genome.

337