



OPEN

DATA DESCRIPTOR

# *Ulmus minor* response to Dutch elm disease: *de novo* transcriptome assembly and annotation

V. Chano<sup>1,2</sup>✉, J. Sobrino-Plata<sup>1,3</sup>, C. Martínez-Arias<sup>1,4</sup>, C. Collada<sup>1</sup>, J. Rodríguez-Calcerrada<sup>1</sup> & J. A. Martín<sup>1,5</sup>✉

Dutch elm disease (DED), caused by *Ophiostoma novo-ulmi* (ONU), has devastated elm populations across Europe and North America since the 20<sup>th</sup> century. In this work, a *de novo* transcriptome assembly of *Ulmus minor* in response to ONU is presented. We used two DED-resistant genotypes, MDV2.3 and VAD2, and one DED-susceptible genotype, MDV1, to capture responses to ONU at four time points post-inoculation (6, 24, 72, and 144 hours). RNA from collected samples was isolated and sequenced producing 60.88 M 100 bp paired-end reads per sample. We performed a *de novo* transcriptome assembly combining data from the three genotypes. The assembly was functionally annotated and validated through differential gene expression analysis of the response. This dataset provides a valuable resource for studying molecular mechanisms of DED resistance in elms, contributing to broadening our understanding of tree immunity and facilitating potential applications in functional annotation of future genome assemblies.

## Background

The molecular defensive mechanisms and plant immune system have been widely studied in the plant model species *Arabidopsis thaliana*, together with other crop model species of great agronomic importance such as *Vitis vinifera*, *Oryza sativa* or *Solanum* sp.<sup>1,2</sup>. However, the scarcity of model species in forestry has hampered research on tree immune responses, while new climate conditions, together with the emergence of invasive pests and diseases, have evidenced the importance of forest health. Currently, improved and more cost-effective sequencing methods have enabled advances in tree breeding through genomic-assisted selection strategies, where a deeper understanding of the role of candidate genes in defensive pathways is crucial<sup>3</sup>. Recent efforts in transcriptomics have tried to close this gap providing evidence of tree immunity in both gymnosperms<sup>4–7</sup> and angiosperms<sup>8–12</sup>.

One of the first reported cases of invasive alien diseases in forest species was the Dutch elm disease (DED), which has devastated elm populations in Europe and North America<sup>13</sup>. DED was initially caused by the vascular pathogen *Ophiostoma ulmi* during the first half of the 20<sup>th</sup> century, later replaced by the more aggressive species *Ophiostoma novo-ulmi* (ONU), which is still very aggressive on the remaining elm populations. The pathogen is effectively transmitted by elm bark beetles of the genus *Scolytus* and *Hylurgopinus* during maturation feeding<sup>14</sup>. *Ulmus minor* in Europe and *Ulmus americana* in North America have been the most severely affected elm species, as nearly 100% of the populations of these two species are highly susceptible to ONU<sup>13,15,16</sup>.

Increasing the resistance to DED has been one of the primary objectives of different breeding programs<sup>17</sup>. This has been achieved by crossbreeding European elms with the more resistant Asian elm species, or through the selection and crossing of resistant native elms. This last strategy is usually preferred by environmental authorities to protect autochthonous genetic resources in reforestation<sup>18</sup>. The Spanish elm breeding program

<sup>1</sup>Departamento de Sistemas y Recursos Naturales, ETSI Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid. José Antonio Novais 10, 28040, Madrid, Spain. <sup>2</sup>Department of Forest Genetics and Forest Tree Breeding, Georg-August-Universität Göttingen. Büsgenweg 2, 37077, Göttingen, Germany. <sup>3</sup>Departamento de Genética, Fisiología y Microbiología, Facultad de Ciencias Biológicas, Universidad Complutense de Madrid. José Antonio Novais 12, 28040, Madrid, Spain. <sup>4</sup>Departamento de Biotecnología-Biología Vegetal, ETSI Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Avda. Puerta Hierro 2-4, 28040, Madrid, Spain. <sup>5</sup>Centro de I + D + i para la Conservación de la Biodiversidad y el Desarrollo Sostenible (CBDS), Universidad Politécnica de Madrid. José Antonio Novais 10, 28040, Madrid, Spain. ✉e-mail: [victor.chano@uni-goettingen.de](mailto:victor.chano@uni-goettingen.de); [juan.martin.garcia@upm.es](mailto:juan.martin.garcia@upm.es)

selected five native DED-resistant *U. minor* genotypes, which were registered as forest reproductive material<sup>19</sup>. These registered cultivars have opened new prospects to decipher and understand the molecular basis of DED resistance and elm immunity, which are crucial to restore elm forests. Transcriptomics and differential gene expression analysis have already been conducted on DED research. For instance, Perdiguero *et al.*<sup>20</sup> provided a DED-sensitive reference transcriptome, with 58,429 unigenes (*de novo* assembly) from almost 1 M of high-quality reads and the identification of 22,500 single nucleotide polymorphisms (SNPs) among three *U. minor* genotypes (one susceptible and two resistant). This genetic resource was used for the design of a cDNA-microarray for later performing a time-course gene expression analysis of the DED-susceptible elm cultivar 'Atinia' (*U. minor* var. *vulgaris*<sup>21</sup>), in response to ONU<sup>22</sup>. More recently, Islam *et al.*<sup>10</sup> described the transcriptomic responses of a susceptible and a resistant genotype of *U. americana* to ONU, providing 5,329 Differentially Expressed Genes (DEGs) in response to the infection. In this work, we aimed to create a reference transcriptome for elm through a time-course design that captures the local and systemic responses at four time points after *U. minor* inoculation with ONU. To this end, we used two resistant Spanish genotypes, MDV2.3 and VAD2, which show differences in their response to the disease<sup>23</sup>, along with the susceptible Spanish genotype MDV1.

## Material and Methods

**Plant material and experimental design.** Three *U. minor* genotypes were used in this work. Two of them were registered in 2014 as Spanish forest reproductive material under the names "Dehesa de Amaniël" (hereinafter referred to as "MDV2.3", from Dehesa de la Villa, Madrid, Spain) and "Ademuz" (VAD2 hereinafter, from Ademuz, Valencia, Spain). MDV2.3 and VAD2 are highly resistant to DED<sup>19</sup>. The third *U. minor* genotype used was MDV1 (from Dehesa de la Villa, Madrid, Spain), which is highly susceptible to the disease<sup>19</sup>. Buds from the three genotypes were used as starting material for *in vitro* micropropagation in growing chambers according to Conde *et al.*<sup>24</sup>. One-month-old plantlets were transplanted to 500 mL pots filled with peat (Kekkilä®, Vantaa, Finland) and perlite in volumetric proportions 3:1 as substrate. The pots were placed in a growth chamber with 16 h photoperiod using fluorescent white light with approximately 240  $\mu\text{mol m}^{-2} \text{s}^{-1}$  photosynthetically active radiation at 24/18 °C day/night temperature and 60% air relative humidity. Plants were regularly watered with filtrated tap water for one month.

In planta inoculations were performed on 2-month-old plants using a ONU bud-cell water suspension at a concentration of  $10^7$  blastospores  $\text{ml}^{-1}$ . Pathogen spores were delivered into the xylem sap stream through a transverse cut made with a sharp blade in the base of the stem<sup>25</sup>. To obtain the ONU blastospores, mycelial plugs from the *O. novo-ulmi* ssp. *americana* isolate Z-BU1 were grown in Erlenmeyer flasks containing Tchernoff's liquid medium<sup>26</sup> at 22 °C in the dark under constant shaking to induce sporulation. Three days later, blastospores were collected by centrifugation and the concentration adjusted to  $10^7$  blastospores  $\text{ml}^{-1}$  using a haemocytometer. A total of 32 plants from each genotype were used, half of them inoculated with 10  $\mu\text{l}$  of the bud-cell suspension *O. novo-ulmi* ssp. *americana* isolate Z-BU1<sup>17</sup>, and the other half mock-inoculated with 10  $\mu\text{l}$  of sterilized distilled water. Samples were collected following a time series with four sampling points: 6 hours post inoculation (hpi hereinafter), 24 hpi, 72 hpi and 144 hpi. Four clonal replicates (ramets) were used as biological replicates for each experimental group. Additionally, two stem segments were collected per plant at each sampling time: one 4-cm-long segment including the inoculation point (local response) and a similar segment taken 10 cm above the inoculation point (systemic response). Plant material was immediately frozen in liquid nitrogen after collection, grinded into a fine powder using a Retsch® mill and sterilized stainless steel balls and stored at -80 °C until use.

**RNA isolation and sequencing.** Total RNA was extracted from the 192 samples using a Plant/Fungi total RNA Purification Kit (Norgen Biotek Corp., Thorold, Ontario), following the instructions from the manufacturer. The protocol included a DNase treatment step for removal of residual DNA with the Norgen RNase-free DNase I Kit (Norgen Biotek Corp., Thorold, Ontario). The integrity of isolated RNA was assessed using a 2100 Bioanalyzer (Agilent Technologies). Approximately 40  $\mu\text{l}$  of total RNA per sample were submitted to MacroGen (MacroGen, The Netherlands), and 192 TruSeq mRNA stranded libraries were prepared using polyA selection. One hundred bp paired-end reads were generated with Illumina HiSeq 3000, with a sequencing depth of 50 million reads per sample. Samples were multiplexed in 4 lanes with 6 samples per lane. In order to avoid technical biases due to sample position in the sequence, biological replicates for each sample were randomly distributed among lanes. Reads were demultiplexed based on barcodes included during library preparation.

**Bioinformatic processing of sequencing data.** The bioinformatic pipelines were run on a Linux Ubuntu v18.04 workstation featuring an Intel Xeon Gold 6130 2.1 GHz processor with 32 cores (64 threads) and 256 GB RAM. Storage included a 2TB SSD (35200 MBps reading and 2500 MBps writing speed) and a Redundant Array of Inexpensive Disks (RAID) setup with 4 TB HDDs (7200 rpm, 12 Gb/s). The home system (SSD) hosted the software, while datasets and databases were stored in the RAID.

Initial quality control of raw reads was performed by using FastQC v0.11.7 and MultiQC<sup>27</sup>. We then used Trimmomatic 0.36<sup>28</sup> with a Phred Score below 30, excluding reads with a minimal length of 30 nucleotides, and using a sliding window (nucleotide window size) of 4. We also cut 12 nucleotides from the beginning of reads using the headcrop option and removed possible adapter sequences when present. Finally, the improvement of quality of trimmed and filtered reads was again assessed with FastQC and MultiQC. By using HISAT2<sup>29</sup>, processed fastq files were mapped against the genome available for the ONU reference strain H327<sup>30</sup> (downloaded from the DOE Joint Genome Institute, <https://mycocosm.jgi.doe.gov/Ophnu1/Ophnu1.home.html>) to filter out reads from the pathogen. First, ONU genome was indexed using the routine hisat2-build, and mapping was performed with the option -score-min L,0,-0.6. Mapped reads were saved as bam files and stored for further analysis (data not shown). Unaligned reads were saved as fastq files and mapped against a custom database created ad

hoc and including sequences from bacteria, mostly rRNA downloaded from SILVA (<https://www.arb-silva.de/>), and from endophytes putatively present in plant samples from the FungDB (<https://fungidb.org/fungidb/app>). Those sequences that did not map during any of these mapping steps were considered as *de novo* transcriptome assembly and validation analyses.

**De novo transcriptome assembly and functional annotation.** Reads were *de novo* assembled using Trinity v2.14.0<sup>31</sup>, indicating a kmer size of 32 for the de Bruijn graph construction, which was selected after running Kmergenie v1.7051<sup>32</sup>. Due to the high number of initial reads and, consequently, the significant computational requirements, assembly was first performed individually for each genotype (MDV1, MDV2.3 or VAD2). Afterwards, a single transcriptome was meta-assembled with these three genotype-transcriptomes together with reads from a 454-pyrosequencing conducted in a previous work<sup>19</sup>, and corresponding to the moderately resistant genotypes MDV4.5 (from Dehesa de la Villa, Madrid, Spain) and JCA2 (from Cazorla, Jaén, Spain), and the susceptible TOAL1 (from Algodor, Toledo, Spain). Unlike Illumina sequencing, which results in numerous short-reads, 454-pyrosequencing allows obtaining longer reads that may be useful to fill possible gaps and reduce redundancies. The sra file from this 454-pyrosequencing was first downloaded from NCBI Short Read Archive, accession number SRR1687227, and then converted into fastq by using prefetch and fastq-dump routines, respectively, from SRA-Toolkit v2.10.2 (<https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software>). This fastq file was also processed with Trimmomatic and evaluated with FastQC, as previously indicated for Illumina reads. Compared to individual transcriptomes of the three genotypes, the number of transcripts was highly reduced in the single transcriptome, although it still contained many redundancies as isoforms. Therefore, we included a final step to identify so-called unigenes and generate a single reference transcriptome by using the perl-based tool get\_longest\_isoform\_seq\_per\_trinity\_gene.pl. Contigs included in the final transcriptome assembly were first renamed following the alphabetical order given by Trinity.

Transcripts were annotated using a multi-fasta file as input for local version of BLASTx<sup>33</sup> using TOA v0.66<sup>34</sup> and the RefSeq Viridiplantae section. Parameters selected for BLASTx were an e-value threshold of  $10^{-5}$  and a maximum of 50 hits per query (transcript). The output files (xml format) were then uploaded into the Functional Analysis module (formerly known as Blast2GO) from Omicsbox v4.0<sup>35</sup>. Mapping and Annotation tools from Omicsbox were used for the assignment of Gene Ontology (GO) terms associated to these transcripts, and InterPro scan was also used to improve annotation, by retrieving GOs associated with protein motifs/domains. The GO-Slim tool was then used to filter and reduce these GO terms to those associated with plants. Moreover, information about Enzyme Codes (EC) and pathways from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) was also gathered by means of Omicsbox.

**Counting and differential gene expression analyses for transcriptome validation.** Local samples (i.e. taken at the inoculation point) of ONU-inoculated and mock-inoculated plants from the three genotypes were used for the validation of the transcriptome assembly. The pseudo-aligner Kallisto v0.46<sup>36</sup> was used to index the reference transcriptome of *U. minor* and generate raw counts for all the transcripts per sample library, creating a final count matrix later imported in R for differential gene expression analysis using DESeq2 v1.30<sup>37</sup>. Prior to differential gene expression analysis, counts were normalized using the vst (variance stabilizing transformation) function, and exploratory analyses of normalized gene expression values were carried out by Principal Component Analysis (PCA) and sample-to-sample distances implemented in DESeq2.

Differential gene expression analyses were performed for each genotype and single time point by using Wald's test to compare ONU-inoculated and mock-inoculated samples (including four biological replicates each). A logarithmic Fold Change (LFC hereinafter) value was obtained for each transcript, calculated as the logarithm of the expression ratio between ONU-inoculated and mock-inoculated samples. Therefore, the LFC value for a given transcript resulted in a positive number when the expression was higher in ONU-inoculated plants (referred hereinafter as overexpression, up-regulation or induction) whereas a negative number (referred hereinafter as underexpression, down-regulation or repression) indicated a higher expression in mock-inoculated plants. The model formula for the Wald's test in DESeq2 was indicated as "design = ~ treatment", and differentially expressed genes (DEGs) were selected when the p-value (adjusted for multiple comparisons by False Discovery Rate, FDR<sup>38</sup>) was lower than 0.05 and the LFC greater than the absolute value of 2 (adj p-value < 0.05 and LFC > |2|). Resulting time-specific sets of DEGs were compared among genotypes by creating an upset plot with the R package upsetR v1.4.0<sup>39</sup>.

## Data Records

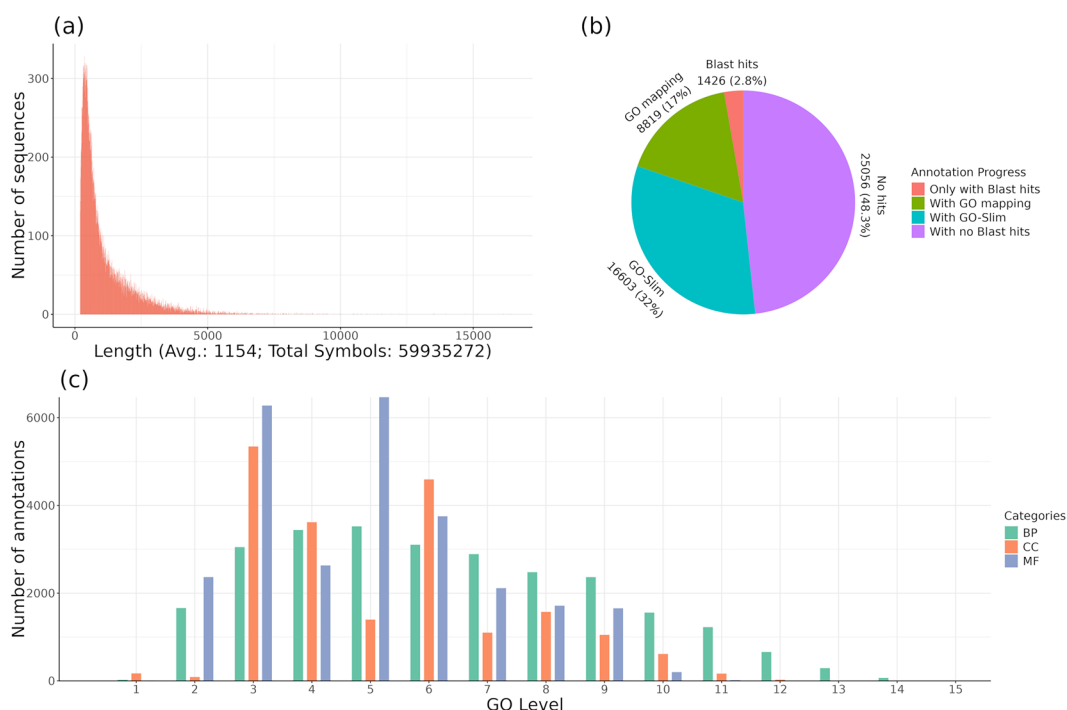
This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GLCK00000000<sup>40</sup>. The version described in this paper is the first version (GLCK00000000.1). Supplementary Tables and multi-fasta file with transcriptome assembly were deposited in Figshare under a CC-BY license<sup>41</sup>. Raw sequencing data (fastq files) of local and distal samples of *U. minor*, ONU-inoculated and mock-inoculated along a time-course (6-, 24-, 72- and 144 hpi) were deposited at the National Center for Biotechnology Information (NCBI) under the BioProject accession no. PRJNA1226172, and the Sequence Read Archive (SRA) accession no. SRP566393<sup>42</sup>.

## Technical Validation

In this work, we performed a *de novo* transcriptome assembly of the *U. minor* response to DED. Based on the initial 11,689.4 M paired-end reads (60.88 M paired-end reads per sample), a first transcriptome pre-assembly for each genotype was performed. These pre-assemblies yielded 228,642 contigs for the DED-susceptible MDV1, 233,398 contigs for the DED-resistant MDV2.3, and 242,861 contigs for the DED-resistant VAD2. The three *de novo* primary transcriptomes were then meta-assembled together with 971,002 reads (350 bp length) derived from a

Parameters	Preliminary assembly	Final assembly
No. of contigs	74,593	51,507
No. contigs > 500 bp	54,995	35,324
No. contigs > 1000 bp	33,658	19,495
Total length (bp)	96 464,235	59 715,254
Largest contig (bp)	16,407	16,407
Average contig length (bp)	1,293.21	1,159.36
Median contig length (bp)	881	733
GC content (%)	38.74	38.27
N50 contig size (bp)	1,961	1,810
L50 (contig no.)	15,439	9,996

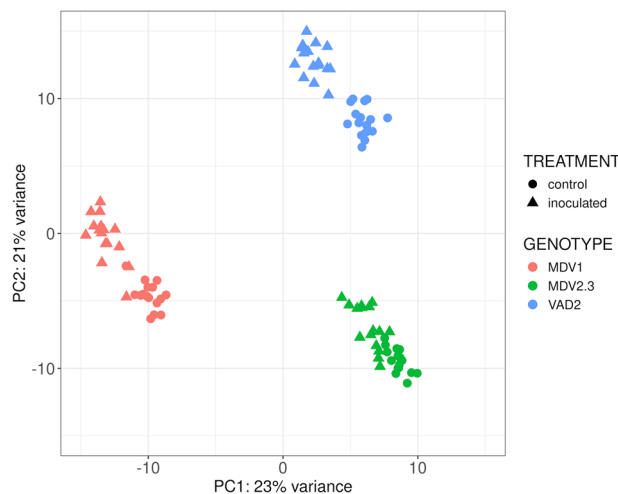
**Table 1.** Summary of transcriptome assembly quality parameters. bp: base pairs; no.: number; GC: guanine-cytosine; N50: sequence length of the shortest contig at 50% of total assembly length; L50: count of smallest number of contigs with sum length over 50% of total assembly length.



**Fig. 1** Final transcriptome assembly. (a) Size distribution of unigenes after generating a single reference transcriptome. (b) Functional annotation progress (c) Distribution of Gene Ontology (GO) levels for the three categories Biological Process (BP), Cellular Component (CC) and Molecular Function (MF).

454-pyrosequencing previously performed for *U. minor*<sup>20</sup> into a preliminary single transcriptome assembly, resulting in 74,593 contigs (including isoforms). The resulting quality parameters were: 54,995 contigs over 500 bp and 33,658 over 1,000 bp, 96 464,235 bp total length, 1,293.21 and 881 bp average and median contig length, respectively, 38.74% GC content, and N50 and L50 values of 1,961 bp and 15,439, respectively (Table 1). The longer transcripts per isoform were identified, reducing redundancies and obtaining a final transcriptome assembly that included 51,507 unigenes<sup>40</sup>. Assembly quality parameters were: 35,324 and 19,495 contigs over 500 and 1,000 bp, respectively, 59 715,254 bp total length, 1,159.36 and 733 bp average and median contig length, respectively, 38.27% GC content, and N50 and L50 values of 1,810 bp and 9,996, respectively (Table 1). Top5 contigs with maximum length were Trinity\_DN1091\_c0\_g1\_i4 (16,407 bp), Trinity\_DN9258\_c0\_g1\_i2 (16,140 bp), Trinity\_DN1582\_c0\_g1\_i1 (15,135 bp), Trinity\_DN1263\_c0\_g2\_i1 (14,705 bp) and Trinity\_DN25886\_c0\_g1\_i1 (13,962 bp). The distribution of the number of sequences by sequence length is shown in Fig. 1a. In Perdiguerio *et al.*<sup>20</sup>, the previous transcriptome resulted in 58,429 unigenes, a slightly higher but very similar number compared to the current work. Similar numbers were also obtained in a more recent study on DED affecting *U. americana*<sup>10</sup>, with 55,499 unigenes identified.

**Functional annotation of *Ulmus minor* transcriptome.** The functional annotation of the final list of unigenes<sup>40</sup> is presented in Supplementary Table S1. From the initial 51,507 sequences, 26,566 (51.6%) matched



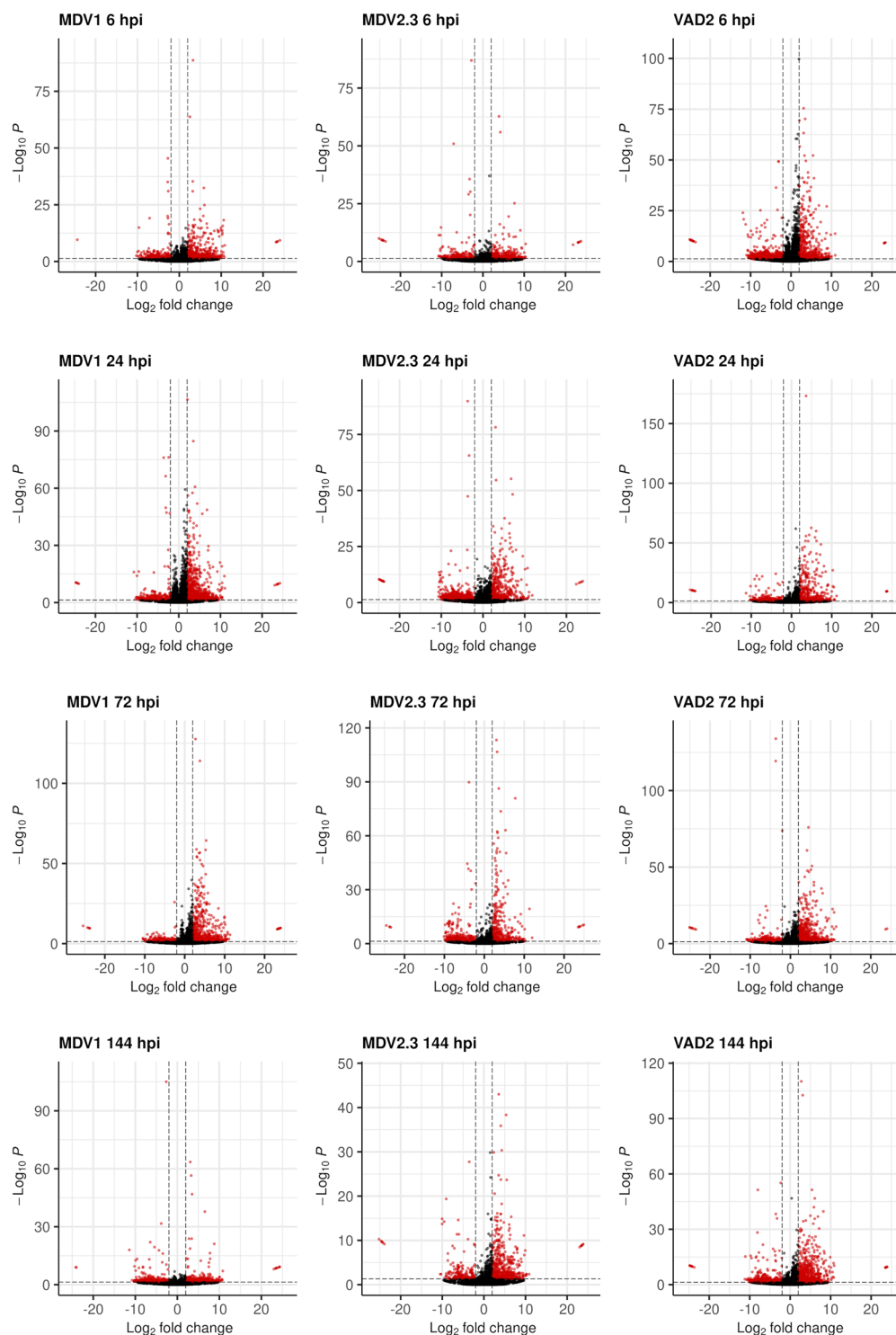
**Fig. 2** Principal Component Analysis (PCA) of transcriptomic local profiles from three *Ulmus minor* genotypes (MDV1, MDV2.3, and VAD2), following inoculation with *Ophiostoma novo-ulmi* or mock-inoculation with sterile water (control). Expression values were normalized by *vst* (variance stabilizing transformation).

to protein records in RefSeq Viridiplantae (Fig. 1b). From these, 1,341 sequences (2.6%) did not yield GO terms associated, while 8,768 sequences (17%) were associated with GOs and 16,457 (32%) passed the GO-Slim step. The percentage of matched sequences was higher than the 42% obtained for *U. americana*<sup>10</sup>. In addition, 24,461 sequences were matched after InterPro Scan, retrieving GOs from 12,473 of them. Thus, the absolute number of identified GO terms was 73,799 distributed throughout three categories, Cellular Component (CC), Molecular Function (MF), and Biological Process (BP), classified up to 15 levels of specificity (Fig. 1c). A direct count within these three categories (data not shown) revealed that the top-five GO terms were: CC “membrane”, “nucleus”, “cytoplasm”, “plasma membrane” and “chloroplast”; MF “protein binding”, “nucleotide binding”, “hydrolase activity”, “catalytic activity” and “transferase activity”; and BP “biosynthetic process”, “nucleobase containing compound metabolic process”, “cellular protein modification process”, “transport” and “cellular process”.

**Differential gene expression analysis: ONU-inoculated vs. mock-inoculated samples.** The gene expression matrix was imported in R and total count values were normalized by means of the *vst* (variance stabilizing transformation) method. Normalized expression values were then inspected through a PCA (Fig. 2). The three genotypes were clearly separated by the first component (PC1; 23% of the variance), with a greater variance between the DED-susceptible MDV1 and the two DED-resistant genotypes MDV2.3 and VAD2. Interestingly, the second component (PC2, 21% of the variance) separated the susceptible MDV1 and the resistant MDV2.3, both from the same origin (Dehesa de la Villa, Madrid), from the resistant VAD2, with a different origin (Ademuz, Valencia). In addition, the two sub-groups corresponding to the inoculation treatment (ONU-inoculated and mock-inoculated samples) were distinguished within each genotype. For the DED-susceptible MDV1 (Supplementary Fig. S1a), ONU-inoculated and mock-inoculated trees were separated by the first component (29% of the variance), while the second component separated the different collecting-points across the time-course (24% of the variance). For the DED-resistant MDV2.3 (Supplementary Fig. S1b), the first component (22% of the variance) separated the time points, while the second component (13% of the variance) separated ONU-inoculated from mock-inoculated samples. Similarly to MDV1, ONU-inoculated and mock-inoculated VAD2 trees were also separated by the first component (24% of the variance), while separation of time points was observed in the second component (18%) (Supplementary Fig. S1c). These PCAs revealed distinct responses among genotypes. Additionally, the results demonstrated the high effectiveness of the inoculation method across all individuals, which also exhibited time-dependent responses. Similar responses were also observed in other tree diseases like ash dieback<sup>11,12</sup>, pine pitch canker in *Pinus pinaster* and *Pinus pinea*<sup>5,7</sup>, or needle bladder rust in Norway spruce<sup>6</sup>.

Validation of the *de novo* transcriptome was performed by using this final assembly as reference for the differential gene expression analysis of the local responses to DED for each genotype and sampling time across the time-course experiment. The distribution of DEGs for the three genotypes and four time points is depicted in Fig. 3. DEGs (red dots) were considered when adjusted p-values < 0.05 and LFC > |2|. Overall, the number of up-regulated and down-regulated genes in response to the infection differed between genotypes and time points, showing different profiles not only between DED-susceptible and DED-resistant genotypes but also between the two DED-resistant genotypes (Fig. 4). The number of induced transcripts in the susceptible genotype MDV1 increased from 6 to 24 hpi, later decreasing, while the number of repressed genes slightly varied during the response (Supplementary Tables S2–S5). Similarly, the number of induced genes in the resistant genotype MDV2.3 also increased during the response from 6 to 72 hpi, decreasing only at 144 hpi (Supplementary Tables S6–S9). In addition, the number of repressed transcripts increased from 6 to 24 hpi, later decreasing (for clarity, the number of down-regulated genes in Fig. 4 are represented as negative numbers). In contrast, the resistant VAD2 showed a different response, with more genes being induced at 6 hpi than at 24 hpi, later increased at 72 hpi and decreased at 144 hpi. Moreover, the profile of repressed genes during the time-course

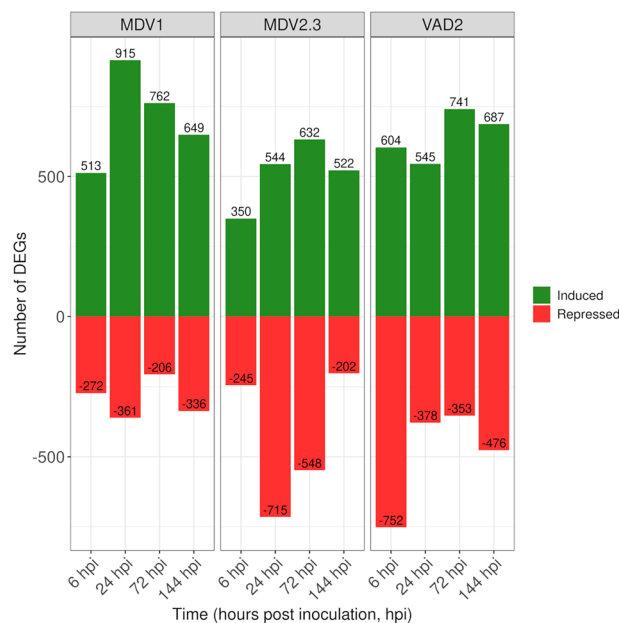




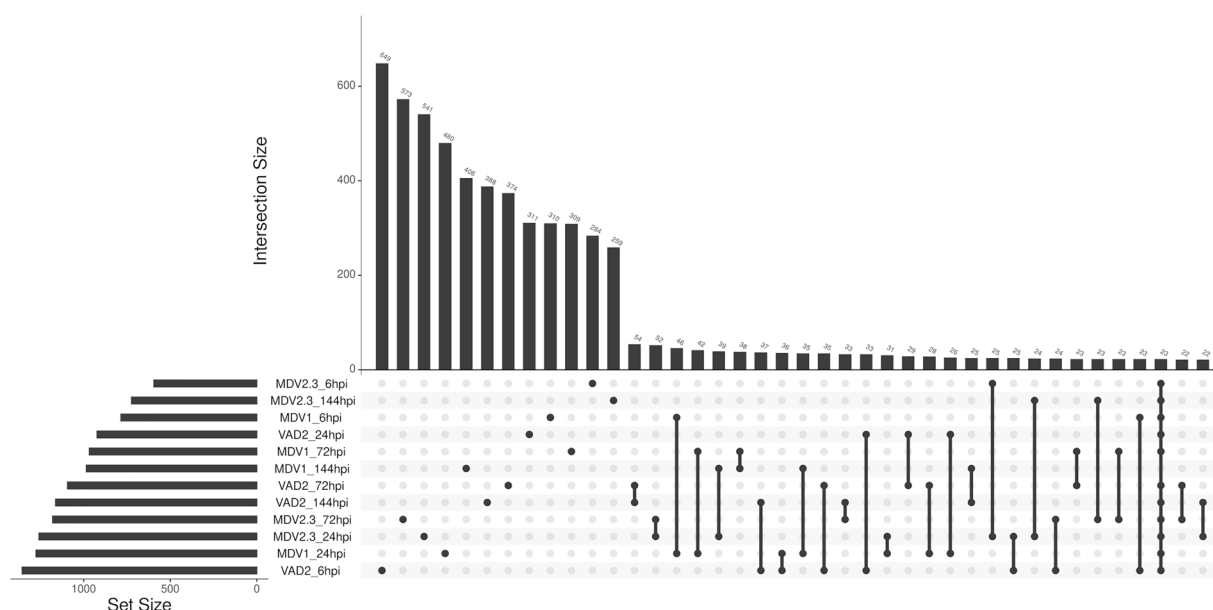
**Fig. 3** Genotype- and time-specific volcano plots showing differentially expressed genes (DEGs, red dots) between *Ophiostoma novo-ulmi*-inoculated and mock-inoculated samples across four time points (6-, 24-, 72-, and 144-hours post-inoculation, hpi) for each *Ulmus minor* genotype (MDV1, MDV2.3, and VAD2). Dashed lines represent significance thresholds (FDR-adjusted  $P < 0.05$ , horizontal) and expression cutoffs ( $\log_2$  fold change  $\geq |2|$ , vertical).

was also different than in MDV1 and MDV2.3, with the highest number of down-regulated genes found at 6 hpi, decreasing later and kept in similar values during the rest of the analysed response (Supplementary Tables S10–S13).

In summary, a first inducible local response to the infection was found for the three genotypes, characterized by a remarkably higher number of up-regulated than down-regulated genes, particularly in the susceptible



**Fig. 4** Bar plots displaying the number of differentially expressed genes (DEGs) per time point (6-, 24-, 72-, and 144-hours post-inoculation, hpi) and *Ulmus minor* genotype (MDV1, MDV2.3, and VAD2). For clarity, upregulated genes are shown as positive values, while downregulated genes are shown as negative values.



**Fig. 5** UpSet plot illustrating shared and exclusive differentially expressed genes (DEGs) across genotypes (MDV1, MDV2.3 and VAD2) and time points (6-, 24-, 72-, and 144-hours post-inoculation, hpi). For visualization purposes, the number of intersections is limited to a maximum of 40 DEG sets.

genotype. The DED-susceptible MDV1 and the DED-resistant MDV2.3 showed a similar number of total DEGs involved in the local response (3,033 and 3,029, respectively) and also a similar response pattern, especially regarding the proportion of induced/repressed genes and their variation across the response. However, the number of total DEGs involved in the local response of the resistant VAD2 was higher (3,325 DEGs). It was also noteworthy that the two DED-resistant genotypes, MDV2.3 and VAD2, showed marked differences in their responses. The expression profile regarding the number of DEGs within a given sampling time was different for VAD2 compared to the other two genotypes, as the greatest transcriptomic response was produced at 24 hpi for MDV1 and MDV2.3, and at 6 hpi for VAD2.

As shown in the top barplot in Fig. 5, intersections were sorted in the x axis by the sizes (number of DEGs) in y axis, indicating a considerably higher number of DEGs that were exclusively found under each time point

and genotype, while the first groups of shared genes were found between two time-points within one genotype (Supplementary Table S14). These findings align with those observed in a longer-term time-course analysis (from 7 to 28 days post-inoculation) of ash dieback response, which identified a higher number of genotype- and time-specific DEGs compared to shared DEGs<sup>11</sup>. The scarcity of common responses, particularly in resistant genotypes, may be explained by a lack of close co-evolution between *U. minor* and DED pathogens, as evidenced by the high susceptibility and mortality of field elms during the two DED pandemics<sup>43</sup>.

## Usage Notes

This dataset is a collection of a transcriptome assembly<sup>40</sup>, functional annotation and expression profiles of the local response to DED in a time-course experiment (at 6-, 24-, 72- and 144 hpi) in three *U. minor* genotypes showing different levels of resistance to the disease. Further research can leverage these raw reads and the annotation file for the identification of genomic features and improve future genome assembly. In addition, researchers can use this assembly as a reference for further transcriptomic profiling of elms in response to DED. The catalogue of candidate genes proposed here, with special focus on genes involved in plant immunity and defensive mechanisms, opens a vast range of opportunities to further investigate the genetic basis of elm susceptibility/resistance to DED through complementary analyses.

## Code availability

Bioinformatic pipelines and R scripts for this work were uploaded in GitHub ([https://github.com/vchano/genesis\\_transcriptome.assembly](https://github.com/vchano/genesis_transcriptome.assembly)).

Received: 12 March 2025; Accepted: 3 July 2025;

Published online: 23 July 2025

## References

- Bentham, A. R. *et al.* A molecular roadmap to the plant immune system. *J. Biol. Chem.* **295**, 14916–14935, <https://doi.org/10.1074/jbc.REV120.010852> (2020).
- Zhou, J.-M. & Zhang, Y. Plant immunity: danger perception and signaling. *Cell* **181**(5), 978–989, <https://doi.org/10.1016/j.cell.2020.04.028> (2020).
- Naidoo, S., Slippers, B., Plett, J. M., Coles, D. & Oates, C. N. The road to resistance in forest trees. *Front. Plant Sci.* **10**, 273, <https://doi.org/10.3389/fpls.2019.00273> (2019).
- Kovalchuk, A. *et al.* Dual RNA-seq analysis provides new insights into interactions between Norway spruce and necrotrophic pathogen *Heterobasidion annosum* s.l. *BMC Plant Biol* **19**, 2, <https://doi.org/10.1186/s12870-018-1602-0> (2019).
- Hernandez-Escribano, L., Visser, E. A., Iturriza, E., Raposo, R. & Naidoo, S. The transcriptome of *Pinus pinaster* under *Fusarium circinatum* challenge. *BMC Genom* **21**, 28, <https://doi.org/10.1186/s12864-019-6444-0> (2020).
- Trujillo-Moya, C. *et al.* RNA-Seq and secondary metabolite analyses reveal a putative defence-transcriptome in Norway spruce (*Picea abies*) against needle bladder rust (*Chrysomyxa rhododendri*) infection. *BMC Genom* **21**, 336, <https://doi.org/10.1186/s12864-020-6587-z> (2020).
- Zamora-Ballesteros, C. *et al.* Dual RNA-sequencing analysis of resistant (*Pinus pinea*) and susceptible (*Pinus radiata*) hosts during *Fusarium circinatum* challenge. *Int. J. Mol. Sci.* **22**(10), 5231, <https://doi.org/10.3390/ijms22105231> (2021).
- Hayden, K. J. *et al.* Dual RNA-seq of the plant pathogen *Phytophthora ramorum* and its tanoak host. *Tree Genet. Genomes* **10**, 489–502, <https://doi.org/10.1007/s11295-014-0698-0> (2014).
- Meyer, F. E. *et al.* Dual RNA-sequencing of *Eucalyptus nitens* during *Phytophthora cinnamomi* challenge reveals pathogen and host factors influencing compatibility. *Front. Plant Sci.* **7**, 191, <https://doi.org/10.3389/fpls.2016.00191> (2016).
- Islam, M. T. *et al.* Deciphering the genome-wide transcriptomic changes during interactions of resistant and susceptible genotypes of American elm with *Ophiostoma novo-ulmi*. *J. Fungi* **8**(2), 120, <https://doi.org/10.3390/jof8020120> (2022).
- Chano, V. *et al.* Transcriptional time-course analysis during ash dieback infection revealed different responses in tolerant and susceptible *Fraxinus excelsior* genotypes. *BMC Plant Biol* **25**, 107, <https://doi.org/10.1186/s12870-025-06074-z> (2025).
- Ferrari, R. C. *et al.* Transcriptional profiling of *Fraxinus excelsior* leaves during the early infection phase of ash dieback. *J. Plant Dis. Prot.* **132**, 48, <https://doi.org/10.1007/s41348-024-01028-3> (2025).
- Martin, J. A. *et al.* Complexities underlying the breeding and deployment of Dutch elm disease resistant elms. *New For* **54**, 661–696, <https://doi.org/10.1007/s11056-021-09865-y> (2023).
- Brasier, C. M. Intercontinental spread and continuing evolution of the Dutch elm disease pathogens, in *The Elms*, ed. C. P. Dunn [https://doi.org/10.1007/978-1-4615-4507-1\\_4](https://doi.org/10.1007/978-1-4615-4507-1_4), 61–72. (Berlin: Springer). (2000).
- Paoletti, M., Buck, K. W. & Brasier, C. M. Cloning and sequence analysis of the MAT-B (MAT-2) genes from the three Dutch elm disease pathogens, *Ophiostoma ulmi*, *O. novo-ulmi* and *O. himal-ulmi*. *Mycol. Res.* **109**, 983–991, <https://doi.org/10.1017/S0953756205003308> (2005).
- Martin, J. A., Fuentes-Utrilla, P., Gil, L. & Witzell, J. Ecological factors in Dutch elm disease complex in Europe – a review. *Ecol. Bullet.* **53**, m209–224, <https://www.jstor.org/stable/41442032> (2010).
- Martin, J. A., Sobrino-Plata, J., Rodríguez-Calcerrada, J., Collada, C. & Gil, L. Breeding and scientific advances in the fight against Dutch elm disease: will they allow the use of elms in forest restoration? *New For* **50**, 83–215, <https://doi.org/10.1007/s11056-018-9640-x> (2019).
- Dominguez, J., Macaya-Sanz, D., Gil, L. & Martin, J. A. Excelling the progenitors: Breeding for resistance to Dutch elm disease from moderately resistant and susceptible native stock. *For. Ecol. Manag.* **511**, 120113, <https://doi.org/10.1016/j.foreco.2022.120113> (2022).
- Martin, J. A. *et al.* Seven *Ulmus minor* clones tolerant to *Ophiostoma novo-ulmi* registered as forest reproductive material in Spain. *IForest* **8**, 172–180, <https://doi.org/10.3832/ifor1224-008> (2015).
- Perdiguerro, P., Venturas, M., Cervera, M. T., Gil, L. & Collada, C. Massive sequencing of *Ulmus minor*'s transcriptome provides new molecular tools for a genus under the constant threat of Dutch elm disease. *Front. Plant Sci.* **6**, 541, <https://doi.org/10.3389/fpls.2015.00541> (2015).
- Gil, L., Fuentes-Utrilla, P., Soto, A., Cervera, M. T. & Collada, C. English elm is a 2,000-year-old Roman clone. *Nature* **431**, 1053–1053, <https://doi.org/10.1038/4311053a> (2004).
- Perdiguerro, P. *et al.* Gene expression trade-offs between defence and growth in English elm induced by *Ophiostoma novo-ulmi*. *Plant Cell Environ* **41**(1), 198–214, <https://doi.org/10.1111/pce.13085> (2018).
- Martínez-Arias, C., Sobrino-Plata, J., Gil, L., Rodríguez-Calcerrada, J. & Martin, J. A. Priming of plant defenses against *Ophiostoma novo-ulmi* by elm (*Ulmus minor* Mill.) fungal endophytes. *J. Fungi* **7**(9), 687, <https://doi.org/10.3390/jof70906> (2021).
- Conde, P., Sousa, A., Costa, A. & Santos, C. A protocol for *Ulmus minor* Mill. Micropropagation and acclimatization. *PCOTC* **92**, 113–119, <https://doi.org/10.1007/s11240-007-9310-8> (2008).



25. Solla, A. *et al.* Screening European elms for resistance to *Ophiostoma novo-ulmi*. *For. Sci.* **51**, 134–141, <https://doi.org/10.1093/forests/51.2.134> (2005).
26. Tchernoff, V. Methods for screening and for the rapid selection of elms for resistance to Dutch elm disease. *Acta Bot. Neerl.* **14**, 409–452, <https://doi.org/10.1111/j.1438-8677.1965.tb00204.x> (1965).
27. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinfo* **32**(19), 3047–3048, <https://doi.org/10.1093/bioinformatics/btw354> (2016).
28. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinfo* **30**(15), 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
29. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915, <https://doi.org/10.1038/s41587-019-0201-4> (2019).
30. Forgetta, V. *et al.* Sequencing of the Dutch elm disease fungus genome using the Roche/454 GS-FLX titanium system in a comparison of multiple genomics core facilities. *J. Biomol. Tech.* **24**, 39–49, <https://doi.org/10.1007/jbt.12-2401-005> (2013).
31. Haas, B. J. *et al.* De novo transcript sequence reconstruction for mRNA-Seq using Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512, <https://doi.org/10.1038/nprot.2013.084> (2013).
32. Chikhi, R. & Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinfo* **30**(1), 31–37, <https://doi.org/10.1093/bioinformatics/btt310> (2013).
33. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinform* **10**, 421, <https://doi.org/10.1186/1471-2105-10-421> (2009).
34. Mora-Marquez, F., Chano, V., Vázquez-Poletti, J. L. & López de Heredia, U. TOA: A software package for automated functional annotation in non-model plant species. *Mol. Ecol. Res.* **21**(2), 621–636, <https://doi.org/10.1111/1755-0998.13285> (2021).
35. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**(10), 3420–3435, <https://doi.org/10.1093/nar/gkn176> (2008).
36. Bray, N., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527, <https://doi.org/10.1038/nbt.3519> (2016).
37. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 *Genome Biol* **15**, 550, <https://doi.org/10.1186/s13059-014-0550-8> (2014).
38. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* **57**(1), 289–300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x> (1995).
39. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinfo* **33**(18), 2938–2940, <https://doi.org/10.1093/bioinformatics/btx364> (2017).
40. Chano, V. *et al.* TSA: *Ulmus minor*, transcriptome shotgun assembly. *Genbank* <https://www.ncbi.nlm.nih.gov/nuccore/GLCK000000000> (2025).
41. Chano, V. *et al.* *Ulmus minor* response to Dutch elm disease: de novo transcriptome assembly and annotation. *Figshare* <https://doi.org/10.6084/m9.figshare.28548182> (2025).
42. NCBI Sequence Read Archive [https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP566393&o=acc\\_s%3Aa](https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP566393&o=acc_s%3Aa) (2025).
43. Brasier, C. M. & Webber, J. F. Is there evidence for post-epidemic attenuation in the Dutch elm disease pathogen *Ophiostoma novo-ulmi*? *Plant Pathol* **68**(5), 921–929, <https://doi.org/10.1111/ppa.13022> (2019).

## Acknowledgements

We thank Sara Ormeño Moncalvillo and all the personnel working for the Spanish elm breeding program for their technical assistance. This research was funded by the Spanish Ministry of Economy and Competitiveness through the project GENESIS (AGL-2015-66952-R; MINECO-FEDER).

## Author contributions

Conceptualization: C.C., J.R.C., J.A.M.; Laboratory work: J.S.P., C.M.A.; Bioinformatics and data analysis: V.C.; Funding acquisition: C.C., J.R.C., J.A.M. Investigation: V.C., J.S.P., J.R.C., J.A.M.; Writing – original draft: V.C.; Writing – review and editing: all.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-05539-y>.

**Correspondence** and requests for materials should be addressed to V.C. or J.A.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025