

# Genes for ash tree resistance to an insect pest identified via comparative genomics

Laura J. Kelly<sup>1,2\*</sup>, William J. Plumb<sup>1,2,3</sup>, David W. Carey<sup>4</sup>, Mary E. Mason<sup>4</sup>, Endymion D. Cooper<sup>1</sup>, William Crowther<sup>1,6</sup>, Alan T. Whittmore<sup>5</sup>, Stephen J. Rossiter<sup>1</sup>, Jennifer L. Koch<sup>4</sup>, Richard J. A. Buggs<sup>1,2\*</sup>

<sup>1</sup>School of Biological and Chemical Sciences, Queen Mary University of London, London, UK. <sup>2</sup>Royal Botanic Gardens, Kew, Richmond, Surrey, UK. <sup>3</sup>Forestry Development Department, Teagasc, Dublin, Republic of Ireland. <sup>4</sup>United States Department of Agriculture, Forest Service, Northern Research Station, Delaware, Ohio, USA. <sup>5</sup>United States Department of Agriculture, Agricultural Research Service, US National Arboretum, Washington, DC, USA. <sup>6</sup>Present address: School of Life Sciences, The University of Warwick, Coventry, UK. \*email: l.kelly@kew.org; r.buggs@kew.org.

**Genome-wide discovery of candidate genes for functional traits within a species typically involves the sequencing of large samples of phenotyped individuals<sup>1</sup>, or linkage analysis through multiple generations<sup>2</sup>. When a trait occurs repeatedly among phylogenetically independent lineages within a genus, a more efficient approach may be to identify genes via detection of amino acid residues shared by species possessing that trait<sup>3,4</sup>. Here, by taking this approach, we identify candidate loci in the genus *Fraxinus* (ash trees) for resistance to the emerald ash borer beetle (EAB; *Agrilus planipennis*), a pest species that appears innocuous to otherwise healthy ash in its native East Asian range<sup>5</sup> but is highly destructive in North America<sup>6</sup> and poses a threat to ash trees in Europe<sup>7</sup>. Assembling whole genome sequences for 24 diploid species and subspecies of ash, and estimating resistance to EAB for 26 taxa from egg bioassays, we find 53 genes containing amino acid variants shared between two or more independent *Fraxinus* lineages with EAB-resistant species, that are unlikely to be due to chance or undetected paralogy. Of these, seven genes have putative roles relating to the phenylpropanoid biosynthesis pathway and 17 are potentially connected to herbivore recognition, defence signalling or programmed cell death. We also find that possible loss-of-function mutations among our 53 candidate genes are more frequent in susceptible species,**

**than in resistant ones. Patterns of polymorphism for the EAB-associated amino acid variants in ash trees representing different European populations suggest that selection may be able to enhance their resistance to EAB.**

EAB has proved more costly than any other invasive forest insect within the USA to date<sup>8</sup>. We assessed resistance to EAB for 26 *Fraxinus* taxa (Supplementary Table 1). Tree resistance was scored according to the instar, health and weight of EAB larvae in the stems of artificially infested trees eight weeks after infestation<sup>9</sup> (Methods, Supplementary Table 1). In *Fraxinus baroniana*, *F. chinensis*, *F. floribunda*, *F. mandshurica*, *F. platypoda* and *Fraxinus* sp. D2006-0159, least squares means (LSM) of the proportion of host killed larvae (number of larvae killed by tree defence response divided by total larvae entering the tree) were >0.75 (Fig. 1a, Supplementary Table 2) and no surviving larvae reached the last developmental instar (L4; Fig. 1b, Supplementary Table 2), indicating that these species are resistant to EAB. In contrast, all other taxa tested had a LSM proportion of larvae killed of 0.58 or less (Fig. 1a) and had LSM for L4 larvae proportion between 0 and 0.89 (Fig. 1b).

To search for candidate genes related to defence response against EAB, we sequenced and assembled the genomes of 28 individuals from 26 diploid taxa representing all sections within the genus<sup>10</sup>, including a common EAB-susceptible accession and a rare putatively EAB-resistant accession<sup>9</sup> for *F. pennsylvanica* (Supplementary Table 3). For all individuals we generated c. 35 to 85X whole genome shotgun coverage with Illumina sequencing platforms (Methods; Supplementary Table 4). On assembly (Methods) these data generated 133,719 to 715,871 scaffolds for each individual, with N50s ranging from 1,987 to 50,545bp (Supplementary Table 4). We annotated genes in these assemblies via a reference based approach (Methods) using the published genome annotation of *F. excelsior*<sup>11</sup>. We clustered the protein sequences of these genes into putative orthologue groups (OGs; Methods), also including protein sequences from the *F. excelsior* reference genome and the published genome annotations of *Olea europaea*<sup>12</sup>, *Erythranthe guttata*<sup>13</sup> and *Solanum lycopersicum*<sup>14</sup>. We found a total of 87,194 OGs, each containing sequences from between two and 32 taxa; 1,403 OGs included a sequence from all 32 taxa.

We generated multiple sequence alignments for the 1,403 OGs including all taxa and inferred gene-trees for each (Methods). In order to generate a species-tree estimate for the

genus *Fraxinus*, we conducted Bayesian concordance analysis (Methods). This resulted in a tree based on 272 phylogenetically informative low copy genes (Fig. 2 and Supplementary Note 1). Within this tree, the EAB-resistant taxa identified from our bioassays occurred in three independent lineages. (1) *F. baroniana*, *F. chinensis*, *F. floribunda* and *Fraxinus* sp. D2006-0159 clustered together, within a larger clade that included most species in section *Ornus*, including susceptible *F. ornus*. (2) *F. mandshurica* occurred within a clade corresponding to section *Fraxinus* that also included susceptible taxa. (3) *F. platypoda* was sister to a clade corresponding to section *Melioides*, which includes most of the susceptible American species.

We searched for amino acid variants putatively convergent between the resistant lineages using an approach that identifies loci with a level of convergence in excess of that likely to be due to chance alone (grand-conv; see Methods). We conducted three pairwise analyses of lineages: (1) *F. mandshurica* versus *F. platypoda*, (2) *F. mandshurica* versus *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, (3) *F. platypoda* versus *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159. In all these analyses we included three outgroups and five *Fraxinus* species with high susceptibility (Methods). Each analysis was based on alignments of OGs found in all of the included taxa: 3,454 OGs in analysis 1, 3,097 OGs in analysis 2 and 3,026 OGs in analysis 3. Our candidate amino acid variants were those identified by grand-conv as convergent (minimum posterior probability of 0.90) within loci predicted to have the highest excess of convergent over divergent substitutions in the resistant lineages (Methods). These loci were then checked for the possible confounding effect of paralogy, as well as gene model and alignment errors, leaving a total of 67 amino acid sites in 53 genes (Supplementary Note 2 and Supplementary Table 5). Phylogenetic analysis of the CDS alignments for the 53 genes revealed that, in all but one case (OG20252; Supplementary Fig. 1a), the pattern of amino acid variation at candidate sites is better explained by a hypothesis of convergent point mutations, rather than introgressive hybridisation or incomplete lineage sorting (Supplementary Fig. 1). For four loci (OG11013, OG20859, OG37870 and OG41448) the state identified as convergent by grand-conv appears more likely to be ancestral within *Fraxinus*, with change occurring in the other direction (i.e. from the “convergent” state identified by grand-conv to the “non-convergent” state; Supplementary Table 5).

Three of the 67 candidate amino acids are predicted to be presence/absence variants for phosphorylation sites (Supplementary Fig. 2a, 2b and Supplementary Table 5), a post-translational modification that plays a key role in regulating plant immune signalling<sup>15</sup>. We also looked for evidence of loss-of-function of the 53 candidate genes, based on the presence of frameshifts, stop codon gains and start codon losses, in any of the *Fraxinus* individuals included in our convergence analyses. Six of our 53 candidate genes appear to lack a fully functional allele in a susceptible taxon, compared with one for resistant taxa (Supplementary Note 3 and Supplementary Table 6), suggesting these susceptible taxa may have impaired function of some genes related to defence against EAB.

Among our 53 candidate genes, seven have putative roles relating to the phenylpropanoid biosynthesis pathway (Supplementary Note 4). This pathway generates antifeedant and cytotoxic compounds, as well as products involved in structural defence, such as lignin<sup>16</sup>; it can contribute to indirect defence by producing volatiles which attract parasitoids or predators<sup>17</sup>. Loci OG15551, OG853 and OG16673 are of particular interest. Four convergent amino acids were identified in OG15551 (Fig. 3), a paralogue of *CYP98A3* (Supplementary Fig. 3a), which encodes a critical phenylpropanoid pathway enzyme<sup>18</sup>. Three of the four residues fall within *CYP98A3* putative substrate recognition sites, with two at positions predicted to contact the substrate<sup>19</sup> including a leucine (sulphur containing)/methionine (non-sulphur containing) variant (Fig. 3). OG853 is apparently orthologous to *MED5a/RFRI* (Supplementary Fig. 3b), a known regulator of the phenylpropanoid pathway<sup>20,21</sup> that seems to be involved in regulation of defence response genes<sup>22</sup>. OG16673 is a likely glycoside hydrolase; putative *Arabidopsis thaliana* homologues belong to glycoside hydrolase family 1 and have beta-glucosidase activity, with functions such as chemical defence against herbivory, lignification and control of phytohormone levels<sup>23</sup>. A role for beta-glucosidases in defence against EAB in individual *Fraxinus* species has been previously suggested on the basis of chemical<sup>24</sup> and transcriptomic<sup>25</sup> data, and several metabolomic studies have indicated that products of the phenylpropanoid pathway could be involved<sup>26</sup>.

We found 15 candidate genes (Supplementary Note 4) with possible roles in perception and signalling relevant to defence response against herbivorous insects<sup>17</sup>. OG4469 is a probable orthologue of AtG-LecRK-1.6, a G-type lectin receptor kinase (LecRK) with ATP binding activity (Supplementary Note 4.3). G-type LecRKs can act as pattern

recognition receptors (PRRs) in the perception of feeding insects<sup>27</sup>; extracellular ATP is a damage-associated molecular pattern (DAMP) whose perception can trigger defence response related genes<sup>27</sup>. OG38407 appears orthologous to *SNIPER4* (Supplementary Fig. 1), a F-box protein encoding gene involved in regulating turnover of defence-response related proteins, for optimal defence activation<sup>28</sup>; the convergent site is in a leucine rich repeat (LRR) region (Supplementary Fig. 2c), which is involved in recognition of substrate proteins for ubiquitination<sup>28,29</sup>.

Several genes appear to relate to phytohormone biosynthesis and signalling, including those with putative functions in jasmonate (JA; OG41448), brassinosteroid (OG43828), cytokinin (OG39275) and abscisic acid (ABA; OG47560) biosynthesis, and GO terms associated with hormone metabolism and biosynthesis are significantly enriched among our set of candidate genes (Supplementary Note 5; Supplementary Table 7). JA signalling is the central regulatory pathway for defence response against insect herbivores<sup>27,30</sup> whereas brassinosteroids and cytokinins can play important roles in insect resistance, via modulation of the JA pathway<sup>30,31</sup>. ABA is induced by herbivory and is a known modulator of resistance to insect herbivores<sup>27,30</sup>. OG11720 is putatively orthologous to *NRT1.5/NPF7.3*, a member of the NRT1/PTR family<sup>32</sup> which is involved in transport of multiple phytohormones (Supplementary Note 4.3); a transcript matching this gene family had decreased expression in response to both mechanical wounding and EAB feeding in *F. pennsylvanica*<sup>25</sup>. Putative functions of further candidates relate to other signalling molecules involved in triggering defence response (Supplementary Note 4.5), including calcium (OG50989)<sup>17,33</sup>, nitric oxide (NO; OG21033)<sup>34,35</sup> and spermine (OG33348)<sup>36</sup>. Increased resistance to EAB can be artificially induced in *Fraxinus* species with otherwise high susceptibility<sup>37</sup>, leading to the suggestion that susceptible species may fail to recognise, or respond quickly enough to, early signs of EAB attack<sup>26</sup>. Our identification of candidate genes putatively involved in perception and signalling underlines the possibility of differences between EAB-resistant and susceptible *Fraxinus* species in both their ability to sense and react to attacking insects.

Hypersensitive response (HR), involving programmed cell death (PCD), is associated with effector-triggered immunity in response to microbial pathogens<sup>38</sup> but can also be induced by insect herbivory<sup>39</sup> and oviposition<sup>40</sup>. OG16739 and OG37870 are candidates with putative roles related to HR-like effects and PCD. OG16739 has homologues that

control cell death in response to wounding, via the induction of ethylene and the expression of defence and senescence related genes<sup>41</sup>. OG37870 may be orthologous to genes that seem to play a role in controlling PCD of xylem elements<sup>42</sup>. Candidate loci whose putative functions lack an obvious link to plant defence response (Supplementary Note 4), could be involved in other phenotypic traits shared between EAB resistant species or may play a role in defence response that is not yet understood. We found that 19 of our 53 candidates match the same *A. thaliana* genes as transcripts that are differentially expressed in response to elm leaf beetle (either in response to simulated egg deposition, or larval feeding)<sup>43</sup>, including genes, such as OG24969, whose putative *A. thaliana* homologues lack a clear defence-related function.

We analysed allelic variation at the 67 amino acid sites within the 53 candidate genes for all sequenced taxa assessed for resistance to EAB. Of the 67 sites, seven only have the EAB-resistance associated state in resistant taxa, and another is only homozygous for the EAB-resistance associated state in resistant taxa (Supplementary Table 8). Of the 53 candidate genes, four are only homozygous in resistant taxa for the EAB-resistance associated state at the candidate amino acid site(s) detected within them (OG853, OG21449, OG36502 and OG37560; Supplementary Table 8). If we omit the genomes of *F. nigra*, *F. excelsior* and the three *F. angustifolia* subspecies (sect. *Fraxinus*), for 24 of the 53 candidate genes we only find the EAB-resistance associated states in resistant taxa, for 48 genes they are only found in taxa with a LS mean proportion of larvae killed of  $\geq 0.25$  and the remaining five genes are only homozygous for the EAB-resistance associated states in taxa with a LS mean proportion of larvae killed of  $\geq 0.25$  (Supplementary Table 8).

Analysis of previously generated whole genome sequence data for 37 *F. excelsior* individuals from different European provenances revealed that for 50 of the 67 candidate amino acid sites (occurring in 41 of the 53 genes) the EAB-resistance associated state was present, with evidence for polymorphism at seven of these sites (Supplementary Table 9). None of the EAB-resistance associated amino acid states were found in the putatively resistant *F. pennsylvanica* genotype (Supplementary Table 8), suggesting that different genes, or different variants within these genes, are involved in the intraspecific variation in susceptibility of this species. Despite this, transcripts inferred to be from 11 of our candidate genes showed evidence for differential expression subsequent to EAB-

feeding in *F. pennsylvanica* (Supplementary Note 6 and Supplementary Table 10) and two gene families that were highlighted as potentially important for response to tissue damage in *F. pennsylvanica*<sup>25</sup> are also represented among our candidates (see above).

We have provided the first evidence based on genome-wide analyses for the types of genes involved in resistance to EAB in the genus *Fraxinus*, indicating that multiple loci, contributing to different defence responses, underlie this trait. Our genome level data may help to target future efforts to increase the resistance of American and European ash species to EAB via breeding or gene editing.

## METHODS

### *Data reporting*

For the emerald ash borer resistance assays, experiments were conducted using a randomised block design. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

### *Plant material*

All plant materials used in this study were sourced from living or seed collections in the UK or USA. Due to biosecurity measures, we were not able to move living materials between the two countries. In our initial selection of material we relied upon species identifications that had already been made in the arboreta or seed banks within which the materials were held. For each of the accessions included in this study, we PCR amplified and Sanger sequenced the nuclear ribosomal internal transcribed spacer (ITS) region, following standard methods; forward and reverse sequences were assembled into contigs using CLC Genomics Workbench v8.5.1 (QIAGEN Aarhus, Denmark). As far as possible the identity of all materials was verified by Eva Wallander using morphology and ITS sequences, according to her classification of the genus<sup>10</sup>. This led to some re-designation of samples: of particular note, one of the three accessions of *F. pennsylvanica* that we genome sequenced was originally sampled as *F. caroliniana*, and the accession that we originally sampled and genome sequenced as *F. bungeana* was determined to be *F. ornus* (Supplementary Table 3). However, subsequent phylogenetic analysis including allele sequences for this latter individual (see below - Distinguishing between different underlying causes of convergent patterns) indicated it to likely be a hybrid between the *F. ornus* lineage and another lineage within sect. *Ornus* and therefore we designate it as *Fraxinus* sp. 1973-6204. We also designated an accession that was originally sampled for genome sequencing as *F. chinensis* as *Fraxinus* sp. D2006-0159 due to uncertainty regarding species delimitation. Furthermore, for genotype vel-4, that was redetermined as *F. pennsylvanica*, we have maintained its original species name (*F. velutina*). A list of all materials used in the study is shown in Supplementary Table 3, including initial identifications and subsequent identifications by Eva Wallander, as well as details of voucher specimens.

### *Emerald ash borer resistance assays*



Twenty six *Fraxinus* taxa (species, subspecies and one taxon of uncertain status) were collected for egg bioassay experiments (Supplementary Table 3). We aimed to test three clonal replicates (grafts or cuttings) of at least two genotypes of each species. For some taxa less than two genotypes were available in the US, and occasionally genotypes did not propagate well by graft or cutting so seedlings from the same seedlot were used instead (details for each taxon are included in Supplementary Table 1). The majority of egg bioassays were conducted in 2015 and 2016 and groups of approximately 20 genotypes were conducted in each set (week within year; Supplementary Table 1) with one grafted ramet, cutting or seedling in each block and all ramets, cuttings and seedlings within the block randomised to location/order of assay. To facilitate comprehensive analysis, the same controls were repeated in each week (susceptible *F. pennsylvanica* genotype pe-37 and/or pe-39 and resistant *F. mandshurica* genotype ‘mancana’).

Trees were treated as uniformly as possible prior to inoculation. Adult beetles were reared and used for egg production as previously described<sup>9</sup>. Inoculations were performed in a greenhouse to keep conditions uniform for the duration of the assay, and to minimise predation of the eggs. We followed the EAB egg transfer bioassay method reported by Koch et al.,<sup>9</sup> that had previously been used on genotypes of *F. pennsylvanica* and *F. mandshurica* with the changes noted below. The egg dose for each tree was determined according to the method of Duan et al.,<sup>44</sup> which takes into account the bark surface area. A target density of 400 eggs per m<sup>2</sup> was used; this density is above that reported to allow host defences to kill larvae in green ash, but is in the range where competition and cannibalism are minimised<sup>44</sup>. Twelve individual eggs, on a small strip cut from the coffee filter paper on which they were laid, were taped to each tree. The spacing was varied between eggs to maintain a consistent target dose (e.g. eggs placed 7.5 cm apart on stem 1.0-1.1cm diameter to eggs placed 3 cm apart on stem 2.5-2.6cm diameter). The portion of the tree where eggs were placed was wrapped in medical gauze to protect from jostling and egg predation. Past experiments have shown that egg assay results are not consistent on stems less than 1 cm in diameter (JLK, unpublished data). Due to size differences between some species, to achieve the target dose and avoid placing eggs where the stem diameter was <1 cm, occasionally less than 12 eggs were placed. A total of 2199 egg bioassays (each egg represents a bioassay) were conducted on 61 different genotypes and a total of 206 ramets, cuttings or seedlings.

Occasional ramets, cuttings and/or seedlings were considered as assay failures if less than three larvae successfully entered the tree (i.e. the effective egg dose was too low), or if there were other problems with the tree (too small diameter overall, cultivation issues, etc.), and that replicate was excluded from analysis (data not shown). Four weeks after egg attachment each egg was inspected to determine if it had successfully hatched, and if there were signs of the larva entering the tree. Larval entry holes, when detected, were marked to assist with future dissection. At eight weeks, dissection of the entry site was performed and galleries made by larval feeding were carefully traced using a grafting knife to determine the outcome of each hatched egg. Health (dead or alive) and weight (in cases when larvae could be recovered intact) was recorded for each larva, and developmental instar was determined using measurements of head capsule and length<sup>45,46</sup>

Preliminary exploratory data analysis indicated that the proportion of “tree killed” (i.e. larvae killed by tree defence response) and the proportion of live L4 larvae (number divided by the number of larvae that entered the tree) were the best variables to distinguish resistance versus susceptibility at the species level. We fitted a generalised linear mixed model to the proportion tree killed and proportion L4 using the GLIMMIX procedure in SAS. The model specification is proportion as a binomial distribution with a logit link function, species as a fixed effect, and block/replicate nested within sequential week (week within year) as a random effect (this allowed for comprehensive analysis over years and weeks with correct variance/covariance restrictions). Non-significant predictors were eliminated from the final model. Least squares means of tree killed or L4 proportion were calculated with confidence intervals on the data scale (proportion).

### ***Genome size estimation by flow cytometry (FC)***

We used FC to estimate the genome size of individuals used for whole genome sequencing (Supplementary Table 4). *Fraxinus* samples from UK collections were prepared and analysed as described in Pellicer et al. (2014), with the exception that ‘general purpose isolation buffer’ (GPB<sup>47</sup>) without the addition of 3% polyvinylpyrrolidone (PVP-40) and LB01 buffer<sup>48</sup> were used for some samples. *Oryza sativa* (‘IR-36’; 1C = 0.50 pg<sup>49</sup>) was used as an internal standard. For each individual analysed, two samples were prepared (from separate leaves or different parts of the same leaf) and two replicates of each sample run. *Fraxinus* samples from US collections were analysed using a Sysmex CyFlow Space flow cytometer, as described in Whittemore and

Xia<sup>50</sup>; *Pisum sativum* ('Ctirad'; 1C = 4.54 pg<sup>51</sup>) and *Glycine max* ('Williams 82'; 1C = 1.13 pg<sup>52</sup>) were used as internal standards. For each individual analysed, six samples were prepared (from separate leaves or different parts of the same leaf) and three samples run with each size standard.

### ***DNA extraction***

Total genomic DNA was extracted from fresh, frozen or silica-dried leaf or cambial material, using either a CTAB extraction protocol modified from<sup>53</sup> or using a Qiagen DNeasy<sup>®</sup> Plant Mini or Maxi kit.

### ***Genome sequencing and assembly***

For each of the 28 diploid individuals selected for whole genome sequencing (Supplementary Table 3), sufficient Illumina sequence data were generated to provide a minimum of c. 30x coverage of the 1C genome size, based on the C-value estimates obtained for the same individuals (see above - Genome size estimation by flow cytometry (FC)), or those of closely related taxa. Libraries with average insert sizes of 300 or 350bp, 500 or 550bp and 800bp were prepared from total genomic DNA by the Genome Centre, at Queen Mary University of London, and the Centre for Genomic Research, at the University of Liverpool. Paired-end reads of 125, 150 or 151 nucleotides were generated using the Illumina NextSeq 500, HiSeq 2500 and HiSeq 4000 platforms (Illumina, San Diego, California, USA); see Supplementary Table 4 for the exact combination of libraries, read lengths and sequencing platforms used for each individual. For selected taxa, chosen to represent different sections within the genus, we also generated data from long mate-pair (LMP) libraries (Supplementary Table 4). LMP libraries with average insert sizes of 3kb and 10kb were prepared from total genomic DNA by the Centre for Genomic Research, at the University of Liverpool, and sequenced on an Illumina HiSeq 2500 to generate reads of 125 nucleotides to a depth of c. 10x coverage of the 1C genome size.

Initial assessment of sequence quality was performed for all read pairs from the short-insert libraries (300-800bp inserts) using FastQC v.0.11.3 or v.0.11.5 ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)). Reads were clipped using the fastx\_trimmer tool in the FASTX-Toolkit v.0.0.14 ([http://hannonlab.cshl.edu/fastx\\_toolkit/index.html](http://hannonlab.cshl.edu/fastx_toolkit/index.html)) to remove the first 5-10 nucleotides

of each read; for the NextSeq reads, the last 5 nucleotides were also clipped. Adapter trimming was performed using cutadapt v.1.8.1<sup>54</sup> with the “O” parameter set to 5 (i.e. minimum overlap of five bases) and using option “b” (i.e. adapters allowed to occur on both the 5’ and 3’ end of reads); default settings were used for all other parameters. Quality trimming and length filtering was performed using Sickle v.1.33<sup>55</sup> with the “pe” option and the following parameter settings: -t sanger -q 20 -l 50 and default settings for other parameters. This yielded quality trimmed paired and singleton reads with a minimum length of 50 nucleotides; only intact read pairs were used for downstream analyses.

For the LMP libraries, duplicate reads were removed using NextClip v.1.3.1 (<https://github.com/richardmleggett/nextclip>) with the --remove\_duplicates parameter specified and default settings for all other parameters. Adapter trimming was performed using cutadapt v.1.10<sup>54</sup>; junction adapters were removed from the start of reads by running option “g”, with the adapter sequence anchored to the beginning of reads with the “^” character, and the following settings for other parameters: -O 10 -n 2 -m 25. Other adapter trimming was performed using option “a”, with the further parameters set to the same values as specified above. Quality trimming was performed with PRINSEQ-lite v.0.20.4<sup>56</sup>, with the following parameter settings: -trim\_qual\_left 20 -trim\_qual\_right 20 -trim\_qual\_window 20 -trim\_tail\_left 101 -trim\_tail\_right 101 -trim\_ns\_left 1 -trim\_ns\_right 1 -min\_len 25 -min\_qual\_mean 20 -out\_format 3.

*De novo* genome assembly was performed for each individual using CLC Genomics Workbench v8.5.1 (QIAGEN Aarhus, Denmark). All trimmed read pairs from the short-insert libraries were used for assembly under the following parameter settings: automatic optimization of word (k-mer) size; maximum size of bubble to try to resolve=5000; minimum contig length=200bp. Assembled contigs were joined to form scaffolds using SSPACE (version 3.0<sup>57</sup>) with default parameters, incorporating data from mate-pair libraries with 3kb and 10kb insert sizes where available. Library insert lengths were specified with a broad error range (i.e.  $\pm 40\%$ ). Gaps in the SSPACE scaffolds were filled using GapCloser (version 1.12<sup>58</sup>) with default parameters. The average library insert lengths were specified using the estimates produced by SSPACE during scaffolding. Scaffolding and gap filling was not performed for individuals that lacked data from libraries with insert size of  $\geq 500$ bp (only a single insert size library was available for

these taxa; Supplementary Table 4). We did not attempt to extract sequences of organellar origin from the assemblies, or to separately assemble the plastid and mitochondrial genomes.

Sequences within the assemblies that correspond to the Illumina PhiX control library were identified via BLAST. A PhiX bacteriophage reference sequence (GenBank accession number CP004084) was used as a query for BLASTN searches, implemented with the BLAST+ package (v.2.5.0+<sup>59</sup>), against the genome assembly for each taxon with an *E* value cut-off of  $1 \times 10^{-10}$ . Sequences that matched the PhiX reference sequence at this threshold were removed from the assemblies. We used the `assemblathon_stats.pl` script ([https://github.com/ucdavis-bioinformatics/assemblathon2-analysis/blob/master/assemblathon\\_stats.pl](https://github.com/ucdavis-bioinformatics/assemblathon2-analysis/blob/master/assemblathon_stats.pl)) with default settings to obtain standard genome assembly metrics, such as N50. BUSCO v2.0<sup>60</sup> was used to assess the content of the genome assemblies. The “embryophyta\_odb9” lineage was used and analyses run with the following parameter settings: `--mode genome -c 8 -e 1e-05 -sp tomato`.

### ***Gene annotation and orthologue inference***

To annotate genes in the newly assembled *Fraxinus* genomes, we used a similarity based approach implemented in GeMoMa<sup>61</sup>, with genes predicted in the *F. excelsior* BATG0.5 assembly as a reference set. We used the “Full Annotation” gff file for BATG0.5 (*Fraxinus\_excelsior\_38873\_TGAC\_v2.longestCDStranscript.gff3*; available from <http://www.ashgenome.org/transcriptomes>), which contains the annotation for the single longest splice variant for each gene model. This annotation file also includes preliminary annotations for genes within the organellar sequences (gene models FRAEX38873\_v2\_000400370-FRAEX38873\_v2\_000401330) which were not reported in the publication of the reference genome<sup>11</sup>; none of the sets of putative orthologues used for the species-tree inference or molecular convergence analysis (see below) include these preliminary organellar models from the BATG0.5 reference assembly. The Extractor tool from GeMoMa v.1.3.2 was used to format the data from the reference genome (gff and assembly files), with the following parameter settings: `v=true f=false r=true Ambiguity=AMBIGUOUS`. To obtain information on similarity between the reference gene models and sequences in the newly assembled *Fraxinus* genomes, we performed TBLASTN searches of individual exons (i.e. the “cds-parts” file generated by Extractor) against the assembly file for each individual with BLAST+ (v.2.2.29+<sup>59</sup>).

makeblastdb was used to format each assembly file into a BLAST database with the following parameter settings: `-out ./blastdb -hash_index -dbtype nucl. tblastn` was then run with the “`cds-parts.fasta`” file as the query, with the following parameter settings: `-num_threads 24 -db ./blastdb -evaluate 1e-5 -outfmt "6 std sallseqid score nident positive gaps ppos qframe sframe qseq sseq qlen slen salltitles" -db_gencode 1 -matrix BLOSUM62 -seg no -word_size 3 -comp_based_stats F -gapopen 11 -gapextend 1 -max_hsps 0`. Finally, the GeMoMa tool itself was run for each individual, with the TBLASTN output, `cds-parts` file and *de novo* assembly file as input, with the “`e`” parameter set to “`1e-5`” and default settings for all other parameters. Because GeMoMa generates predictions for each reference gene model separately, the output may contain gene models that are at identical, or overlapping, positions, especially for genes that belong to multi-gene families (<http://www.jstacs.de/index.php/GeMoMa#FAQs>). As the presence of these redundant gene models does not prevent the correct inference of sets of orthologues with OMA (see below), we opted to retain the predicted proteins from all gene model predictions generated by GeMoMa for input into OMA. The `gffread` utility from `cufflinks v.2.2.1`<sup>62</sup> was used to generate the CDS for each gene model; `getfasta` from `bedtools v.2.26.0`<sup>63</sup> used to generate full-length gene sequences (i.e. including introns where present), with the “`-name`” and “`-s`” options invoked.

To identify sets of putatively orthologous sequences, we used OMA standalone `v.2.0.0`<sup>64,65</sup> to infer OMA groups (OGs) and hierarchical orthologous groups (HOGs). To the protein sets from the 29 diploid *Fraxinus* genome assemblies (the 28 newly generated assemblies, plus the existing reference assembly for *F. excelsior*), we added proteomes from three outgroup species: *Olea europaea* (olive), which belongs to the same family as *Fraxinus* (Oleaceae); *Erythranthe guttata* (monkey flower; formerly known as *Mimulus guttatus*), which belongs to the same order as *Fraxinus* (Lamiales); *Solanum lycopersicum* (tomato), which belongs to the same major eudicot clade as *Fraxinus* (lamiids). For *O. europaea*, we used the annotation for v6 of the genome assembly<sup>12</sup>; the file containing proteins for the single longest transcript per gene (`OE6A.longestpeptide.fa`) was downloaded from: <http://denovo.cnag.cat/genomes/olive/download/?directory=%2FOe6%2F>. For *E. guttata* we used the annotation for v2.0 of the genome assembly<sup>13</sup>; the file containing proteins for the primary transcript per gene (`Mguttatus_256_v2.0.protein_primaryTranscriptOnly.fa`) was downloaded from Phytozome 12<sup>66</sup>. For *S. lycopersicum* we used the annotation for

the vITAG2.4 genome assembly; the file containing proteins for the primary transcript per gene (Slycopersicum\_390\_ITAG2.4.protein\_primaryTranscriptOnly.fa) was downloaded from Phytozome 12.

Fasta formatted files containing the protein sequences from all 32 taxa were used to generate an OMA formatted database. An initial run of OMA was performed using the option to estimate the species-tree from the OGs (option 'estimate' for the SpeciesTree parameter); we set the InputDataType parameter to 'AA' and left all other parameters with the default settings. The species-tree topology from the initial run was then modified in FigTree v.1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>) to reroot it on *S. lycopersicum*; nodes within the main clades of *Fraxinus* species (which corresponded to sections recognised in the taxonomic classification<sup>10</sup>) were also collapsed and relationships between these major clades and any individual *Fraxinus* not placed into a clade, were collapsed. OMA was then rerun with the modified species-tree topology specified in Newick format through the SpeciesTree parameter; the species-tree topology is used during the inference of HOGs (<https://omabrowser.org/standalone/>), and does not influence the OGs obtained.

### ***Species-tree inference***

To obtain a more robust estimate of the species-tree for *Fraxinus* (compared with that estimated by OMA, see above, or existing species-tree estimates based on very few independent loci<sup>67,68</sup>), we selected clusters of putatively orthologous sequences from the results of the OMA analysis. OGs containing protein sequences from all 32 taxa (29 diploid *Fraxinus* and three outgroups) were identified and corresponding CDSs aligned with MUSCLE<sup>69</sup> via GUIDANCE2<sup>70</sup> with the following parameter settings: --program GUIDANCE --msaProgram MUSCLE --seqType codon --bootstraps 100, and default settings for other parameters. Datasets where sequences were removed during the alignment process (identified by GUIDANCE2 as being unreliably aligned) or which failed to align due to the presence of incomplete codons (i.e. the sequence length was not divisible by three) were discarded. Alignment files with unreliably aligned codons removed (i.e. including only codons with GUIDANCE scores above the 0.93 threshold for the "colCutoff" parameter) were used for downstream analyses. A custom Perl script was used to identify alignments shorter than 300 characters in length or which included

sequences with <10% non-gap characters; these datasets were excluded from further analysis.

The remaining alignment files were converted from fasta to nexus format with the `secret` tool from EMBOSS v.6.6.0<sup>71</sup>. MrBayes v. 3.2.6<sup>72</sup> was used to estimate gene-trees with the following parameter settings: `lset nst=mixed rates=gamma; prset statefreqpr=dirichlet(1,1,1,1); mcmc nruns = 2 nchains = 4 ngen = 5000000 samplefreq = 1000; Sumt Burninfrac = 0.10 Contype = Allcompat; Sump Burninfrac = 0.10.` Diagnostics (average standard deviation of split frequencies (ASDSF) and for post burnin samples, potential scale reduction factor (PSRF) for branch and node parameters and effective sample size (ESS) for tree length) were examined to ensure that runs for a given locus had reached convergence and that a sufficient number of independent samples had been taken. We discarded datasets where the ASDSF was  $\geq 0.010$ ; all remaining datasets had an ESS for tree length in excess of 500.

We used BUCKy v.1.4.4<sup>73,74</sup> to infer a species-tree for *Fraxinus* via Bayesian concordance analysis, which allows for the possibility of gene-tree heterogeneity (arising from biological processes such as hybridisation, that is reported to occur within *Fraxinus*<sup>67</sup>) but which makes no assumptions regarding the reason for discordance between different genes<sup>74</sup>. First, `mbsum` v.1.4.4 (distributed with BUCKy) was run on the MrBayes tree files for each locus, removing the trees sampled during the first 500,000 generations of each run as a burn-in. We then used the output from `mbsum` to select the most informative loci on the basis of the number of distinct tree topologies represented in the sample of trees from MrBayes; loci with a maximum of 2000 distinct topologies were retained. BUCKy was then run on the combined set of `mbsum` output files for all retained loci under ten different values for the  $\alpha$  parameter (0.1, 1, 2, 5, 10, 20, 100, 500, 1000,  $\infty$ ), which specifies the *a priori* level of discordance expected between loci. Each analysis with a different  $\alpha$  setting was performed with different random seed numbers (parameters `-s1` and `-s2`) and the following other parameter settings: `-k 4 -m 10 -n 1000000 -c 2`. Run outputs were checked to ensure that the average standard deviation of the mean sample-wide concordance factors (CF; the sample-wide CF is the proportion of loci in the sample with that clade<sup>73</sup>) was  $< 0.01$ . The same primary concordance tree (PCT; the PCT comprises compatible clades found in the highest proportion of loci and represents the



main vertical phylogenetic signal<sup>74</sup>) and CFs for each node were obtained for all settings of  $\alpha$ .

We also repeated the species-tree inference using full-length sequences (i.e. including introns, where present); alignment and gene-tree inference were carried out as described above, with the exception that the `--seqType` parameter in GUIDANCE2 was set to “nuc”. BUCKy analyses were performed as described above. The same PCT was obtained for all settings of  $\alpha$ , with only minor differences in the mean sample-wide concordance factors. The PCT inferred from the full-length datasets was also identical to that obtained from the CDS analyses; we based our final species-tree estimate on the output of the full-length analyses due to the presence of a larger number of informative loci within these datasets.

One of the informative loci from the CDS analyses (i.e. those with  $\leq 2000$  distinct topologies within their gene-tree sample), and three of those from the full-length analyses, were subsequently found to be among our filtered set of candidate loci with evidence of convergence between EAB-resistant taxa (see below). To test whether the signal from these loci had an undue influence on the species-tree estimation, we excluded them and repeated the BUCKy analyses for the CDS and full-length datasets as described above, with the exception that only an  $\alpha$  parameter setting of 1 was used. The PCTs obtained from these analyses were identical to those inferred when including all datasets, with minor (i.e. 0.01) differences in CFs.

In addition to the analysis including all taxa, we also performed BUCKy analyses for 13 taxa selected for inclusion in the grand-conv analyses and for the subsets of 10-12 taxa for each of the three grand-conv pairwise comparisons (see below - Analysis of patterns of sequence variation consistent with molecular convergence). OGs containing protein sequences from all 13 taxa (10 *Fraxinus* and three outgroups) were identified and corresponding CDSs for these 13 taxa aligned with MUSCLE via GUIDANCE2, as described above. We also identified and aligned CDSs for all additional OGs which did not include all 13 taxa, but which contained proteins for all taxa in one of the subsets used for the pairwise comparisons. Filtering of alignments and gene-tree inference were carried out as described above for the full set of 32 taxa. BUCKy was run separately with the MrBayes tree sample samples for loci including all 13 taxa and for additional loci

including taxa for each of the three smaller subsets for the grand-conv pairwise comparisons. BUCKy analyses were performed as described above, with the exception that no filtering of loci on the basis of number of distinct gene-tree topologies was performed, a value of between 2 and 35 was used for the -m parameter, and only a single  $\alpha$  parameter setting, of 0.1, was used. Topologies for the PCTs for the set of 13 taxa and subsets of 10-12 taxa were both congruent with each other and with the PCTs inferred from analyses including all taxa, for all nodes with a CF of  $\geq 0.38$ .

### ***Analysis of patterns of sequence variation consistent with molecular convergence***

To test for signatures of putative molecular convergence in protein sequences we used a set of diploid taxa representing the extremes of variation in susceptibility to EAB, as assessed by our egg bioassays. By limiting our analysis at this stage to this subset of taxa we could also maximise the number of genes analysed, as with increasing taxon sampling the number of OGs for which all taxa are represented decreases. This set comprised: five highly susceptible taxa (*F. americana*, *F. latifolia*, *F. ornus*, *F. pennsylvanica* [susceptible genotype], and *F. velutina*), five resistant taxa (*F. baroniana*, *F. floribunda*, *F. mandshurica*, *F. platypoda* and *Fraxinus* sp. D2006-0159) and three outgroups (*O. europaea*, *E. guttata* and *S. lycopersicum*).

We did not include *F. nigra* as one of the highly susceptible species, although it has been used in comparisons of resistant and susceptible species in several previous studies (e.g. <sup>75,76</sup>), because detailed analysis of the phenotype of *F. nigra* individuals using EAB bioassays indicates evidence of a possible widespread hypersensitive-like response, which rather than being a beneficial defence response may actually accelerate tree death (JLK, unpublished data), similar to what has been proposed for *Tsuga canadensis* in response to hemlock woolly adelgid<sup>39</sup>. This is consistent with reports of mortality of *F. nigra* occurring at lower densities of EAB in the field compared with other highly susceptible North American species (N. Siegert and D. McCullough, pers. comm.).

Of the ten *Fraxinus* genome assemblies included in the convergence analysis, six were from genotypes that were also included in the EAB resistance assays outlined above (*F. americana* am-6, *F. baroniana* bar-2, *F. floribunda* flor-ins-12, *F. pennsylvanica* pe-48 and *F. platypoda* spa-1 and *Fraxinus* sp. D2006-0159 F-unk-1). For the other four, we

could not test the exact individual that we sequenced the genome of, but relied upon results of bioassays of other individuals in the same species.

We used the following three pairwise comparisons to seek loci showing amino acid convergence between resistant taxa:

1. *F. mandshurica* (sect. *Fraxinus*) versus *F. platypoda* (*incertae sedis*)
2. *F. mandshurica* (sect. *Fraxinus*) versus *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159 (sect. *Ornus*)
3. *F. platypoda* (*incertae sedis*) versus *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159 (sect. *Ornus*)

The more divergent homologous amino acid sequences are between species, the more likely it is that a convergent amino acid state will occur by chance<sup>4</sup>. In order to account for this, we compared the posterior expected numbers of convergent versus divergent substitutions across all pairs of independent branches of the *Fraxinus*-species-tree for the selected taxa using a beta-release of the software Grand-Convergence v0.8.0 (hereafter referred to as grand-conv; <https://github.com/dekoning-lab/grand-conv>). This software is based on PAML 4.8<sup>77</sup> and is a development of a method used by Castoe et al<sup>4</sup>. It has also been recently used, for example, to detect convergence among *Laverania* species<sup>78</sup>.

For input into grand-conv, we used the same OGs that were the basis of the BUCKY analyses of the 13 taxa selected for the grand-conv analyses, and analyses of the subsets of taxa for the pairwise comparisons (see above - Species-tree inference). Therefore, as well as meeting the criterion of including all relevant taxa, the OGs analysed with grand-conv had also successfully passed the alignment, filtering and gene-tree inference steps. A set of input files was created for each of the three pairwise comparisons that, where present, removed any taxa from the other resistant lineage from the alignments generated by GUIDANCE2. Alignment files were then converted from fasta to phylip format using the Fasta2Phylip.pl script [<https://github.com/josephhughes/Sequence-manipulation>]. To ensure that sequences from each taxon appeared in a consistent order across all datasets (which is necessary for automating the generation of site-specific posterior probabilities for selected branch pairs with grand-conv), the phylip formatted files were sorted using the unix command “sort” prior to input into grand-conv. Species-tree files for each of the

pairwise comparisons were created from the Newick formatted PCTs for the relevant taxon sets generated with BUCKy, with the trees edited to root them on *S. lycopersicum*.

For the grand-conv analysis, “gc-estimate” was first run on the full set of input alignment files for each pairwise comparison, with the following parameters settings: --gencode=0 --aa-model=lg --free-bl=1, specifying the appropriate species-tree file for each of the pairwise comparisons. Next, “gc-discover” was run to generate site-specific values for the posterior probability (PP) of divergence or convergence for the branch pairs of interest; the numbers for the branch pairs relating to the resistant taxa were established from an initial run of “gc-estimate” and “gc-discover” on a single input file, and then specified when running “gc-discover” on all input files using the --branch-pairs parameter. A custom Perl script was then used to filter the output files containing the site-specific posterior probabilities to identify loci with at least one amino acid site where the PP of convergence was higher than divergence and passed a  $\geq 0.9000$  threshold. For this filtered set of datasets with significant evidence of convergence at at least one site, we checked if the “excess” convergence, as measured from the residual values from the non-parametric errors-in-variables regression calculated by grand-conv, was higher for the branch pair of interest than for any other independent pair of branches within the species-tree (i.e. the highest residual was found for the resistant branch pair). Only loci where the highest excess convergence was found in the resistant branch-pair were retained for further analysis.

### ***Refining the initial list of candidate loci identified with grand-conv***

For the set of loci with evidence of convergence between at least one pair of resistant lineages from the grand-conv analyses, we applied additional tests to assess the robustness of the pattern of shared amino acid states. Specifically, we checked for the potential impact of alignment uncertainty and orthology/paralogy conflation. For each candidate locus, we identified the Hierarchical Orthologous Group (HOG) from the OMA analysis, to which the sequences for the candidate locus belong. These HOGs include sequences for an expanded set of taxa (see above - Species-tree inference) and may represent a single gene for all taxa (i.e. a set of orthologous sequences) or several closely related paralogues<sup>64</sup>. Protein sequences for HOGs were aligned with GUIDANCE and gene-tree inference conducted with MrBayes, as described above for the OMA putative orthologous groups with the exception that the --seqType parameter in GUIDANCE was

set to “aa” and in MrBayes the prset parameter was set to “prset aamodelpr = mixed”. Any MrBayes analyses that had not converged after 5M generations (average standard deviation of split frequencies >0.01) were run for an additional 5M generations. The multiple sequence alignment and gene-tree estimates were then used to refine the initial list of candidate loci. Loci were dropped from initial list of candidates if either of the following applied:

1. If in the filtered MSA alignment generated by GUIDANCE, the site/sites where convergence was detected were not present, indicating they were in a part of the protein sequences that can not be aligned reliably.
2. If in the consensus gene-tree estimated by MrBayes, there was evidence that the sequences within which convergence was initially detected (i.e. those belonging to the ten *Fraxinus* species included in the grand-conv analysis) belong to different paralogues and that the pattern of convergence could be explained by sequences with the “convergent” state belonging to one paralogue and the “non-convergent” state belonging to another paralogue. We also excluded two loci that belong to large gene families (>10 copies) for which the MrBayes analyses failed to reach convergence within a reasonable time ( $\leq 10$ M generations) and for which orthology/paralogy conflation could therefore not be excluded.

Additionally, for the set of loci remaining, we checked for errors in the estimation of gene models (including in the reference models from *F. excelsior*) that might impact the results of the grand-conv analyses. Specifically, we dropped from our list any loci where the amino acid sites with evidence of convergence were found to be outside of an exon, or outside of the gene itself, following manual correction of the gene model prediction.

### ***Analysis of variants within candidate loci***

To assess the possible impact of allelic variants (i.e. those not represented in the genome assemblies) on patterns of amino acid variation associated with the level of EAB susceptibility in *Fraxinus*, we called variants (SNPs and indels) and predicted their functional effects. For each sequenced *Fraxinus* individual, trimmed read pairs from the short insert Illumina libraries were mapped to the *de novo* genome assembly for the same individual using Bowtie 2 v.2.3.0<sup>79</sup> with the “very-sensitive” preset and setting “maxins” to between and 1000 and 1400, depending on the libraries available for that individual. Read mappings were converted to BAM format and sorted using the “view” and “sort”

functions in samtools v.1.4.1<sup>80</sup>. Prior to variant calling, duplicate reads were marked and read group information added to the BAM files using the “MarkDuplicates” and “AddOrReplaceReadGroups” functions in picard tools v.1.139 (<http://broadinstitute.github.io/picard>).

Variant discovery was performed with gatk v.3.8<sup>81</sup>. BAM files were first processed to realign INDELS using the “RealignerTargetCreator” and “IndelRealigner” tools. An initial set of variants was called for each individual using the “HaplotypeCaller” tool, setting the `-stand_call_conf` parameter to 30. VCF files from the initial variant calling were then hard filtered to identify low confidence calls by running the “VariantFiltration” tool with the `-filterExpression` parameter set as follows: `"QD < 5.0 || FS > 20.0 || MQ < 30.0 || MQRankSum < -8.0 || MQRankSum > 8.0 || ReadPosRankSum < -2.0 || ReadPosRankSum > 2.0"` (hard filtering thresholds were selected by first plotting the values for FS, MQ, MQRankSum, QD and ReadPosRankSum from the initial set of variant calls for selected individuals, representing the range of different sequence and library types used, to visualise their distribution and then modifying the default hard filtering thresholds in line with the guidance provided in the gatk document “Understanding and adapting the generic hard-filtering recommendations” (<https://software.broadinstitute.org/gatk/documentation/article.php?id=6925>).

Variants passing the gatk hard filtering step (excluding those where alleles had not been called; GT field = `./.`) were further analysed using SnpEff v.4.3u<sup>82</sup>, in order to predict the impact of any variants within genes identified from the convergence analyses. Custom genome databases were built for each individual using the SnpEff command “build” with option “`-gtf22`”; a gtf file containing the annotation for all genes, as well as fasta files containing the genome assembly, CDS and protein sequences, were used as input. Annotation of the impact of variants was performed by running SnpEff with genes of interest specified using the `-onlyTr` parameter and the `-ud` parameter set to “0” to deactivate annotation of up or downstream variants. For each variant predicted to alter the protein sequence, the position of the change was checked to see whether it occurred at a site at which evidence for convergence had also been detected and, if so, whether it involved a change to or from the state identified as being convergent between resistant taxa.

We also used the SnpEff results to check for evidence of mutations that could indicate the presence of non-functional gene copies in certain taxa. Variants annotated as “stop gained”, “start lost” or “frameshift” in the 10 ingroup taxa included in the convergence analysis were manually examined to confirm that they would result in a disruption to the expected protein product, and that they were not false positives caused by errors in gene model estimation (e.g. misspecification of intron/exon boundaries). We checked further for evidence of truncation of sequences or errors in the GeMoMa gene model estimation that might be caused by loss-of-function mutations outside of the predicted exon boundaries (e.g. such as the loss of a start codon, which could cause GeMoMa to predict an incomplete gene model if an alternative possible start codon was present downstream). Such putative loss-of-function mutations would not be detected as such by SnpEff because they would be interpreted as low impact intergenic or intron variants.

We used WhatsHap v0.15<sup>83</sup> to perform read-based phasing of alleles for loci with evidence of multiple variants within them. Input files for phasing in each taxon consisted of the fasta formatted genome assembly, VCF file containing variants passing the gatk hard-filtering step and BAM file from Bowtie 2 with duplicates marked and indels realigned (i.e. as input into variant calling with gatk - see above). The WhatsHap “phase” tool was run with the following parameter settings: --max-coverage 20 --indels; only contigs/scaffolds containing the genes of interest were phased (specified using the “--chromosome” option). For loci with evidence of variants within the CDS, we used the output of WhatsHap to generate fully or partially phased allele sequences. The SnpSift tool from SnpEff v.4.3u<sup>82</sup> was used to select variants that alter the CDS from the annotated VCF file generated by SnpEff and the positions of these variants checked against the WhatsHap output to see if they fell within phased blocks. For each gene, the number and size (i.e. number of phased variants encompassed) of each phased block was found and a custom Perl script used to select the largest (or joint largest) block for genes with at least one block spanning multiple phased variants within the CDS. Details of phased variants impacting the CDS within the selected blocks, and of variants for genes with only a single variant within the CDS (which were not considered for phasing with WhatsHap, but for which the CDS for separate alleles can be generated), were extracted from the WhatsHap output VCF file; these selected variants were then applied to the gene sequences from each genome assembly to generate individual alleles which are fully or partially phased within the CDS. The “faidx” function in samtools v.1.6<sup>80</sup> was used to

extract the relevant subsequences from the genome assembly files, and the “consensus” command in bcftools v.1.4 (<http://www.htslib.org/doc/bcftools-1.4.html>) used to obtain the sequence for each allele with the “-H 1” and “-H 2” options. In cases where the selected phased block also spans unphased variants, both sequences output by bcftools will have the state found in the original genome assembly at these sites, as they will for any variants outside of the selected phased block. The revseq and descseq tools from EMBOSS v.6.6.0<sup>71</sup> were used to reverse complement the sequences for any genes annotated on the minus strand and to rename the output sequences. Phased alleles were used for further phylogenetic analysis of candidate loci (see below); phasing results were also used to check loci with multiple potential loss-of-function mutations within a single individual, to establish whether the mutations are on the same or different alleles. We discounted any cases of potential loss-of-function mutations where multiple frameshifts occurring in close proximity on the same allele resulting in the correct reading frame being maintained.

To check for polymorphism within *F. excelsior* at sites with evidence of convergence, we examined the combined BAM file generated from mapping Illumina HiSeq reads from 37 individuals from different European provenances (the European Diversity Panel) to the *F. excelsior* reference genome (BATG0.5) by Sollars et al<sup>11</sup>. Duplicate reads were removed from the BAM file using the “MarkDuplicates” function in picard tools v.1.139 (<http://broadinstitute.github.io/picard>), with the REMOVE\_DUPLICATES option set to “true”. Selected contigs (containing the genes of interest) were extracted from the BAM file using the “view” function in samtools v.1.6<sup>80</sup> and visualised with Tablet v.1.17.08.17<sup>84</sup>; evidence for polymorphism was observed directly from the reads and variants only recorded if supported by at least 10% of reads at that site. Loci OG39275 and OG46977 were excluded from this analysis due to errors in the reference gene models, possibly arising from misassembly, which meant the sites homologous to those with evidence of convergence between EAB-resistant taxa could not be identified (see Supplementary Table 5 for more details).

### ***Distinguishing between different underlying causes of convergent patterns***

To test whether evidence of convergence found by grand-conv might actually be due to taxa sharing the same amino acid variant as a result of introgression or incomplete lineage sorting (ILS), we conducted phylogenetic analyses of coding DNA sequences for



the candidate loci to infer their gene-trees. If introgression or ILS were the cause of the patterns detected by grand-conv, we would expect sequences from loci with apparently convergent residues to group together within their gene-tree, even when nucleotides encoding those residues are removed. The CDSs for the refined set of candidate loci were aligned with MUSCLE via GUIDANCE2 and alignment files with unreliably aligned codons removed were used for downstream analyses, as described above (see - Species-tree inference); none of datasets had sequences that were identified by GUIDANCE2 as being unreliably aligned. OG40061 failed to align due to the presence of an incomplete codon at the end of the reference gene model from *F. excelsior*; we trimmed the final 2bp from the *F. excelsior* sequence and reran GUIDANCE2 using this modified file.

Phased allele sequences generated using the WhatsHap results (see above - Analysis of variants within candidate loci) were added to the CDS alignments using MAFFT v.7.310<sup>85</sup> with the options "--add" and "--keeplength", in order to splice out any introns present in the phased sequences and maintain the original length of the CDS alignments. For any taxa for which phased sequences had been added, the original unphased sequence was removed from the alignment.

If intragenic recombination has taken place, gene-trees inferred from the CDS alignments may fail to group together the sequences with evidence of convergence even in cases where the convergent pattern is due to ILS or introgressive hybridisation. This is because the phylogenetic signal from any non-recombinant fragments of alleles derived from ILS or introgressive hybridisation may not be sufficiently strong to override that from fragments of alleles that have not been subject to these processes. To account for this possibility, we used hyphy v.2.3.14.20181030beta(MPI)<sup>86</sup> to conduct recombination tests with GARD<sup>87</sup> with the following parameter settings: 012345 "General Discrete" 3. Where GARD found significant evidence for a recombination breakpoint ( $p$ -value < 0.05), we partitioned the alignment into non-recombinant fragments for phylogenetic analysis.

Alignment files were converted to nexus format and gene-trees estimated with MrBayes, as described above (see - Species-tree inference). We checked the ASDSF and used Tracer v.1.6.0 (<http://beast.bio.ed.ac.uk/Tracer>) to inspect the ESS values for each parameter from the post burnin samples and to confirm that the burnin setting (i.e. discarding the first 10% of samples) was sufficient; in cases where runs had not

converged after 5M generations (ASDSF  $\geq 0.010$ ), additional generations were run until an ASDSF of  $< 0.010$  was reached. We examined the consensus trees generated by MrBayes to look for evidence that sequences sharing the amino acid states inferred as convergent by grand-conv cluster together in the gene-tree, in conflict with relationships inferred in the species-tree for *Fraxinus*. In cases where evidence of such clustering was found, the codon(s) corresponding to the amino acid site(s) at which evidence of convergence was detected were excluded and the MrBayes analysis repeated. In cases where sequences that have the “convergent” amino acid group together in the gene-tree even after the codon(s) for the relevant site(s) have been excluded, we concluded that the evidence of convergence detected by grand-conv is more likely due to introgressive hybridisation or ILS. We also examined the gene-tree topologies to assess whether any of the amino acid states identified as convergent by grand-conv is more likely to be the ancestral state for *Fraxinus*.

#### ***Further characterisation of candidate loci***

To identify the gene from *Arabidopsis thaliana* that best matches each of the candidate loci in our refined set, we conducted a TBLASTN search of the *F. excelsior* protein sequence belonging to the relevant OGs against the *A. thaliana* sequences in the nr/nt database in GenBank<sup>88</sup> and selected the hit with the lowest *E* value. In cases where the OG lacked a sequence from *F. excelsior*, we used the protein sequence from *F. mandshurica* as the query for the TBLASTN search instead.

We also checked for the presence of the *F. excelsior* sequences within the OrthoMCL clusters generated by Sollars et al.<sup>11</sup> to see if they were associated with the same *A. thaliana* genes as identified by BLAST. We obtained information on the function of the best matching *A. thaliana* genes from The Arabidopsis Information Resource (TAIR; <https://www.arabidopsis.org>) and the literature. The OrthoMCL analysis conducted by Sollars et al.<sup>11</sup> also included a range of other plant species, including *S. lycopersicum* (tomato) and the tree species *Populus trichocarpa* (poplar). As tomato is much more closely related to *F. excelsior* than is *A. thaliana*, and poplar is also a tree species, the function of the genes in these taxa may provide a better guide to the function of the *F. excelsior* genes. We therefore also checked the OrthoMCL clusters containing our candidate *F. excelsior* genes to identify putative orthologues, or close paralogues, from tomato and poplar. In cases where the OrthoMCL cluster included multiple tomato or

poplar genes, we focused attention on the tomato sequence that also belonged to the OMA group, as the putative orthologue of our *F. excelsior* gene in that species. For poplar, we looked for information on all sequences, unless there were a large number in the cluster (>4). We searched for literature on the function of the tomato and poplar genes, using the gene identifiers from the versions of the genome annotations used for the OrthoMCL analysis<sup>11</sup> and also looked for information on PhytoMine, in Phytozome 12<sup>66</sup>.

To further clarify the orthology/paralogy relationships between our candidates and genes from other species, we conducted phylogenetic analysis of the relevant OrthoMCL clusters from Sollars et al<sup>11</sup> for selected loci. Protein sequences belonging to each OrthoMCL cluster were aligned and gene-trees inferred using GUIDANCE2 and MrBayes respectively, as described above for the OGs and HOGs. For the OrthoMCL cluster relating to OG15551, following an initial MrBayes analysis, we removed two incomplete sequences (Migut.O00792.1.p and GSVIVT01025800001, which were missing >25% of characters in the alignment) and two divergent sequences from *A. thaliana* (AT1G74540 and AT1G74550) which are known to derive from a Brassicales-specific retroposition event and subsequent Brassicaceae-specific tandem duplication<sup>89</sup>; the alignment and phylogenetic analysis was then repeated for the reduced dataset.

For OG15551, we generated a sequence logo for regions of the protein containing sites at which evidence of convergence was detected. We obtained putatively homologous sequences by downloading the fasta file for the OMA group (OMA Browser fingerprint YGPIYSF<sup>90</sup>) containing the *A. thaliana* *CYP98A3* gene (AT2G40890); the sequences were filtered to include only those from angiosperms, with a maximum of one sequence per genus retained (29 genera in total). To this dataset, we added the OG15551 protein sequences for *F. mandshurica* and *F. pennsylvanica* pe-48 and manually aligned the regions containing the relevant sites (positions 208-218 and 474-482 in the *F. excelsior* FRAEX38873\_v2\_000261700 reference protein). We used WebLogo v.3.7.3<sup>91</sup> without compositional adjustment to generate logos for each of these regions.

### ***GO term enrichment analysis***

To test for the possibility of overrepresentation of particular functional categories among the candidate loci in our refined set, compared with the complete set of genes used as input for the convergence analyses, we conducted gene ontology (GO) enrichment tests.

Fisher's exact tests with the "weight" and "elim" algorithms, which take into account the GO graph topology<sup>92</sup>, were run using the topGO package<sup>93</sup> (v.2.32.0) in R v.3.5.1<sup>94</sup>. We created a "genes-to-GOs" file for the complete set of *F. excelsior* gene models included in the grand-conv analyses, using GO terms from the existing functional annotation for the reference genome (Sollars et al.<sup>11</sup>); only the single longest transcript per gene (see <http://www.ashgenome.org/transcriptomes>) was included and for any OMA groups that lacked an *F. excelsior* sequence we used the reference model referred to by the majority of other *Fraxinus* sequences in the group (i.e. as indicated in the GeMoMa gene model names). We also created a list of *F. excelsior* reference model genes belonging to our refined set of candidate loci; again, for any OMA groups that lacked an *F. excelsior* sequence, we used the reference model referred to by the majority of other *Fraxinus* sequences in the group. The complete list of *F. excelsior* reference genes included in the grand-conv analyses, and their associated GO terms, was used as the background against which the list of gene models from the refined set of candidate loci was tested. Fisher's exact test was run separately, with each of the algorithms, to check for enrichment of terms within the biological process (BP), molecular function (MF) and cellular component (CC) domains.

### ***Protein modelling***

The SignalP 5.0 server<sup>95</sup> and Phobius server<sup>96</sup> (<http://phobius.sbc.su.se/index.html>) were used to detect the presence of signal peptides; for SignalP the organism group was set to "Eukarya" and for Phobius the "normal prediction" method was used. All *Fraxinus* sequences belonging to the OMA groups were used as input for the signal peptide analyses; we only concluded that a signal peptide was present if it was predicted by both methods. The NetPhos 3.1 Server (<http://www.cbs.dtu.dk/services/NetPhos/>) was used with default settings to identify candidate phosphorylation sites for loci where the amino acid variant observed at a site with evidence of convergence included a serine, threonine or tyrosine. The same protein sequences for low and high susceptibility taxa as used for protein modelling (see below) were input to an initial run of NetPhos 3.1; where evidence for phosphorylation site presence/absence was detected with this initial sequence pair (i.e. present in the sequence with the convergent state and absent from that with the non-convergent state, or vice versa) we reran NetPhos 3.1 on all *Fraxinus* sequences from the relevant OMA groups to test if this difference was consistently associated with the convergent/non-convergent state. We only counted as potential phosphorylation sites

those for which the NetPhos score for phosphorylation potential was  $\geq 0.900$  for all sequences with the putative site.

RaptorX-Binding<sup>97</sup> (<http://raptorx.uchicago.edu/BindingSite/>) was used to generate predicted protein models for each of the candidate genes in our refined set, as well as to outline possible binding sites and candidate ligands. Protein sequences for gene models from the *F. excelsior* reference genome were used for initial protein model and binding site prediction, except in cases where *F. excelsior* was not present in the OMA group or where comparison with the other ingroup and outgroup taxa indicated the *F. excelsior* gene model may be incorrect/incomplete; for these loci, the *F. mandshurica* sequences were used instead as, after the reference, the genome assembly for this taxon is one of the highest quality available. For loci for which a binding site could be successfully predicted (i.e. with at least one potential binding site with a pocket multiplicity value of  $\geq 40$ ), additional models were generated for representative resistant (*F. mandshurica* or *F. platypoda*) and susceptible (*F. ornus* or susceptible *F. pennsylvanica*) taxa using Swiss-model<sup>98</sup> and Phyre2<sup>99</sup> (intensive mode), with the exact taxon selection depending on which grand-conv pairwise comparison the locus was detected in (see above - Analysis of patterns of sequence variation consistent with molecular convergence) and which taxa had complete gene models. Where errors were detected in the predicted protein sequences for resistant or susceptible taxa (i.e. due to errors in the predicted gene model, which were detected through comparison with sequences from other species, including those from outgroups) these were corrected prior to modelling (e.g. by trimming extra sequence resulting from incorrect prediction of the start codon). Models predicted by the three independent methods (RaptorX-Binding, Swiss-modeller and Phyre2) were compared by aligning them using PyMOL v.2.0 with the align function to check for congruence; only those loci whose models displayed congruence and where the convergent site was located within/close to the putative active site were taken forward for predictive ligand docking analysis (using the Phyre2 and RaptorX-Binding models for the docking itself). In addition, any loci with congruent models where the site with evidence of convergence is also a putative phosphorylation site presence/absence variant, or which are within a putative functional domain, were analysed further. Ligand candidates were selected based on relevant literature and/or the RaptorX-Binding output, with SDF files for each of the molecules being obtained from PubChem (<https://pubchem.ncbi.nlm.nih.gov>). SDF files were converted to 3d pdb files using

### Online SMILES Translator and Structure File Generator

(<https://cactus.nci.nih.gov/translate/>), so they could be used with Autodock. Docking analysis was carried out using Autodock Vina v.1.1.2<sup>100</sup> with the GUI PyRx v.0.8<sup>101</sup>. Following docking, ligand binding site coordinates were exported as SDF files from Pyrex and loaded into PyMOL with the corresponding protein model file for the resistant and susceptible taxa. Binding sites were then annotated and the residues at which evidence for convergence had been detected with grand-conc were labelled.

### ***Evidence for differential expression of candidate loci in *F. pennsylvanica****

We used published transcriptome assembly and expression data from *F. pennsylvanica*<sup>25</sup> to look for evidence of differential expression of our candidate loci in response to EAB larval feeding. This dataset comprised six genotypes of *F. pennsylvanica*, four putatively resistant to EAB and two susceptible to EAB. To identify the orthologues of our genes in the protein sequences of this independently assembled transcriptome<sup>25</sup>, we repeated the OMA clustering analysis (see above - Gene annotation and orthologue inference) with the addition of these data, available as “*Fraxinus\_pennsylvanica\_120313\_peptides*” at the Harwood Genomics Project website (<https://hardwoodgenomics.org>). OMA was run as described above, with the SpeciesTree parameter set to ‘estimate’; because we only intended to use the results for the OGs, and not the HOGs, from this analysis, we did not repeat the clustering with a modified species-tree as was done for our main OMA analysis. Having identified the likely orthologous loci from the *F. pennsylvanica* transcriptome<sup>25</sup>, we used the results of the differential expression analysis<sup>25</sup> to check whether our candidate loci had significantly increased or decreased expression post-EAB feeding in this dataset.

### **Data availability**

Underlying data for Figure 1 are available in Supplementary Tables 1 and 2. All Illumina sequence data and genome assemblies will be submitted to the European Nucleotide Archive, and the accession number provided, prior to publication. VCF files containing variants called for *Fraxinus* individuals will be submitted to the European Variation Archive, and the accession number provided, prior to publication. All other data are available from the corresponding authors upon reasonable request.

### **Code availability**

All unpublished code is available upon reasonable request from the corresponding authors.

**Acknowledgements** This research utilised Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT (<http://doi.org/10.5281/zenodo.438045>). We thank J. Carlson for providing *F. pennsylvanica* DNA; T. Baxter, S. Brockington, P. Brownless, D. Crowley, S. Honey, R. Irvine, R. Jinks, P. Jones, T. Kirkham, H. McAllister, I. Parkinson and S. Redstone for help with obtaining *Fraxinus* materials from UK collections; T. Poland for providing EAB eggs; M. Miller for propagating trees for the bioassays; R. Matko for preparation of voucher specimens; J. Pellicer for advice on flow cytometry; P. Howard and M. Struebig for advice on preparation of DNA extractions; J. Keilwagen for help with GeMoMa; K. Davies and J. Parker for help with software for convergence analyses; members of the Evolution Labchat group and Rossiter Lab at Queen Mary University of London, for discussions; R. Rose and J. Sayers for advice on protein modelling analyses. This project was funded by the Living with Environmental Change (LWEC) Tree Health and Plant Biosecurity Initiative - Phase 2 (grant number BB/L012162/1) funded jointly by the BBSRC, Defra, Economic and Social Research Council, Forestry Commission, NERC and the Scottish Government. R.J.A.B. acknowledges support from the DEFRA Future Proofing Plant Health scheme. R.J.A.B. and L.J.K. acknowledge support from the Erica Waltraud Albrecht Endowment Fund.

**Author contributions** R.J.A.B. conceived and oversaw the project. L.J.K. and R.J.A.B. wrote the manuscript, with input from J.L.K., W.J.P and S.J.R. L.J.K. conducted gene annotation, orthologue inference, the convergence analyses, calling and analysis of variants, the GO enrichment analysis and all phylogenetic analyses. L.J.K., W.C. and A.T.W. performed genome size estimation by flow cytometry. L.J.K., W.C., E.D.C. and D.W.C. extracted DNA. L.J.K. and E.D.C. assembled the genomes. J.L.K. conceived and oversaw the emerald ash borer (EAB) bioassays. D.W.C. conducted the EAB bioassays. J.L.K. and M.E.M. analysed data from the EAB bioassays. W.J.P. conducted the protein modelling analyses. S.J.R. advised on convergence analyses.

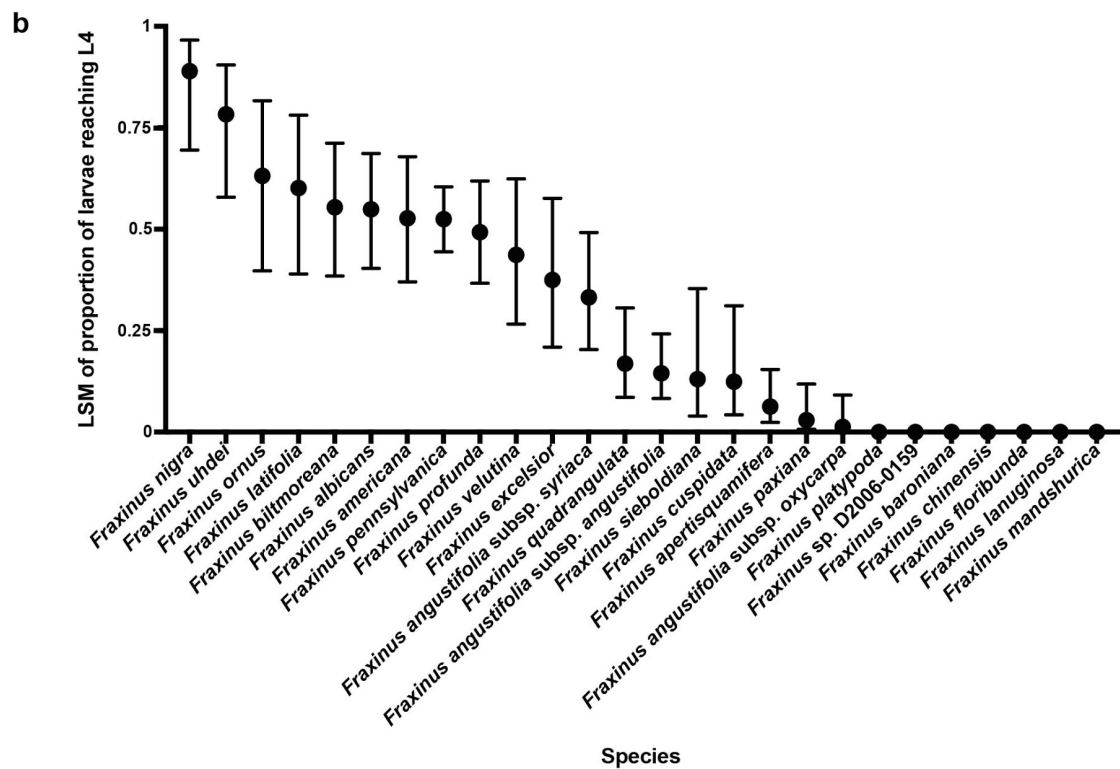
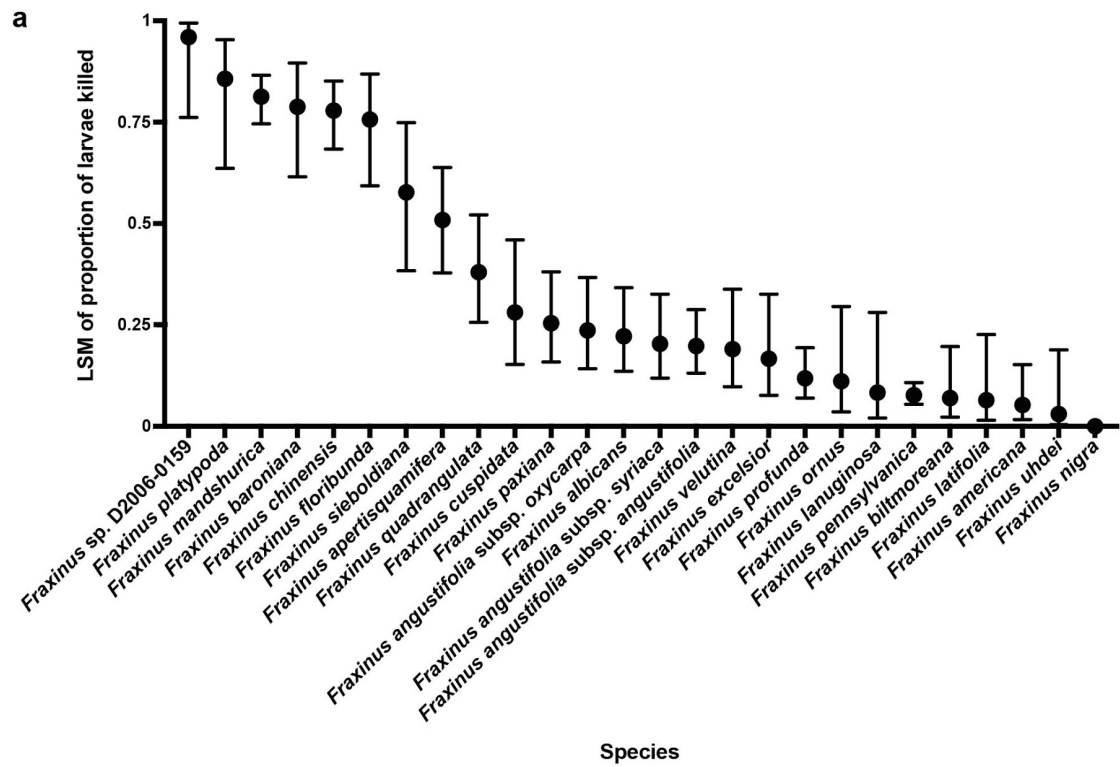
**Competing interests** The authors declare no competing interests.

**Correspondence and requests for materials** should be addressed to L.J.K. and R.J.A.B; requests for DNA samples or plant materials may require the signing of a material transfer agreement, and permission from the original source in cases of intended commercial use.

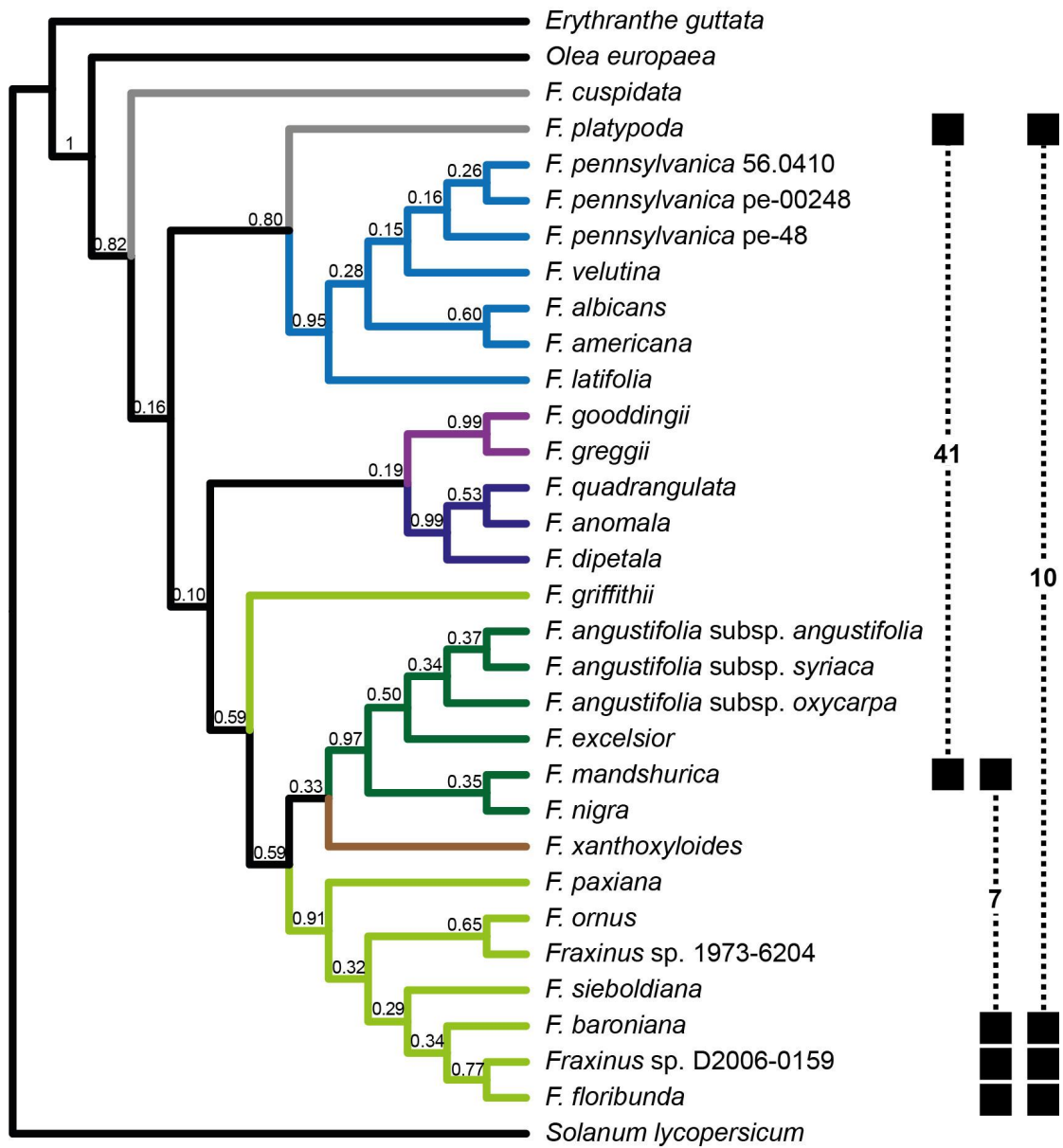


## Figures

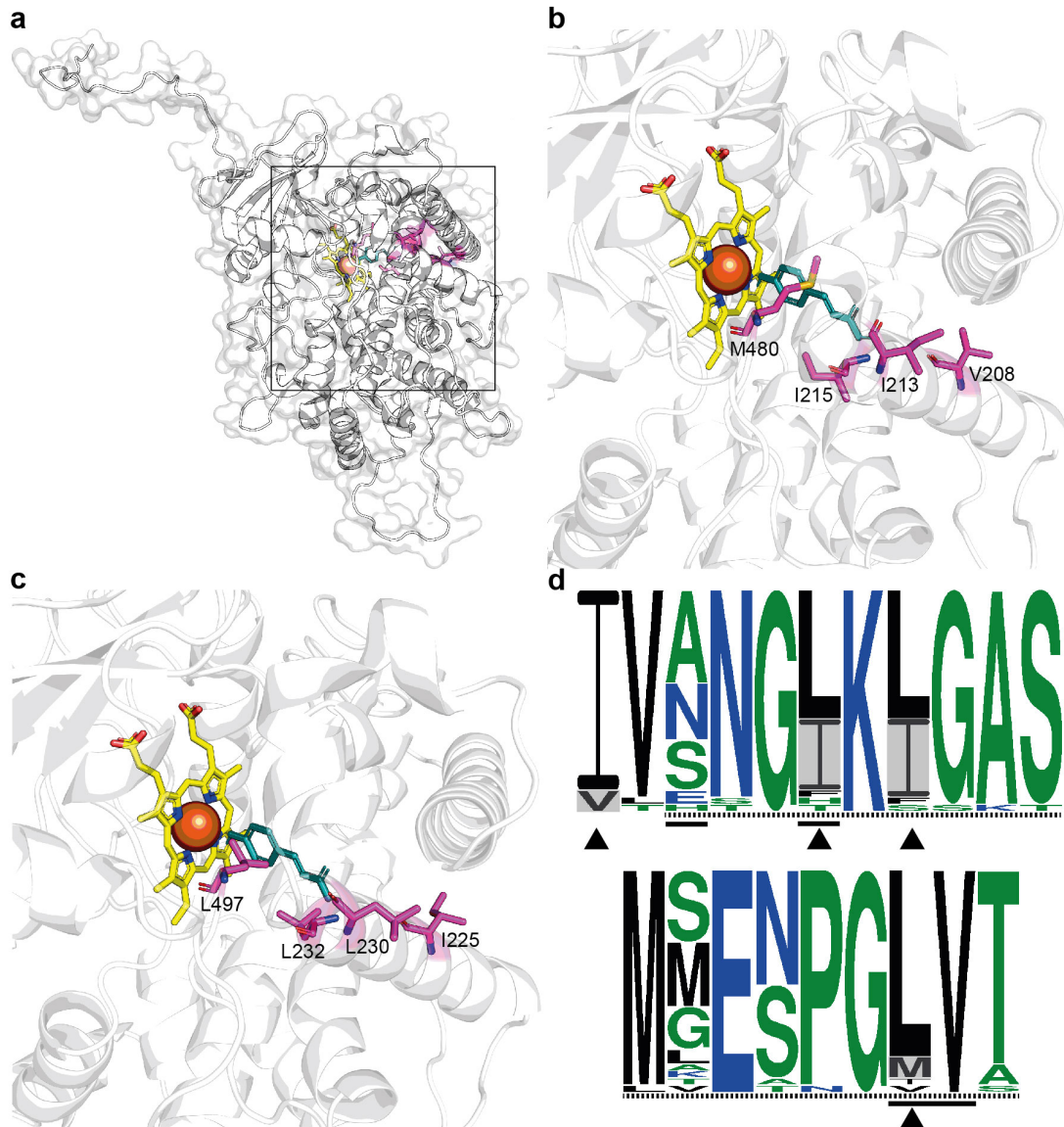
**Figure 1. *Fraxinus* species' resistance to EAB in bioassays. a, b,** Measures of resistance of different *Fraxinus* taxa to EAB larvae. The *x* axis shows taxa tested; the *y* axis shows least squares means (LSM) of the proportion of larvae successfully entering the tree that were killed by a host defence response (**a**) or LSM of the proportion of larvae successfully entering the tree that reached the L4 instar (**b**). The error bars represent 95% confidence intervals. *Fraxinus* sp. D2006-0159 is a genotype from China for which we could not determine a recognised species name. *Fraxinus biltmoreana*, *F. chinensis*, *F. lanuginosa*, *F. profunda* and *F. uhdei* are polyploids and were not included in the genomic analyses; *F. apertisquamifera* was also not included.



**Figure 2. Species-tree for the genus *Fraxinus*.** Primary concordance tree inferred from 272 phylogenetically informative loci found in all taxa, inferred via Bayesian concordance analysis with BUCKy. Taxonomic sections within *Fraxinus*, according to Wallander<sup>10</sup>, are shown in different colours: dark blue - section *Dipetalae*; dark green - section *Fraxinus*; light blue - section *Melioides*; light green - section *Ornus*; purple - section *Pauciflorae*; brown - section *Sciadhanthus*. *Fraxinus* species not placed into a specific section (*incertae sedis*) are coloured grey and outgroups black. Numbers above the branches are sample-wide concordance factors. Filled squares (linked by dashed lines) indicate the resistant taxa included in the three pairwise convergence analyses, with the number of candidate genes found from that comparison shown; numbers do not sum to 53 (i.e. the total number of candidate genes) because some genes were identified by more than one pairwise comparison.



**Figure 3. Predicted protein structure for OG15551.** **a**, Predicted structure for OG15551, modelled using the protein sequence for the EAB-resistant species *Fraxinus mandshurica*. The black box indicates the region containing the active site, which is enlarged in **b** and **c**. **b**, Region containing the predicted active site in *F. mandshurica*, showing the four amino acid sites at which evidence for convergence between EAB-resistant species was detected. The putative substrate, p-Coumarate, is shown in blue and the heme cofactor in yellow. **c**, Region containing the predicted active site in the EAB-susceptible *F. pennsylvanica* pe-48, showing the amino acid states found at the four sites at which evidence for convergence between EAB-resistant species was detected; the putative substrate and cofactor are shown as in **b**. **d**, Sequence logos for OG15551 and putatively homologous sequences from other angiosperms for regions containing sites at which evidence of convergence was detected (positions 208-218 (top) and 474-482 (bottom) in the *F. excelsior* reference protein), showing the degree of sequence conservation across 30 genera. The height of each residue indicates its relative frequency at that site; amino acids are coloured according to their hydrophobicity (blue = hydrophilic; black = hydrophobic; green = neutral). Dashed lines indicate substrate recognition sites and solid lines residues that are predicted to contact the substrate in the *A. thaliana* CYP98A3 protein<sup>19</sup>; arrowheads indicate sites at which evidence of convergence between EAB-resistant taxa was detected and grey shading shows the amino acid states associated with resistance.



## **SUPPLEMENTARY INFORMATION**

### **Supplementary Tables:**

Supplementary Table 1. Results from EAB egg bioassays

Supplementary Table 2. Results of statistical analysis of bioassay results

Supplementary Table 3. Details of plant material used for whole genome sequencing and egg bioassays

Supplementary Table 4. Summary of genome size estimates for individuals used for whole genome sequencing, Illumina sequence data generated and genome assembly statistics

Supplementary Table 5. Details for the refined set of 53 candidate loci identified by the grand-conv analyses

Supplementary Table 6. Evidence for potential loss-of-function mutations in the candidate genes

Supplementary Table 7. Significantly enriched GO terms among the candidate genes

Supplementary Table 8. Presence of amino acid states associated with EAB-resistance in the 53 candidate genes across 29 *Fraxinus* genomes

Supplementary Table 9. Positions of candidate amino acid sites in *Fraxinus excelsior* diversity panel and evidence for polymorphism

Supplementary Table 10. Putative orthologues identified from the *Fraxinus pennsylvanica* transcriptome and evidence for differential expression of transcripts

### **Supplementary Figures:**

Supplementary Figure 1. Gene-trees for the 53 candidate loci inferred from CDS alignments including phased alleles and accounting for evidence of recombination

Supplementary Figure 2. Predicted protein structures for selected candidate loci

Supplementary Figure 3. Gene-trees for OrthoMCL clusters including *Fraxinus excelsior* sequences from selected candidate genes

### **Supplementary Notes:**

Supplementary Note 1. Species-tree for *Fraxinus*.

Supplementary Note 2. Results of grand-conv analyses.

Supplementary Note 3. Evidence for loss-of-function mutations.

Supplementary Note 4. Details of candidate loci.

Supplementary Note 5. Significant enrichment of GO terms.

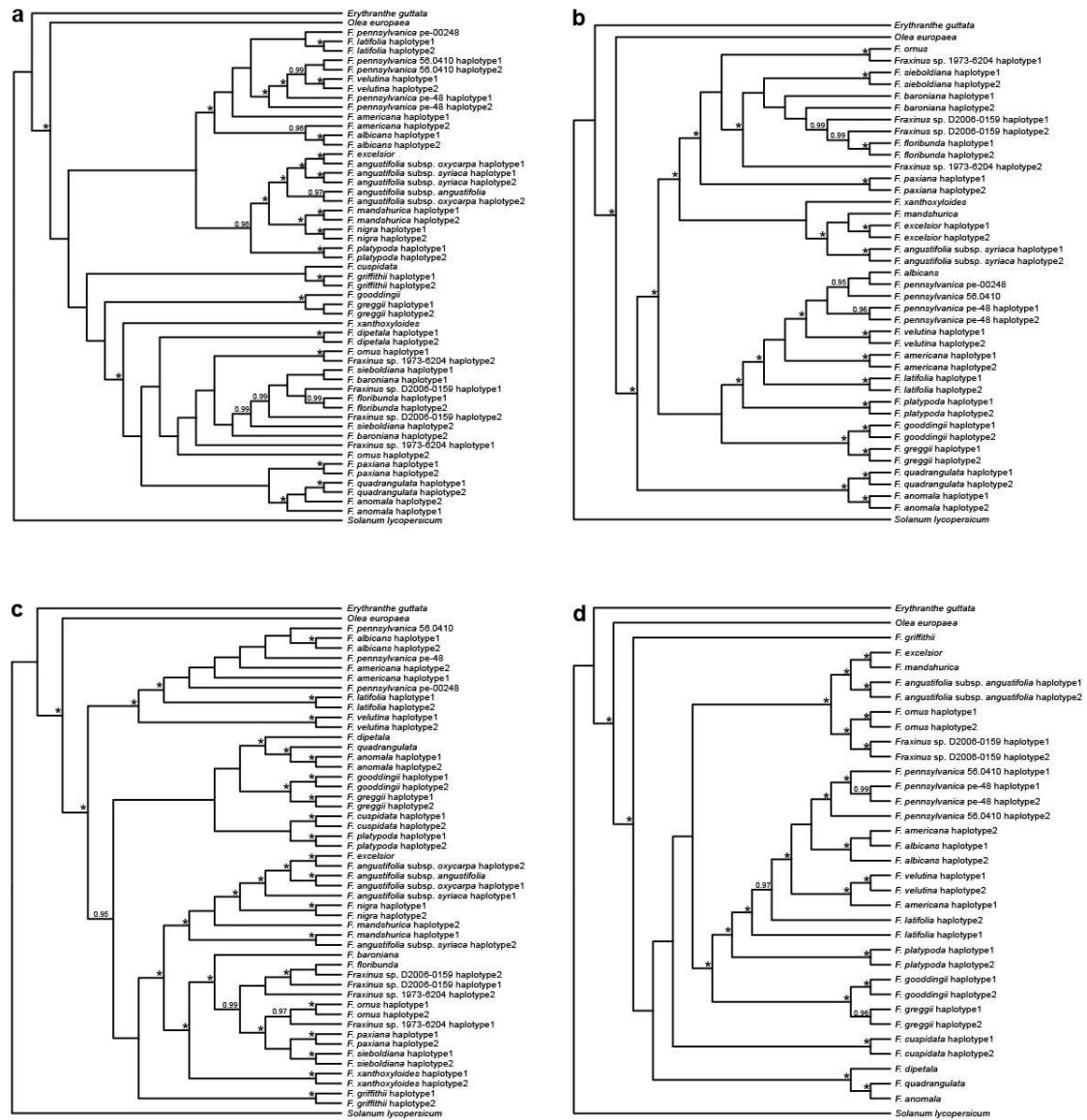
## Supplementary Note 6. Differential expression in response to EAB-feeding.

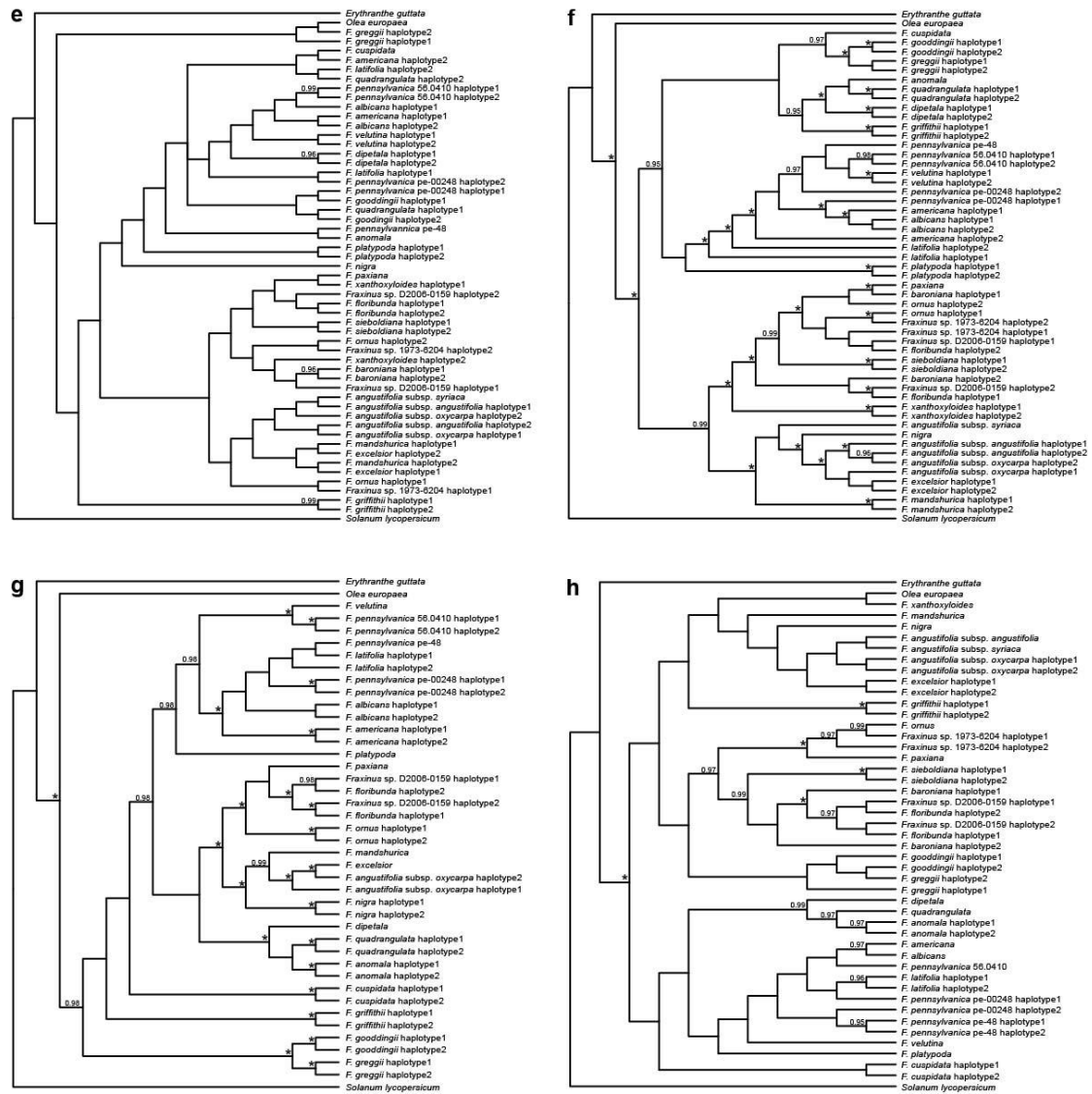


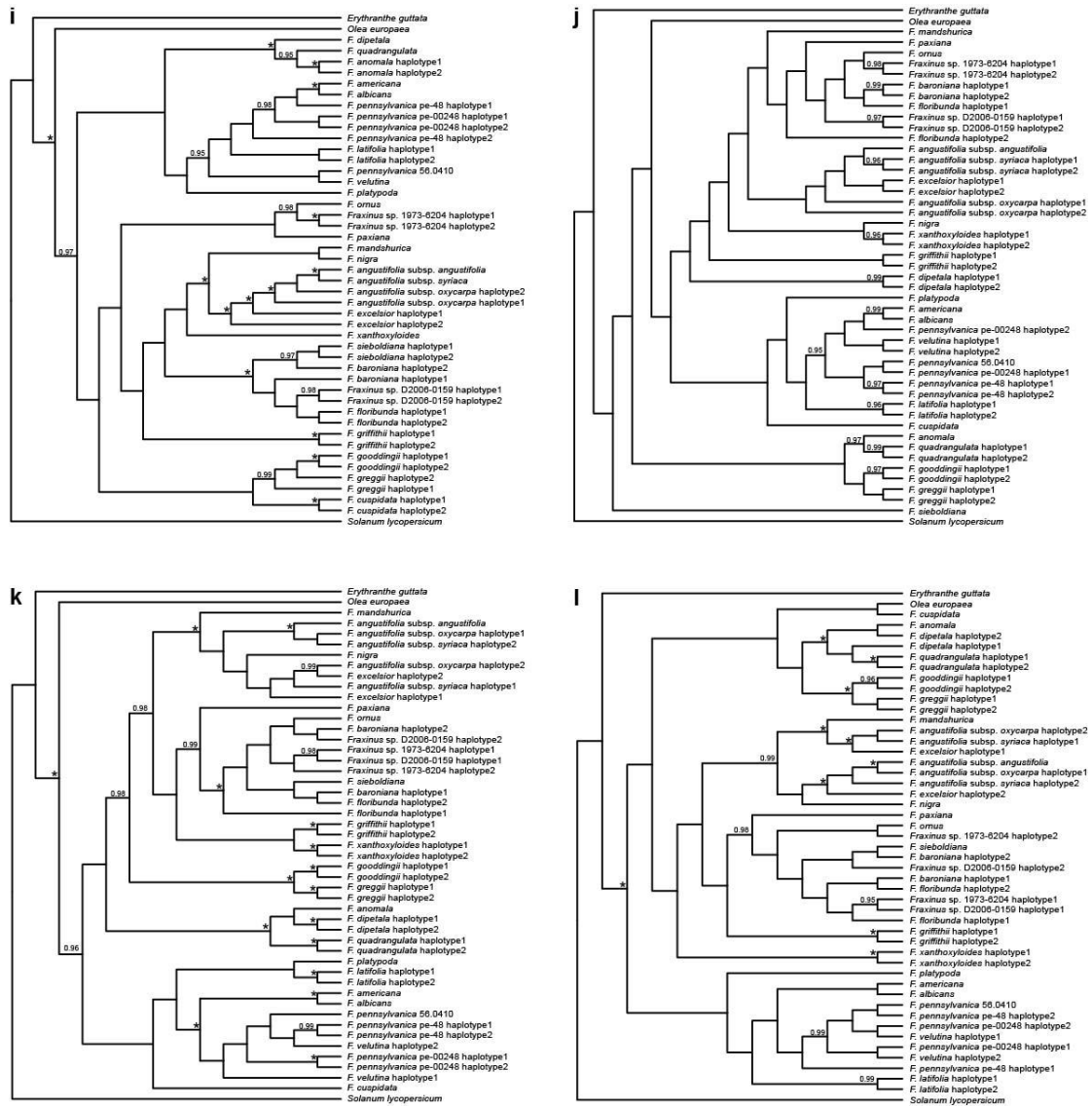
**Supplementary Figure 1. Gene-trees for the 53 candidate loci inferred from CDS alignments including phased alleles and accounting for evidence of recombination.**

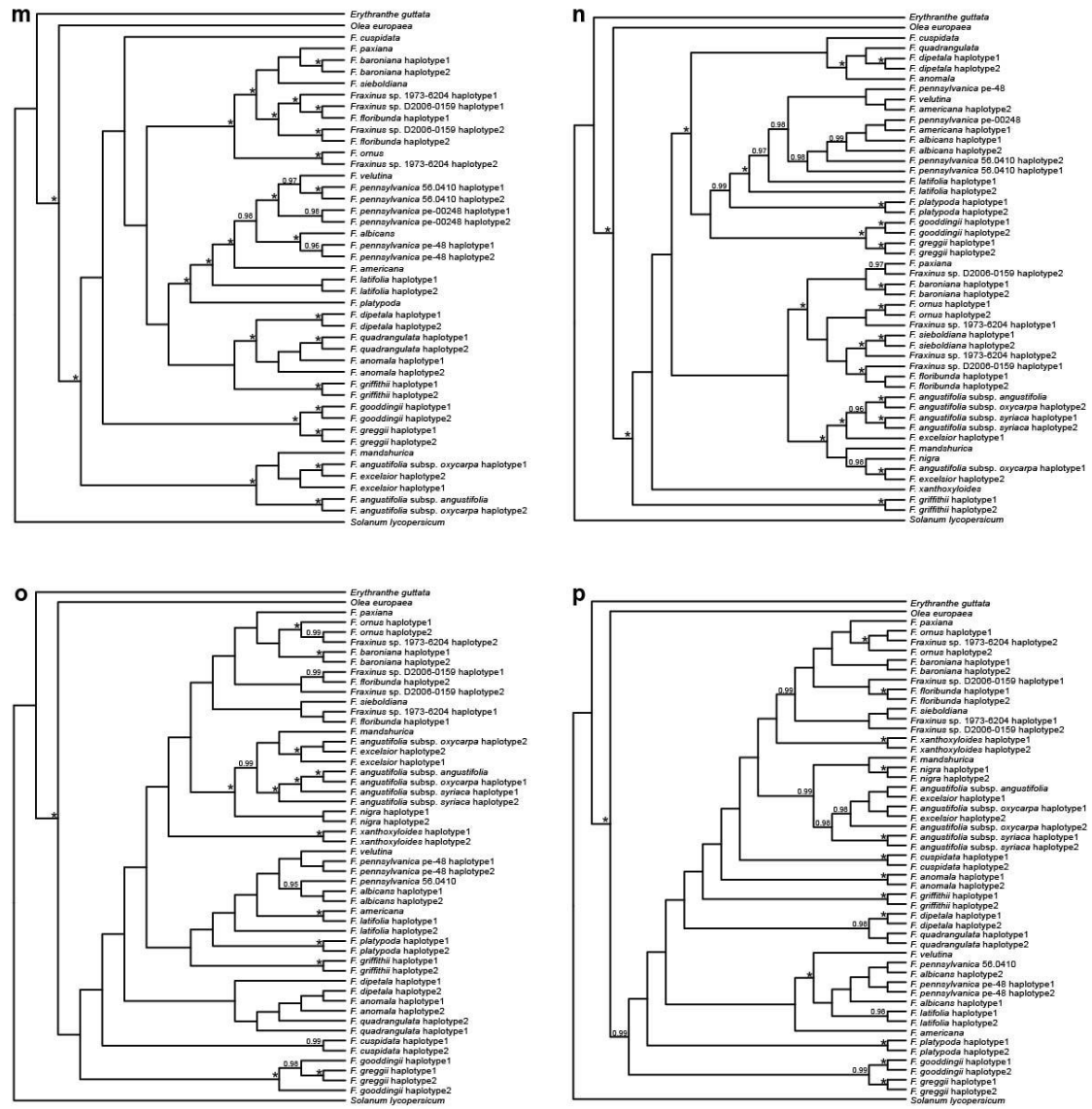
**a**, Gene-tree for OG20252 inferred from positions 378-1524 of the CDS alignment excluding the codon for the amino acid site with evidence of convergence (the gene-tree inferred from positions 1-377 is not shown because the majority of nodes are poorly supported and the codon for the convergent site is within the other fragment). **b**, Gene-tree for OG853. **c**, Gene-tree for OG2897. **d**, Gene-tree for OG4372. **e**, Gene-tree for OG4469 inferred from positions 1-250 of the CDS alignment. **f**, Gene-tree for OG4469 inferred from positions 251-2595 of the CDS alignment. **g**, Gene-tree for OG5539. **h**, Gene-tree for OG6935 inferred from positions 1-794 of the CDS alignment. **i**, Gene-tree for OG6935 inferred from positions 795-2208 of the CDS alignment. **j**, Gene-tree for OG7454 inferred from positions 1-557 of the CDS alignment. **k**, Gene-tree for OG7454 inferred from positions 558-1405 of the CDS alignment. **l**, Gene-tree for OG7454 inferred from positions 1406-2352 of the CDS alignment excluding the codons for the amino acid sites with evidence of convergence. **m**, Gene-tree for OG10762 inferred from the CDS alignment with the codon for the amino acid site with evidence of convergence excluded. **n**, Gene-tree for OG11013. **o**, Gene-tree for OG11720 inferred from positions 1-868 of the CDS alignment excluding the codon for the amino acid site with evidence of convergence. **p**, Gene-tree for OG11720 inferred from positions 869-1773 of the CDS alignment. **q**, Gene-tree for OG13887 inferred from positions 1-269 of the CDS alignment. **r**, Gene-tree for OG13887 inferred from positions 270-1803 of the CDS alignment. **s**, Gene-tree for OG15551 inferred from the CDS alignment with the codons for the amino acid sites with evidence of convergence excluded. **t**, Gene-tree for OG16673 inferred from the CDS alignment with the codon for the amino acid site with evidence of convergence excluded. **u**, Gene-tree for OG16739. **v**, Gene-tree for OG17252. **w**, Gene-tree for OG19104. **x**, Gene-tree for OG20520 inferred from positions 1-372 of the CDS alignment. **y**, Gene-tree for OG20520 inferred from positions 373-1452 of the CDS alignment. **z**, Gene-tree for OG20859. **aa**, Gene-tree for OG21033 inferred from positions 1-327 of the CDS alignment. **ab**, Gene-tree for OG21033 inferred from positions 328-1539 of the CDS alignment. **ac**, Gene-tree for OG21449 inferred from positions 1-455 of the CDS alignment. **ad**, Gene-tree for OG21449 inferred from positions 456-1209 of the CDS alignment. **ae**, Gene-tree for OG23214 inferred from positions 1-836 of the CDS alignment. **af**, Gene-tree for OG23214 inferred from positions 837-1401 of the CDS alignment. **ag**, Gene-tree for OG23284. **ah**, Gene-tree for

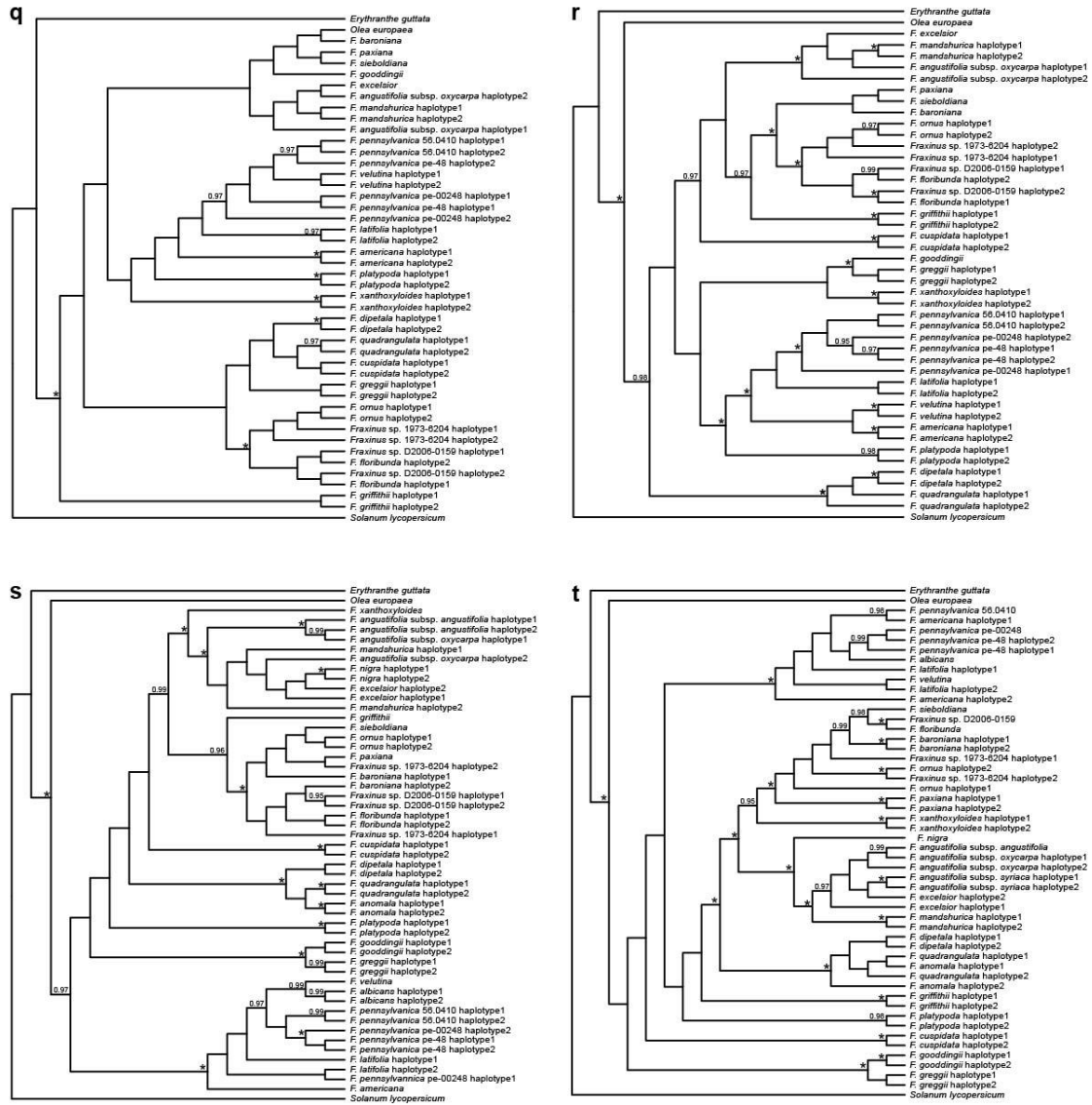
OG24614. **ai**, Gene-tree for OG24969. **aj**, Gene-tree for OG26964 inferred from positions 1-205 of the CDS alignment. **ak**, Gene-tree for OG26964 inferred from positions 206-1092 of the CDS alignment. **al**, Gene-tree for OG27080. **am**, Gene-tree for OG27693. **an**, Gene-tree for OG27838 inferred from positions 1-381 of the CDS alignment. **ao**, Gene-tree for OG27838 inferred from positions 382-1707 of the CDS alignment. **ap**, Gene-tree for OG28712. **aq**, Gene-tree for OG30208. **ar**, Gene-tree for OG32176 inferred from the CDS alignment with the codons for the amino acid sites with evidence of convergence excluded. **as**, Gene-tree for OG33348. **at**, Gene-tree for OG35707. **au**, Gene-tree for OG36502. **av**, Gene-tree for OG37560. **aw**, Gene-tree for OG37870. **ax**, Gene-tree for OG38407. **ay**, Gene-tree for OG38543. **Az**, Gene-tree for OG39275. **aaa**, Gene-tree for OG40061. **aab**, Gene-tree for OG41448 inferred from positions 1-243 of the CDS alignment. **aac**, Gene-tree for OG41448 inferred from positions 244-981 of the CDS alignment excluding the codon for the amino acid site with evidence of convergence. **aad**, Gene-tree for OG41488. **aae**, Gene-tree for OG43828. **aaf**, Gene-tree for OG46977 inferred from positions 1-291 of the CDS alignment. **aag**, Gene-tree for OG46977 inferred from positions 292-804 of the CDS alignment. **aaH**, Gene-tree for OG47560 inferred from positions 1-280 of the CDS alignment. **aaI**, Gene-tree for OG47560 inferred from positions 281-852 of the CDS alignment excluding the codons for the amino acid sites with evidence of convergence. **aaJ**, Gene-tree for OG47629. **aaK**, Gene-tree for OG49074. **aaL**, Gene-tree for OG50989. **aaM**, Gene-tree for OG56563. **aaN**, Gene-tree for OG59564 inferred from positions 1-189 of the CDS alignment. **aaO**, Gene-tree for OG59564 inferred from positions 190-558 of the CDS alignment excluding the codon for the amino acid site with evidence of convergence. **aaP**, Gene-tree for OG60899. **aaQ**, Gene-tree for OG64545. Gene-trees were inferred with MrBayes and rooted on *Solanum lycopersicum*. Numbers above branches are posterior probabilities (PP) of  $\geq 0.95$ ; asterisks indicate nodes with PP=1.

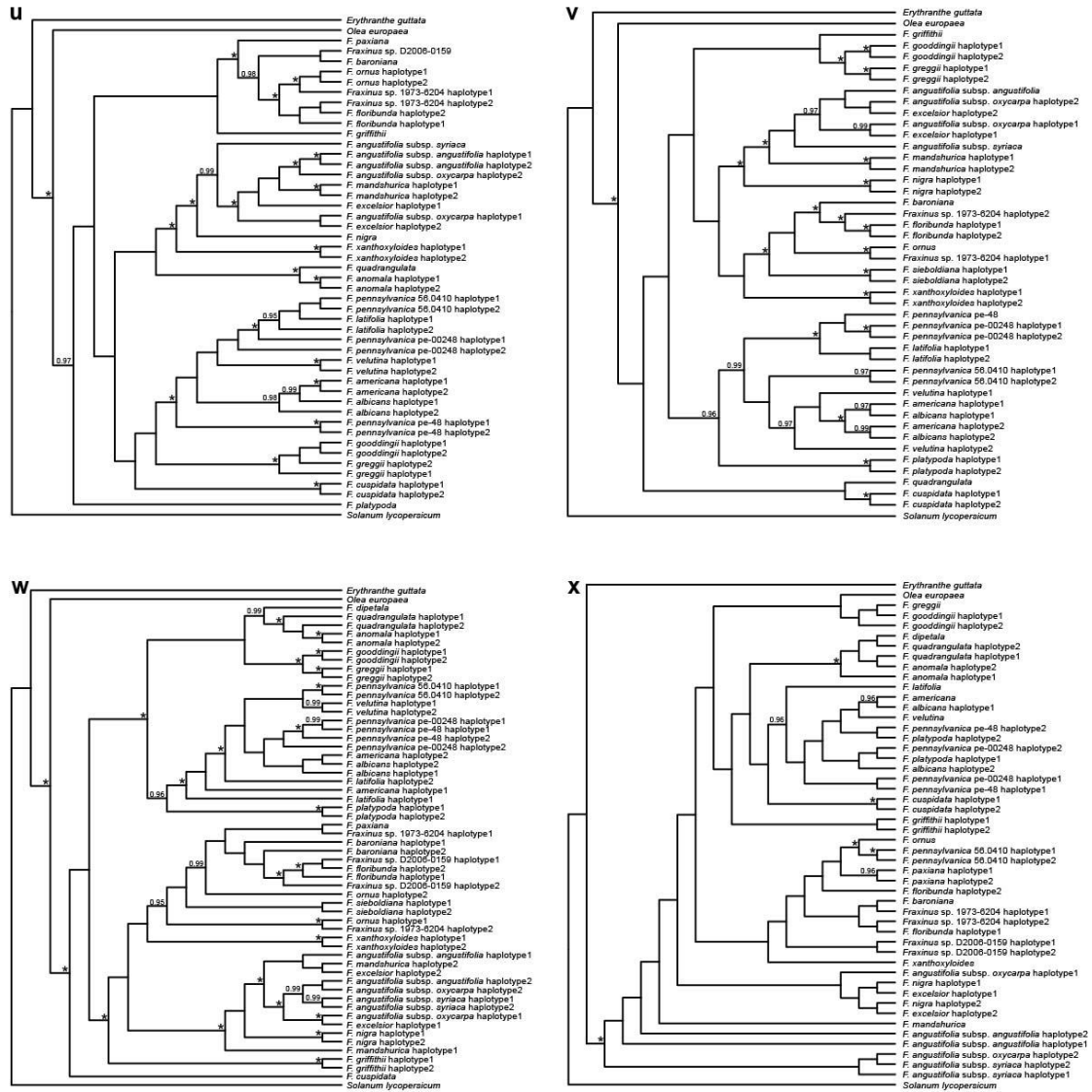




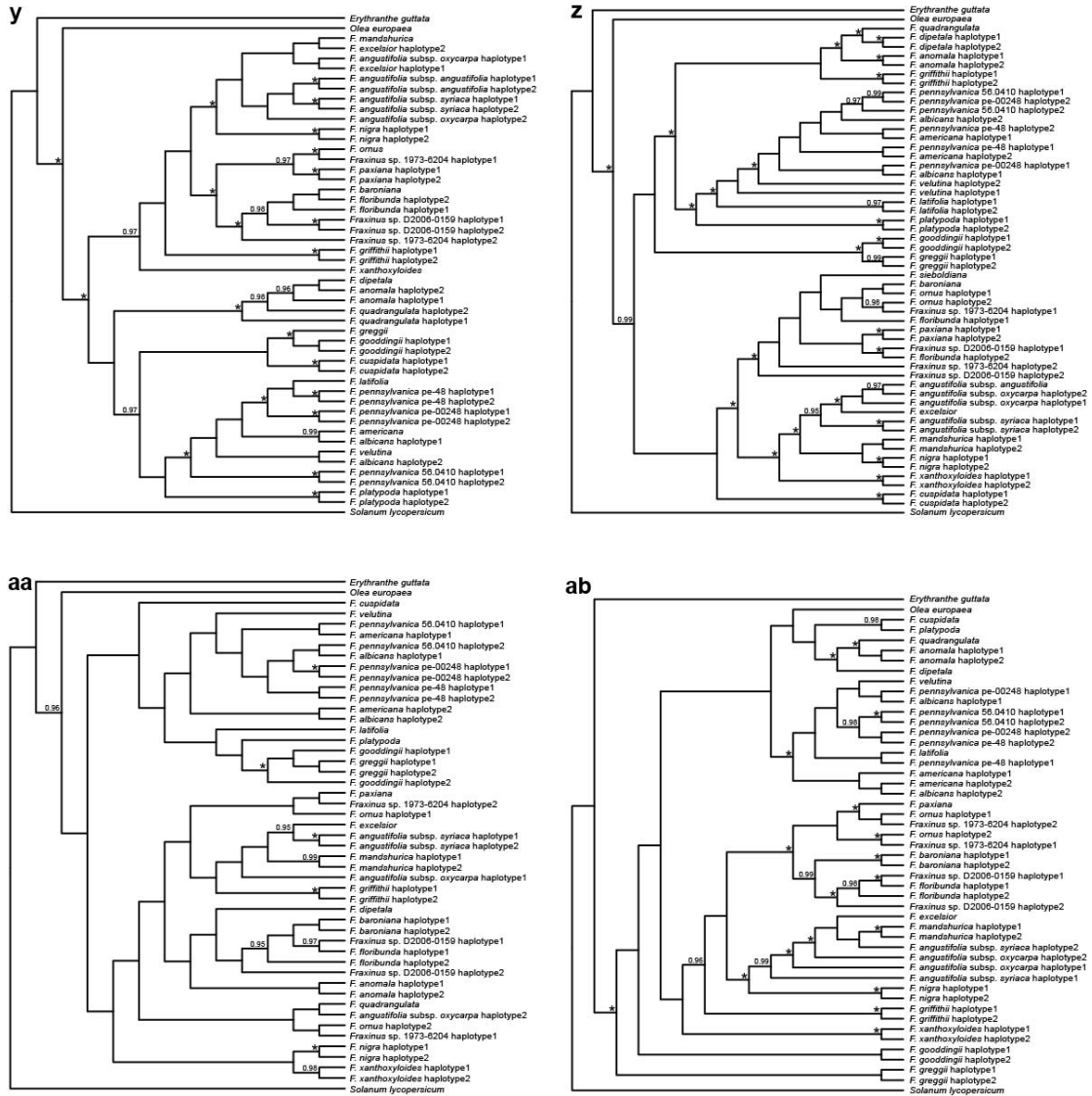


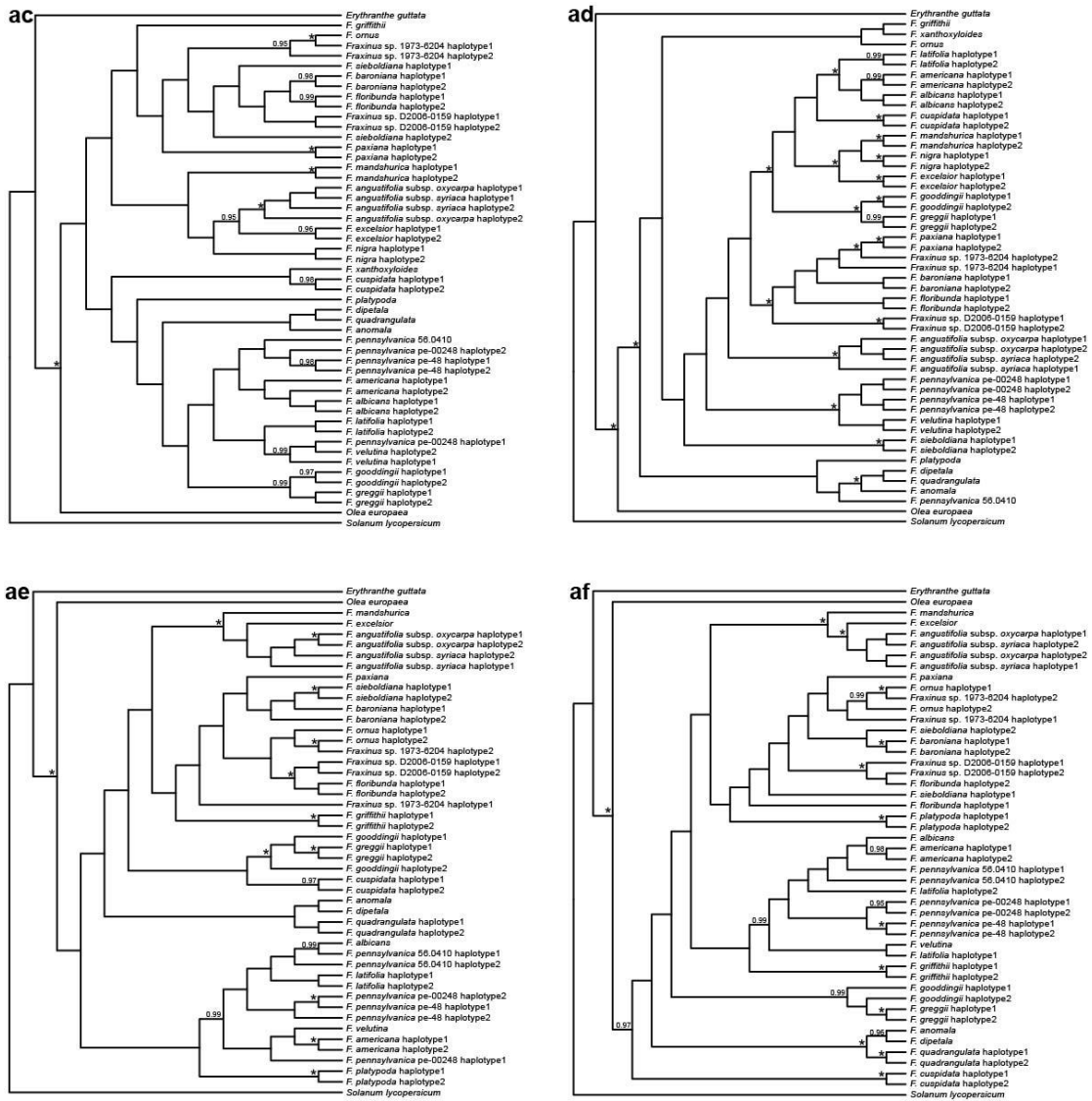


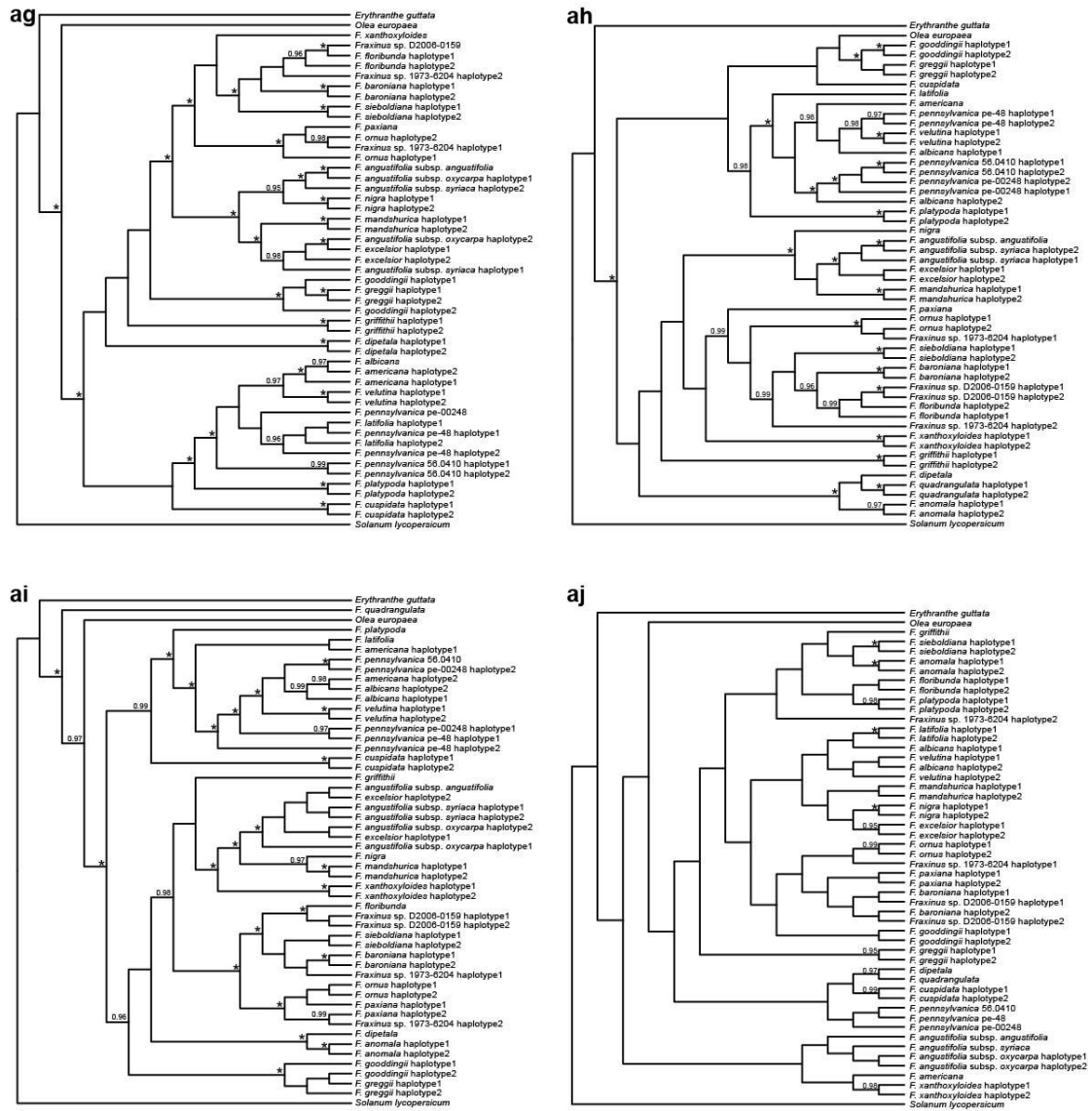


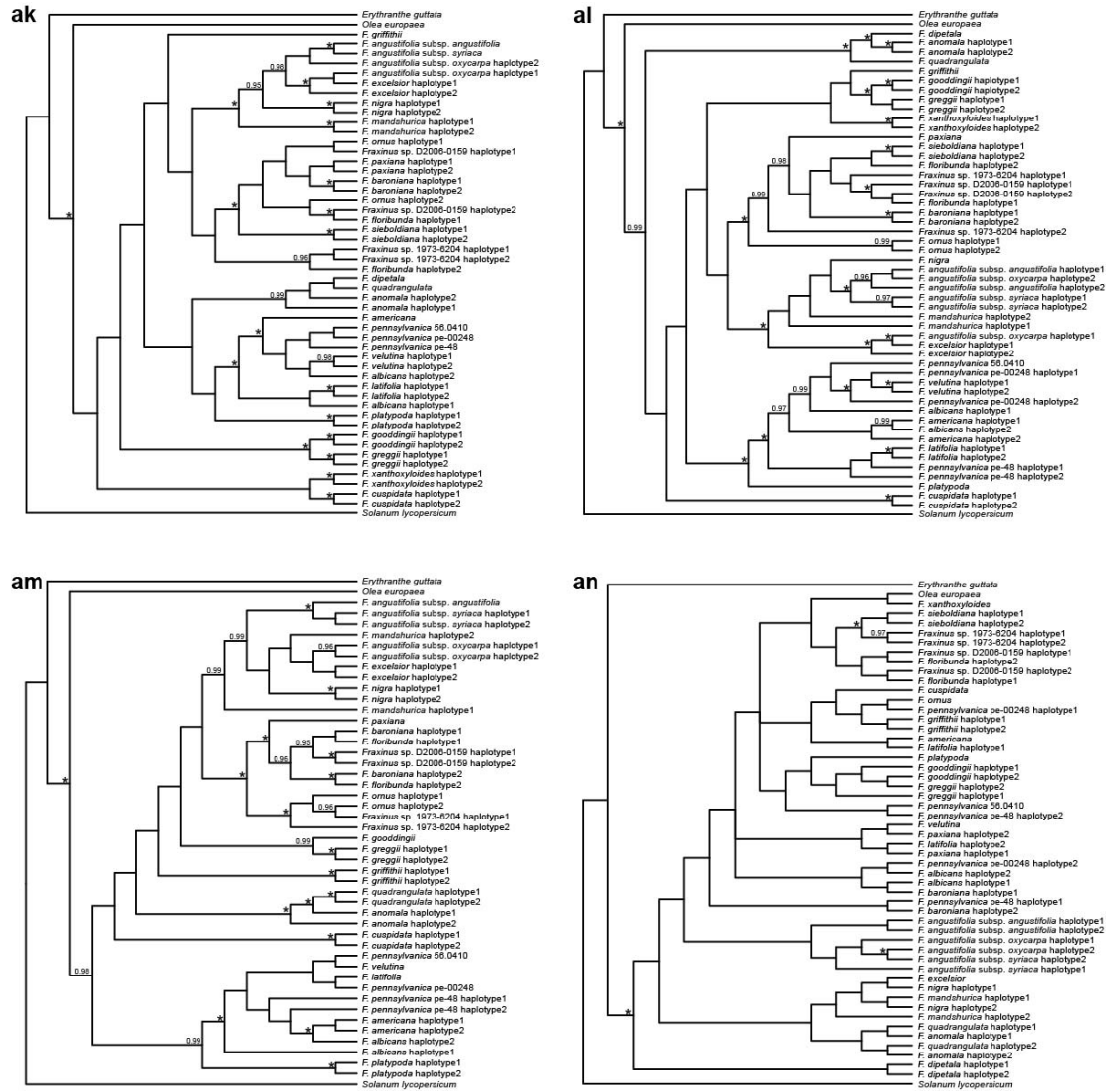


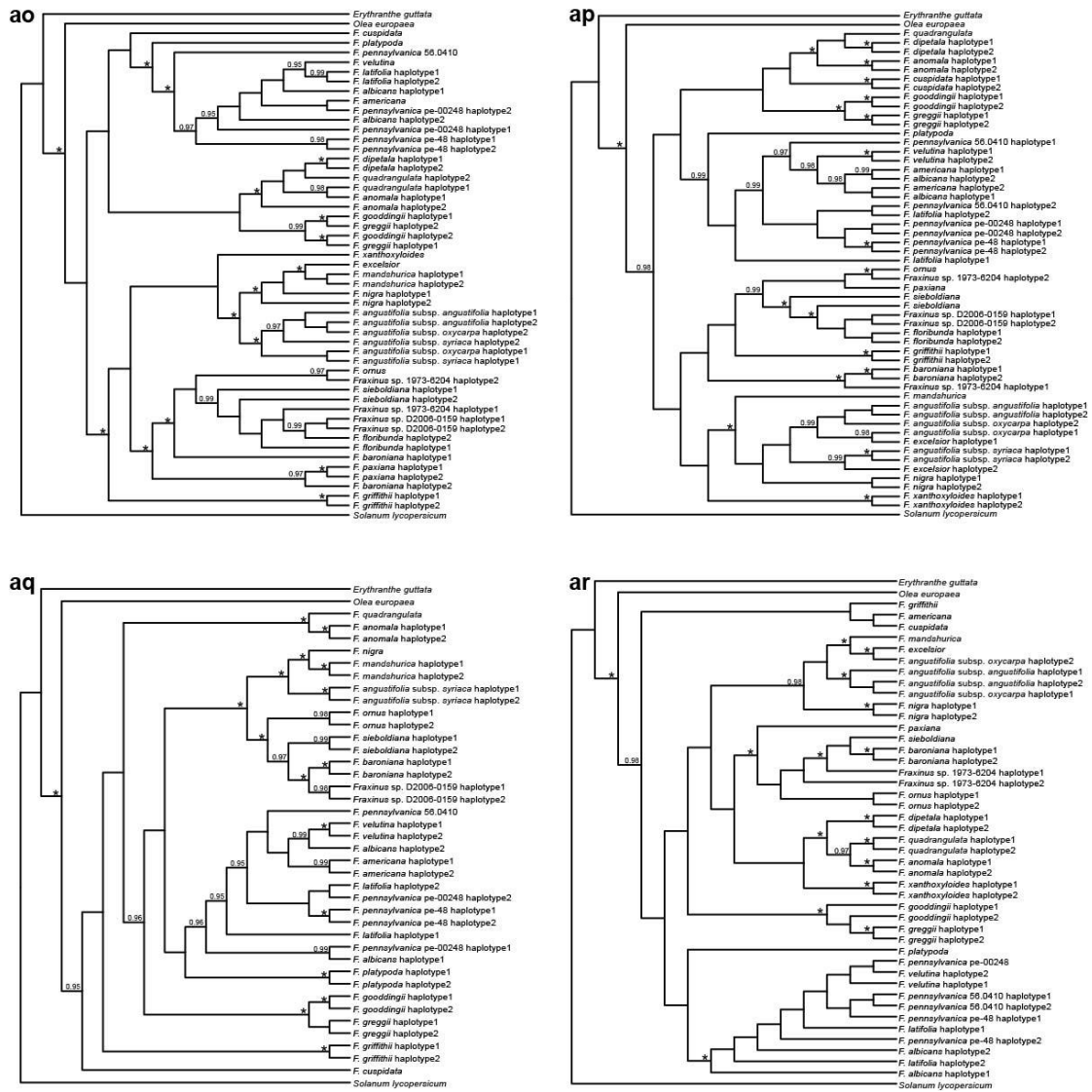


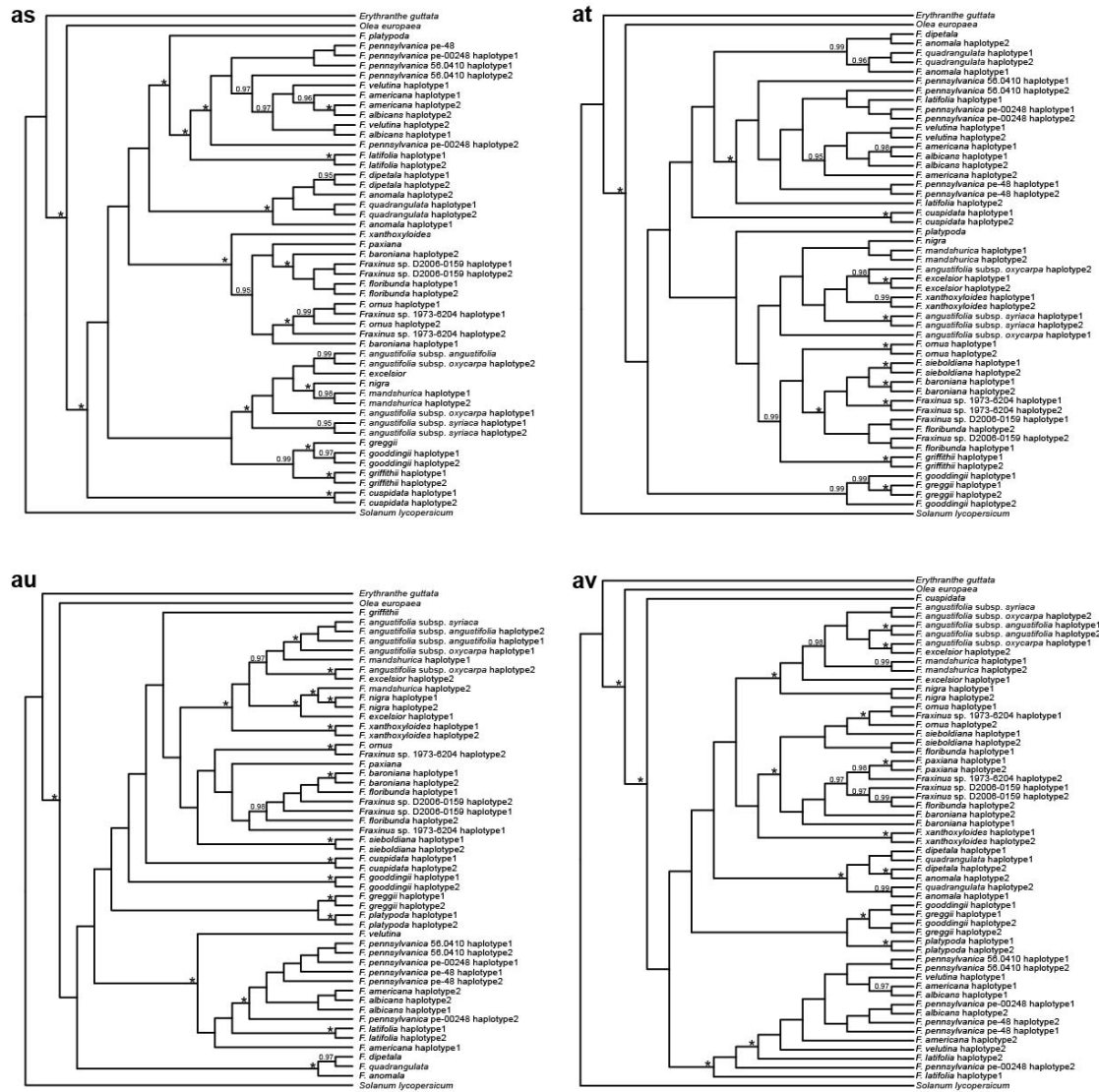


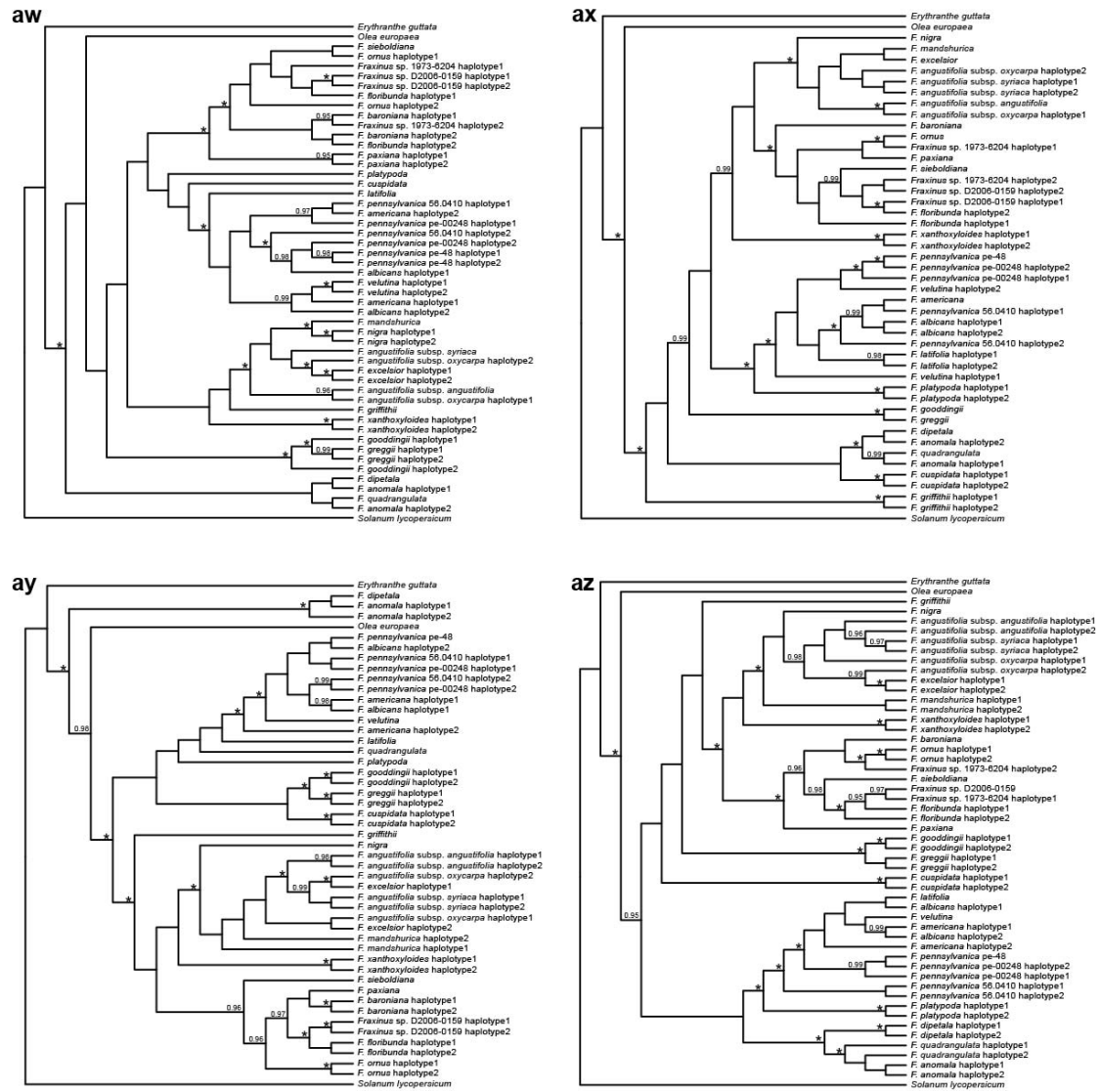


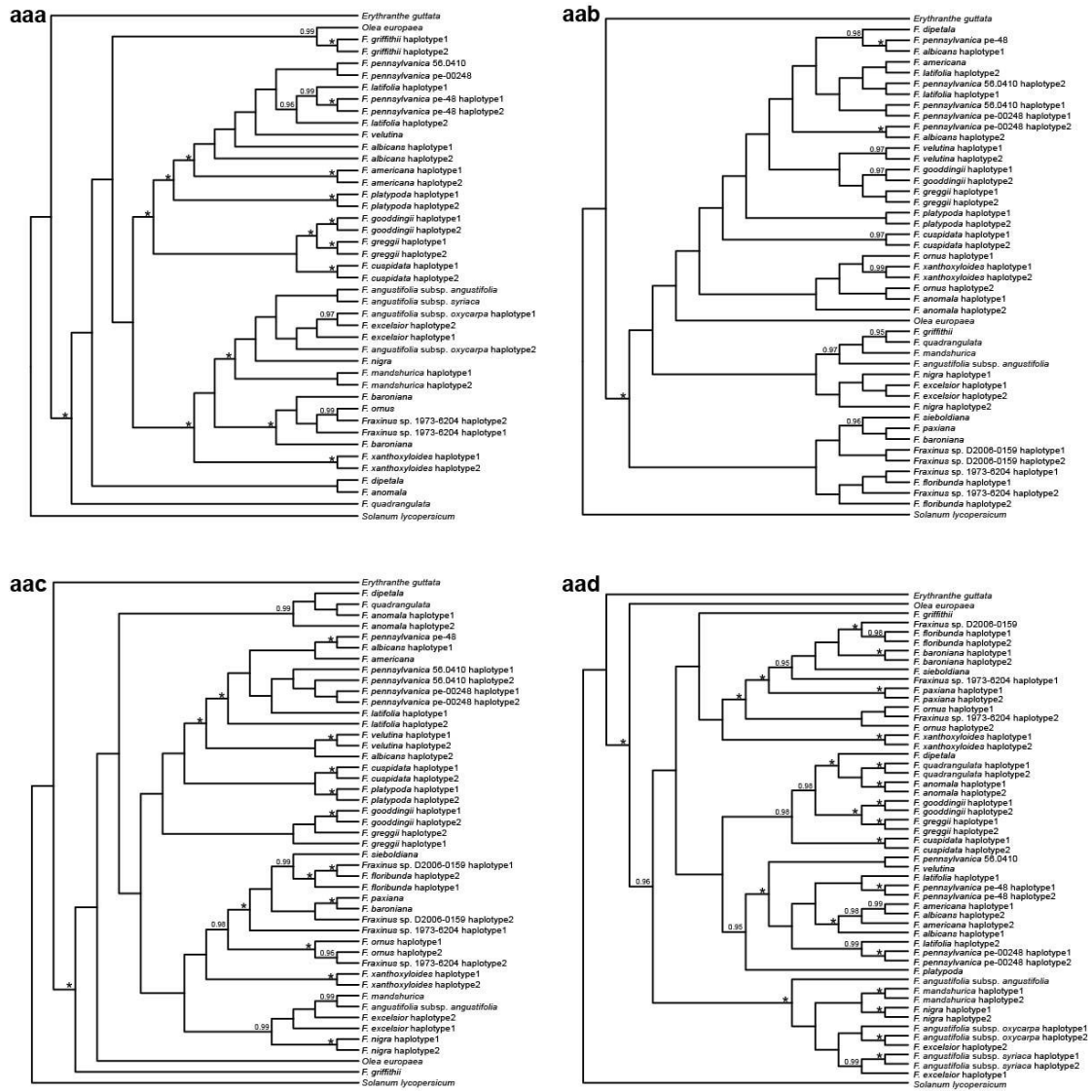




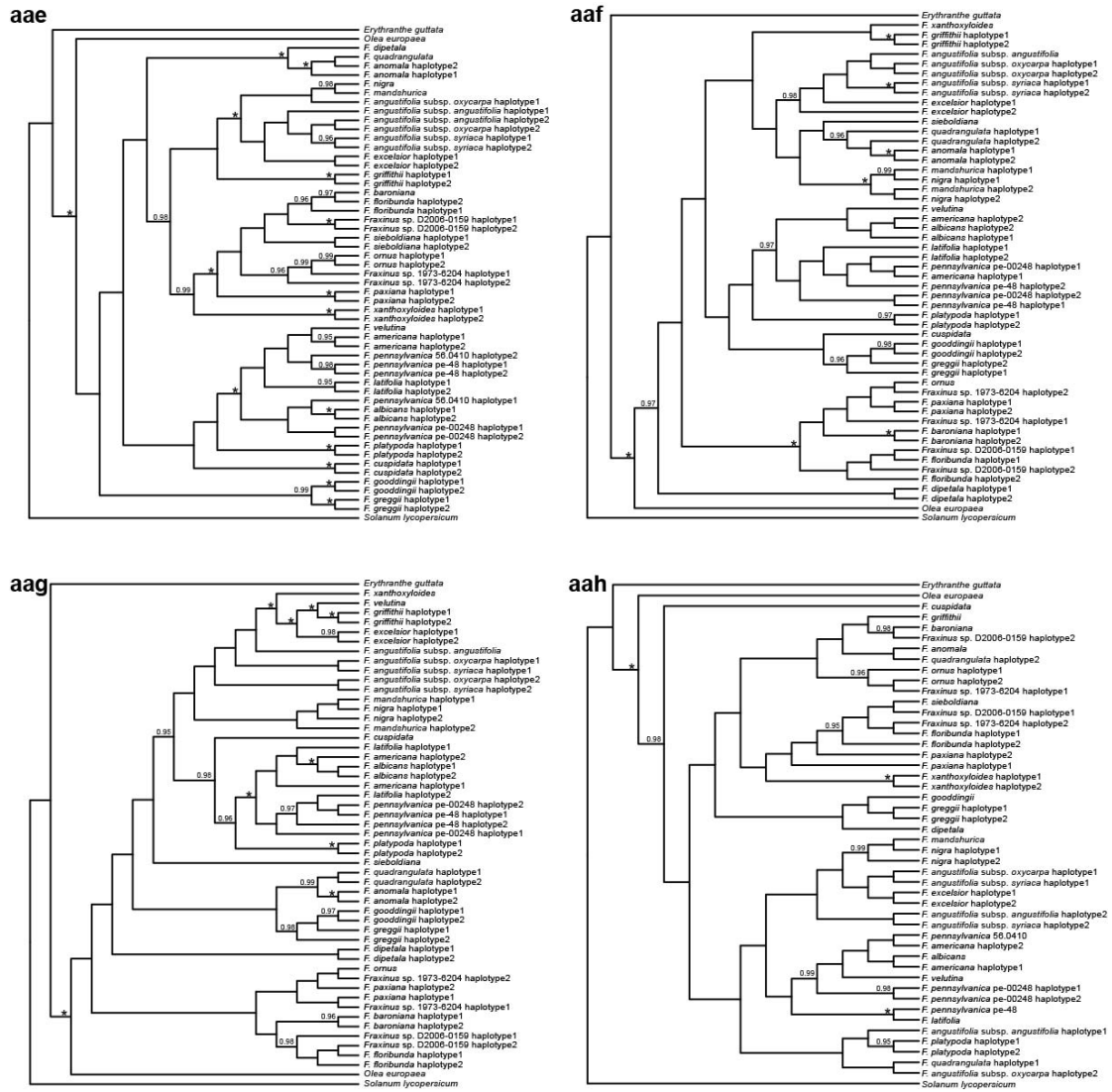


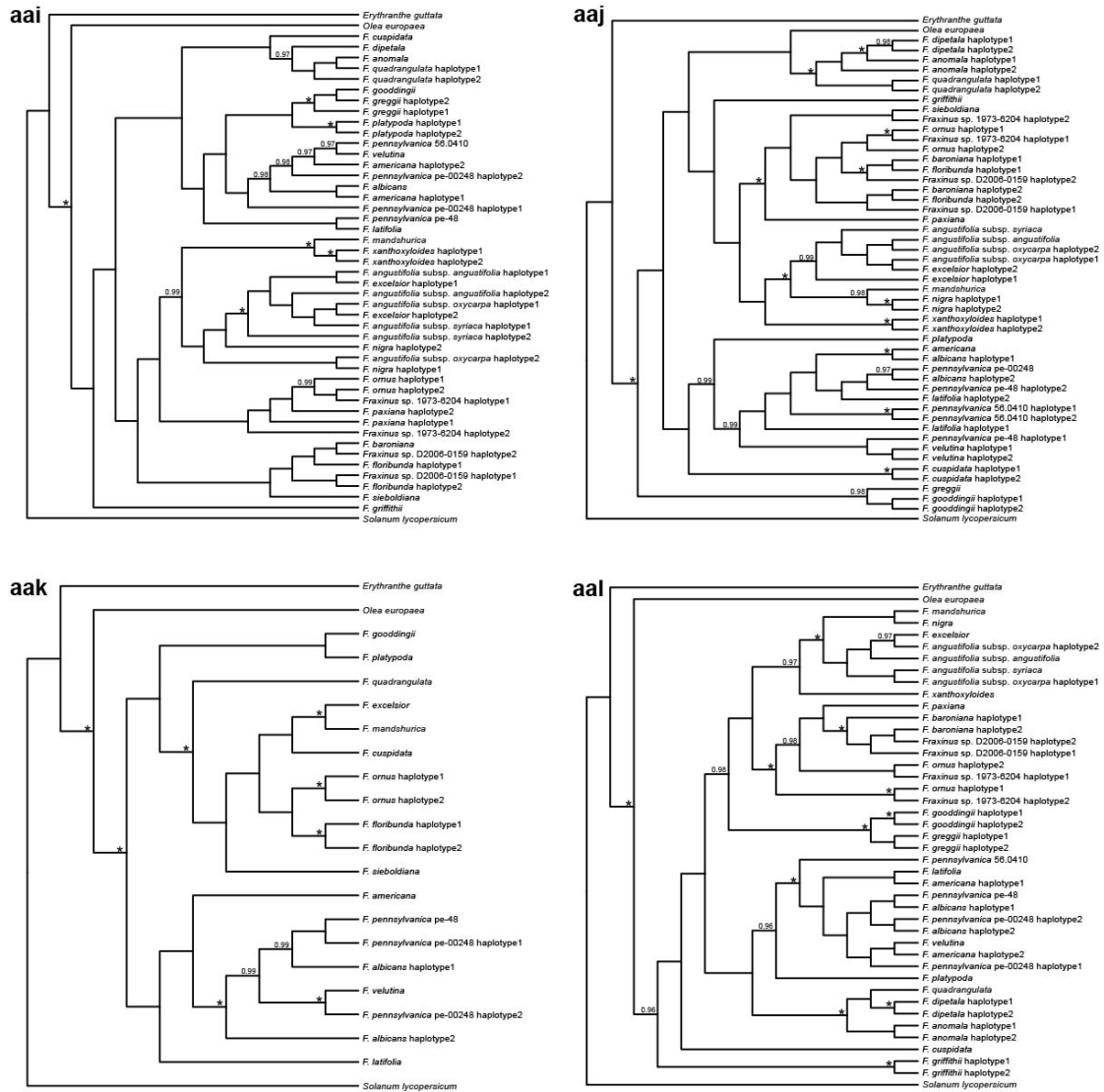


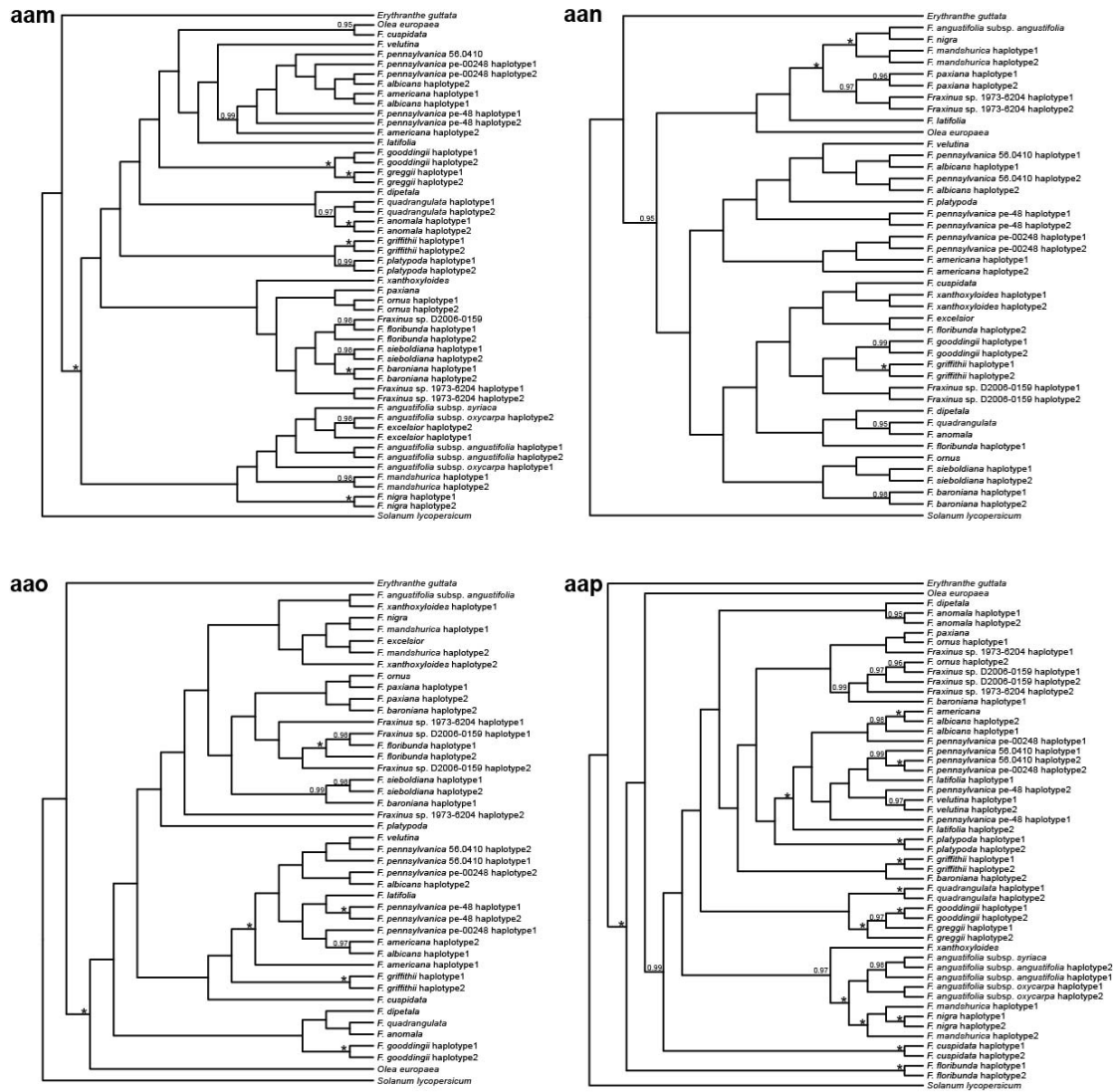


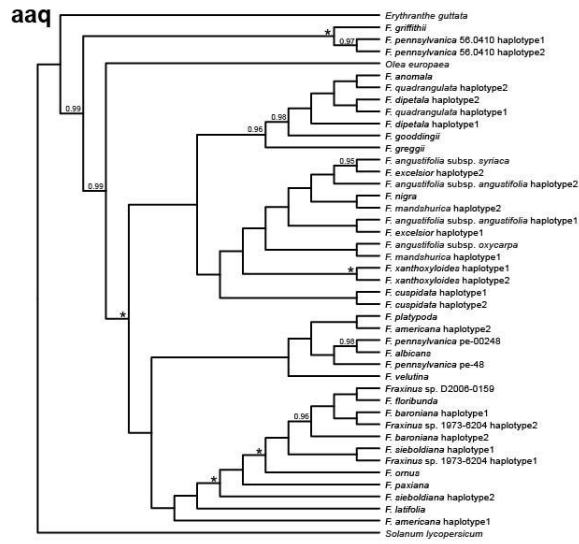




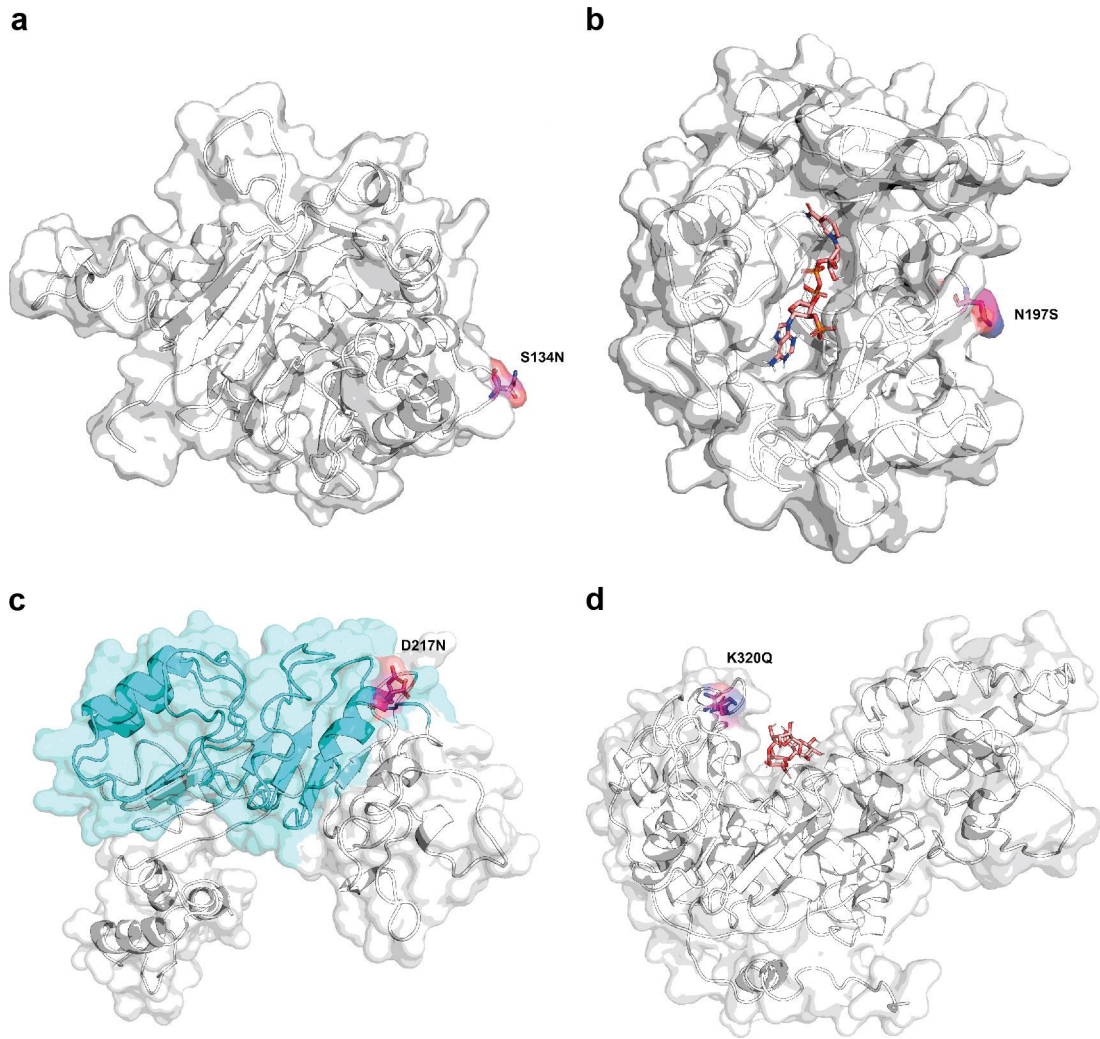








**Supplementary Figure 2. Predicted protein structures for selected candidate loci. a,** Predicted protein structure for OG36502, modelled using the protein sequence for *Fraxinus platypoda*. The serine/asparagine variant at the site where convergence was detected is highlighted; the serine is a putative phosphorylation site. **b,** Predicted protein structure for OG40061, modelled using the protein sequence for *F. mandshurica*. The asparagine/serine variant at the site where convergence was detected is highlighted; the serine is a putative phosphorylation site. The putative substrate, NADP, is shown docked within the predicted active site. **c,** Predicted protein structure for OG38407, modelled using the protein sequence for *F. mandshurica*. The aspartic acid/asparagine variant at the site where convergence was detected is highlighted; the site falls within a leucine rich repeat region (LRR; shaded blue) which is predicted to span from position 111-237 within the protein sequence (detected using the GenomeNet MOTIF tool ([www.genome.jp/tools/motif/](http://www.genome.jp/tools/motif/)), searching against the NCBI-CDD and Pfam databases with default parameters; the LRR region was identified as positions 111-237 with an e-value of 1e-05). **d,** Predicted protein structure for OG21033, modelled using the protein sequence for *F. platypoda*. The lysine/glutamine at the site where convergence was detected is highlighted. The putative substrate,  $\beta$ -D-Glcp-(1 $\rightarrow$ 3)- $\beta$ -D-GlcpA-(1 $\rightarrow$ 4)- $\beta$ -D-Glcp, is shown docked within the predicted active site.

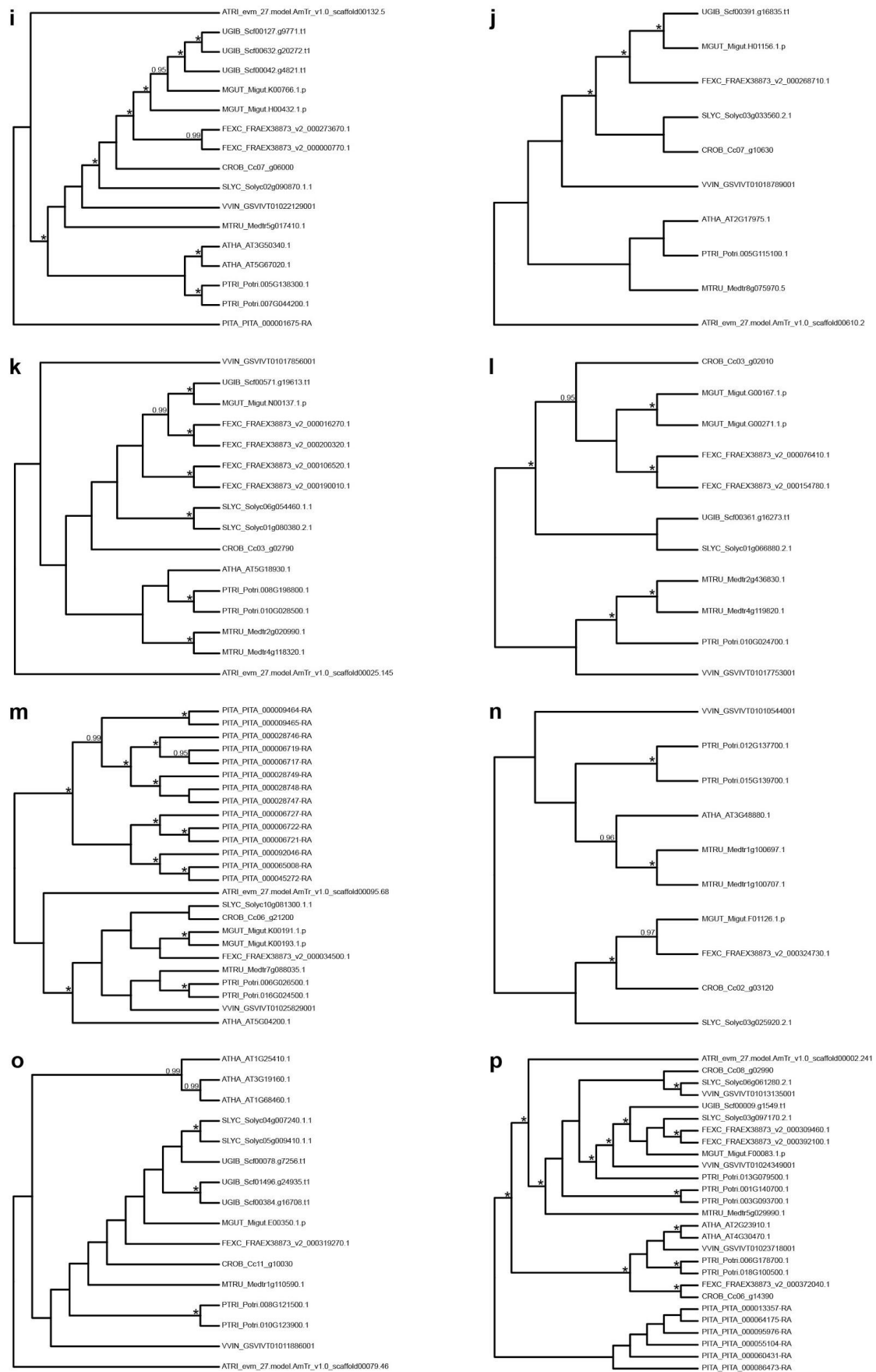


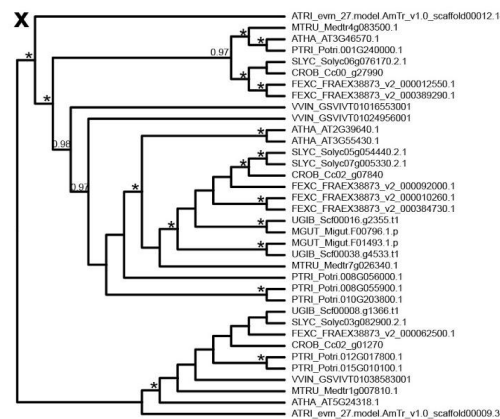
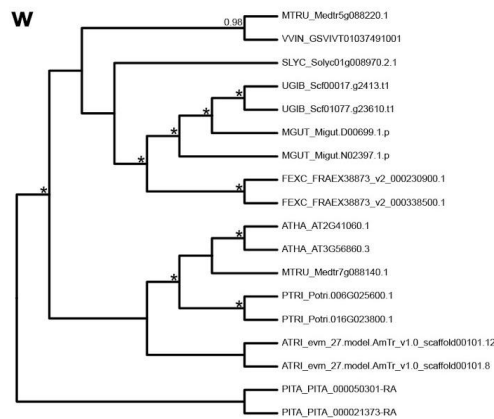
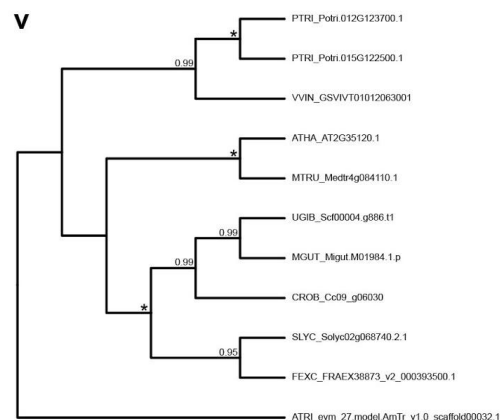
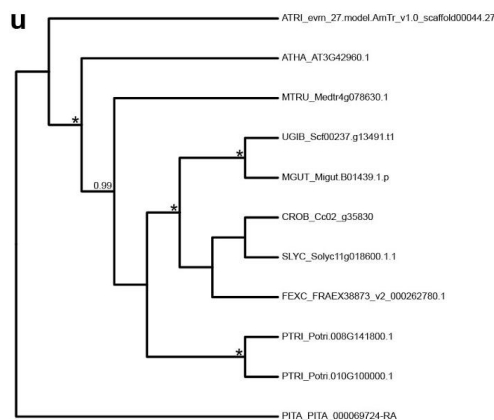
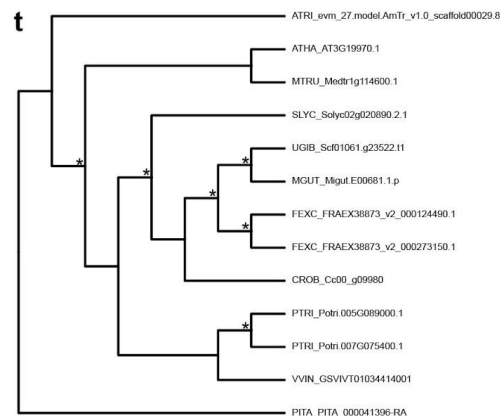
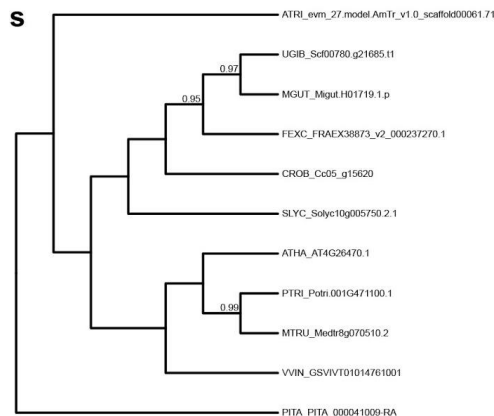
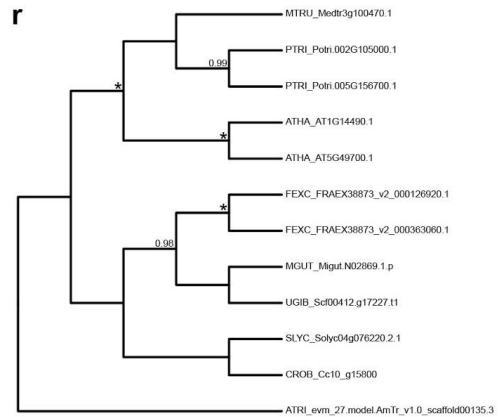
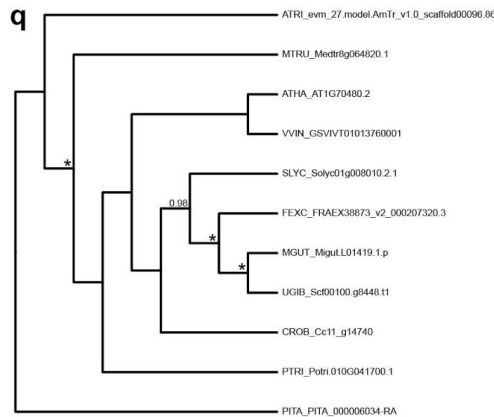
**Supplementary Figure 3. Gene-trees for OrthoMCL clusters including *Fraxinus***

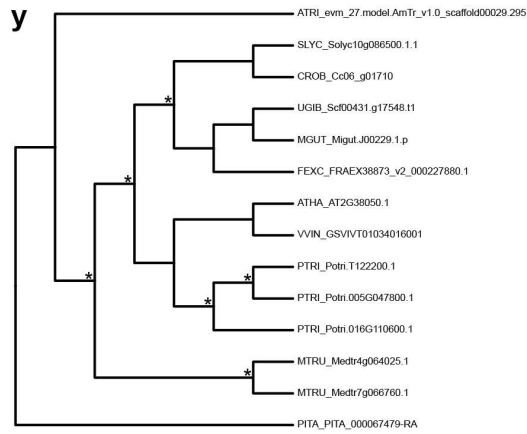
***excelsior* sequences from selected candidate genes.** Gene-trees for OrthoMCL clusters including FRAEX38873\_v2\_000261700 (a), FRAEX38873\_v2\_000200250 (b), FRAEX38873\_v2\_000386580 (c), FRAEX38873\_v2\_000122730 (d), FRAEX38873\_v2\_000111170 (e), FRAEX38873\_v2\_000248940 (f), FRAEX38873\_v2\_000080400 (g), FRAEX38873\_v2\_000093260 (h), FRAEX38873\_v2\_000000770 (i), FRAEX38873\_v2\_000268710 (j), FRAEX38873\_v2\_000106520 (k), FRAEX38873\_v2\_000076410 (l), FRAEX38873\_v2\_000034500 (m), FRAEX38873\_v2\_000324730 (n), FRAEX38873\_v2\_000319270 (o), FRAEX38873\_v2\_000309460 (p), FRAEX38873\_v2\_000207320 (q), FRAEX38873\_v2\_000363060 (r), FRAEX38873\_v2\_000237270 (s), FRAEX38873\_v2\_000382300 (t), FRAEX38873\_v2\_000262780 (u), FRAEX38873\_v2\_000393500 (v), FRAEX38873\_v2\_000338500 (w), FRAEX38873\_v2\_000092000 (x) and FRAEX38873\_v2\_000227880 (y). Gene-trees were inferred with MrBayes and rooted on either *Pinus taeda*, *Amborella trichopoda* or midpoint rooted (in cases where branch lengths viewed when the tree was unrooted did not clearly indicate that sequences from either *P. taeda* or *A. trichopoda* represented suitable outgroups). Numbers above branches are posterior probabilities (PP) of  $\geq 0.95$ ; asterisks indicate nodes with PP=1. Species names are abbreviated as follows: ATHA, *Arabidopsis thaliana*; ATRI, *Amborella trichopoda*; CROB, *Coffea canephora*; FEXC, *Fraxinus excelsior*; MGUT, *Erythranthe guttata*; MTRU, *Medicago truncatula*; PITA, *Pinus taeda*; PTRI, *Populus trichocarpa*; SLYC, *Solanum lycopersicum*; UGIB, *Utricularia gibba*; VVIN, *Vitis vinifera*. Gene model names follow species abbreviations.











## Supplementary Note 1

### 1. *Species-tree for Fraxinus*

The primary concordance tree inferred with BUCKy (Fig. 2) suggests evidence for significant incongruence between individual gene-tree topologies, as demonstrated by the low concordance factors (CFs) for many of the nodes. Nevertheless, with the exception of sect. *Ornus*, all taxonomic sections represented by more than one taxon are resolved as monophyletic with a CF of  $\geq 0.95$ , indicating that the majority of genes support these relationships. All species from sect. *Ornus* form a monophyletic group (CF=0.91) except for *F. griffithii*, which was also found to fall outside of the section on the basis of data from the plastid genome<sup>68</sup>.

## Supplementary Note 2

### 2. *Results of grand-conv analyses*

From grand-conv analyses of susceptible versus resistant *Fraxinus* lineages, we identified 64 loci with significant evidence of excess convergence (defined as loci with at least one amino acid site with a PP  $\geq 0.90$  of convergence and where the resistant branch pair has higher excess convergence than any other independent branch pair within the species-tree) in at least one of the three pairwise combinations of resistant lineages. The majority of loci (45) only show evidence of convergence between *F. mandshurica* and *F. platypoda*. Seven loci are only convergent between *F. mandshurica* and the sect. *Ornus* taxa (*F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159), and nine between *F. platypoda* and the sect. *Ornus* taxa. Of the remaining three loci, one shows evidence for convergence in two of the three pairwise comparisons, with the final two loci having evidence of convergence in all three pairs of lineages. For all three loci with evidence of convergence between more than one pair of lineages, the same residues were inferred as being convergent in each case. Filtering of this initial list of 64 loci according to the criteria specified in the Methods (see section “Refining the initial list of candidate loci identified with grand-conv”), resulted in a refined list of 53 loci. Further details of each of these loci is provided in Supplementary Note 4.

## Supplementary Note 3

### 3. *Evidence for loss-of-function mutations*

Nineteen of our 53 loci have evidence for potential loss-of-function mutations (start codon losses, stop codon gains/premature stop codons or frameshift variants) that cannot

be discounted by errors in gene model estimation (Supplementary Table 6). For ten of these loci, possible loss-of-function mutations are found only in susceptible taxa, for seven loci they are found only in resistant taxa, and for two loci they are found both in susceptible and resistant taxa. Three susceptible taxa (*F. latifolia*, *F. ornus* and *F. pennsylvanica* (susceptible genotype)) lack an allele without potential loss-of-function mutations (either they are homozygous for the mutation, or phasing indicates there are separate mutations on each allele) for two loci each (Supplementary Table 6) suggesting these genes could be non-functional, or generate a truncated protein product. For a single resistant taxon (*F. floribunda*) both alleles for OG60899 have potential loss-of-function mutations. In all other cases, there is either one allele that appears fully functional or it is uncertain whether multiple possible loss-of-function mutations are on the same or different alleles (Supplementary Table 6).

#### **Supplementary Note 4**

##### ***4. Details of candidate loci***

Seven candidate genes have putative roles relating to the phenylpropanoid biosynthesis pathway (OG853, OG15551, OG16673, OG19104, OG27080, OG40061 and OG64545), 15 have possible roles in perception and signalling that are relevant to defence response against insect herbivores (OG4469, OG11720, OG21033, OG23214, OG23284, OG24614, OG32176, OG33348, OG38407, OG39275, OG41448, OG41488, OG43828, OG47560 and OG50989) and two (OG16739 and OG37870) are candidates with putative roles related to hypersensitive like-response and programmed cell death. Further candidates also have potential roles relating to other aspects of defence response, including OG27693 and OG59564, which may play a role in response to oxidative stress, OG11013, an apparent member of a protein superfamily whose members have roles in plant stress and immunity, OG37560, a probable member of the *HIPP* gene family that may be involved in plant-pathogen interactions, and OG47629, which has a putative orthologue in *S. lycopersicum* that has been classified as a "disease-specific gene".

Below we provide details of the individual candidate genes, including information from literature searches and results obtained from any phylogenetic analyses and protein modelling. Gene model numbers from the BATG0.5 reference genome for *F. excelsior* are shown in parentheses after the OMA group names.

#### 4.1. Candidates found in all three grand-conv pairwise comparisons

**OG6935 (FRAEX38873\_v2\_000218910).** Grand-conv detected evidence of a convergent methionine in all five resistant taxa, at a site where the other *Fraxinus* species in the analysis have a threonine (Supplementary Table 5). The top match for this gene in *Arabidopsis thaliana* is an RNA helicase family protein (AT3G62310) which is in the DExD/H-box RH family and has a putative orthologue in rice (OS03G0314100; both belong to the same OMA group (fingerprint: QWCVDFA), [omabrowser.org/oma/info/ARATH24986/](http://omabrowser.org/oma/info/ARATH24986/)) that is drought inducible<sup>102</sup>. Expression of AT3G62310 has been shown to be induced in response to salt acclimation<sup>103</sup>.

**OG11013 (FRAEX38873\_v2\_000029110).** Grand-conv detected evidence of a convergent arginine in all five resistant taxa, at a site where the other *Fraxinus* species in the analysis have a glutamine (Supplementary Table 5). The alignment of all sequences for OG11013 showed some susceptible species also have an arginine at the relevant amino acid site. However, the glutamine was only present in species that are susceptible to EAB, including *F. ornus* (sect. *Ornus*) and members of sect. *Melioides* (Supplementary Table 8), indicating possible convergence between a subset of the taxa with high susceptibility. It should be noted that both *F. pennsylvanica* individuals (susceptible and putatively resistant) have a glutamine at the relevant position. Among the outgroups, *Mimulus* had an arginine, *Olea* a glutamine and *Solanum* a histidine. The top match for this gene in *A. thaliana* is a tetratricopeptide repeat (TPR)-like superfamily protein encoding gene (AT3G49142). Although the function of AT3G49142 has not been fully characterised, the functions of some individual TPR proteins have been deduced, and include roles in plant stress, hormone signaling and immunity<sup>104</sup>.

#### 4.2. Candidate found in grand-conv analyses 1 and 3

**OG27838 (FRAEX38873\_v2\_000134340).** Grand-conv detected evidence of a convergent arginine in the five resistant taxa, at a site where the susceptible *Fraxinus* species in the analysis have a glutamine. Although grand-conv identified the arginine as convergent between *F. mandshurica* and *F. platypoda*, and between *F. platypoda* and *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, it was not detected as convergent in the pairwise comparison between *F. mandshurica* and *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159 (Supplementary Table 5). The top match for this gene in *A. thaliana* is *REDUCED DORMANCY 2* (AT2G38560), or *RDO2*, which is also known as

*TRANSCRIPT ELONGATION FACTOR IIS (TFIIS)*. AT2G38560/*RDO2/TFIIS* encodes a transcription factor that is needed for RNA polymerase II (RNAPII) processivity, stimulating transcription elongation by RNAPII, and may also play a role in the control of alternative splicing<sup>105</sup>.

**4.3. Candidates found in *F. mandshurica* and *F. platypoda* (grand-conv analysis 1) OG853 (FRAEX38873\_v2\_000200250).** Grand-conv detected evidence of a convergent phenylalanine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an isoleucine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *REF4-RELATED 1* (AT3G23590), or *RFRI*; phylogenetic analysis of the OrthoMCL cluster containing both FRAEX38873\_v2\_000200250 and AT3G23590 indicates that they are orthologues (Supplementary Fig. 3b). AT3G23590/*RFRI* is also known as *MED5a* or *MED33a* and encodes a protein that belongs to the Mediator complex, which is a transcriptional regulator of nearly all cellular pathways<sup>20</sup>. AT3G23590/*RFRI/MED5a* acts in a partially redundant manner with its paralogue *REF4/MED5b* (AT2G48110) to limit phenylpropanoid accumulation<sup>20</sup>. Disruption of both the *MED5a* and *MED5b* genes in the lignin-deficient *ref8* mutant of *A. thaliana*, which has reduced activity of *REF8/CYP98A3* (AT2G40890; the gene encoding *p*-coumaroylshikimate 3'-hydroxylase (C3'H) and the best matching *A. thaliana* gene for one of our other candidates, OG15551/FRAEX38873\_v2\_000261700), restores lignin deposition to wild-type levels<sup>21</sup>. However, the disruption of the *MED5* genes does not restore the synthesis of guaiacyl and syringyl lignin subunits; the lignin content of the *med5a/5b ref8* triple mutant is almost entirely composed of *p*-hydroxyphenyl subunits, which account for < 2% of the lignin content in wild-type plants<sup>21</sup>. It has also been found that *med5a/5b* loss-of-function mutant plants show increased expression of numerous genes involved in the phenylpropanoid biosynthesis pathway, including those encoding PAL, C4H, 4CL, C3'H, CCR, CAD and CSE, and that 'phenylpropanoid biosynthesis' was the most significantly enriched GO term for genes that were overexpressed in *med5a/5b* plants<sup>21</sup>. Although the direct targets of the *MED5* genes are yet to be identified, it has been suggested that *MED5a* and *MED5b* may limit phenylpropanoid biosynthesis via the transcriptional activation of genes that negatively regulate the pathway<sup>106</sup>. Further analysis of *med5a/5b* loss-of-function mutant plants found that genes with significantly decreased expression were enriched for GO terms related to defence response, such as "response to jasmonic acid", "response to wounding" and "defence response to insect",

suggesting a role for *MED5a* and *MED5b* in regulation of defence response genes<sup>22</sup>. The orthologue of FRAEX38873\_v2\_000200250 in *S. lycopersicum* is Solyc01g080200 (Supplementary Fig. 3b), and has been shown to be upregulated in response to tomato yellow leaf curl virus<sup>107</sup>.

**OG2897 (FRAEX38873\_v2\_000138720).** Grand-conv detected evidence of a convergent isoleucine and glutamic acid in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have a phenylalanine and glycine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a tetratricopeptide repeat (TPR)-like superfamily protein (AT4G30825), also known as *PPRb*<sup>108</sup>. AT4G30825/*PPRb* is involved in nucleotide-excision repair, phosphorylation and regulation of transcription by RNA polymerase II, amongst other processes ([www.arabidopsis.org/servlets/TairObject?id=500231848&type=locus](http://www.arabidopsis.org/servlets/TairObject?id=500231848&type=locus)). AT4G30825 is co-expressed with the disease resistance regulator *OVEREXPRESSION OF CATIONIC PEROXIDASE3 (OCP3)*<sup>108</sup>, which also plays a role in drought tolerance<sup>109</sup>.

**OG4372 (FRAEX38873\_v2\_000149840).** Grand-conv detected evidence of a convergent isoleucine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a serine or aspartic acid (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a transcription regulator NOT2/NOT3/NOT5 family protein (AT5G18230). NOT2 and NOT3/5 form a module that is part of the CCR4-NOT complex, which is a conserved multi-subunit complex that regulates gene expression at different levels<sup>110</sup>. The NOT2–3/5 module appears to be involved in coordination of transcriptional regulation and mRNA degradation, as well as assembly of cellular complexes such as the proteasome<sup>110</sup>.

**OG4469 (FRAEX38873\_v2\_000386580).** Grand-conv detected evidence of a convergent threonine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a proline (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a protein kinase superfamily protein (AT5G24080); phylogenetic analysis of the OrthoMCL cluster containing FRAEX38873\_v2\_000386580 and AT5G24080 indicates they are likely to be orthologous (Supplementary Fig. 3c). The protein encoded by AT5G24080 is reported to be a G-type lectin S-receptor-like serine/threonine-protein kinase with an ATP binding



site, belonging to the Ser/Thr protein kinase family ([www.uniprot.org/uniprot/Q9FLV4](http://www.uniprot.org/uniprot/Q9FLV4)). G-type lectin receptor kinases have various functions, including in self-incompatibility, with some having a role in plant defence<sup>111,112</sup> such as in response to the tobacco hawk moth (*Manduca sexta*) in *Nicotiana attenuata*<sup>113</sup> or to the brown planthopper (*Nilaparvata lugens*) in rice<sup>114</sup>. Extracellular ATP is associated with cell damage, and perception of ATP may play a role in defence response against insect herbivores via stimulation of the JA pathway<sup>27</sup>. AT5G24080 has also been named *AtG-LecRK-I.6* by Teixeira et al.<sup>111</sup>, who suggest it may be defective because its protein has an incomplete kinase domain and lacks some of the subdomains typically found in kinases. However, such proteins may still be functional even if they lack kinase activity<sup>111</sup>. Putative orthologues of OG4469/FRAEX38873\_v2\_000386580 in *S. lycopersicum* (Solyc03g007790 and Solyc06g036470) have been found to encode proteins that include the S-locus glycoprotein domain, transmembrane domain and kinase domains<sup>111</sup>, although their precise functions have not been characterised.

**OG5539 (FRAEX38873\_v2\_000290300)**. Grand-conv detected evidence of a convergent leucine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a phenylalanine (Supplementary Table 5). However, the leucine is also found in the alternate allele for susceptible *F. americana* (Supplementary Table 8). The top match for this gene in *A. thaliana* is *CHLORIDE CHANNEL G* (AT5G33280), *CLCG*. AT5G33280/*CLCG* is also known as *CBSCLC6* ([www.uniprot.org/uniprot/P60300](http://www.uniprot.org/uniprot/P60300)) and is suggested to function in chloride homeostasis during NaCl stress<sup>115</sup>.

**OG7454 (FRAEX38873\_v2\_000218430)**. Grand-conv detected evidence of a convergent histidine and lysine in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have arginine and glutamic acid (Supplementary Table 5). The top match for this gene in *A. thaliana* is *EMBRYO DEFECTIVE 3141* (AT5G50390), or *EMB3141*. AT5G50390/*EMB3141* is also known as *PCMP-H58* and encodes a pentatricopeptide repeat containing protein ([www.uniprot.org/uniprot/Q9FK33](http://www.uniprot.org/uniprot/Q9FK33)); this gene appears to be required for embryo development<sup>116</sup>.

**OG10762 (FRAEX38873\_v2\_000122730).** Grand-conv detected evidence of a convergent serine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an alanine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *WITH NO LYSINE (K) KINASE 4* (AT5G58350), or *WNK4*. Phylogenetic analysis of the OrthoMCL cluster containing both FRAEX38873\_v2\_000122730 and AT5G58350 indicates that they are orthologues (Supplementary Fig. 3d). AT5G58350/*WNK4* is also known as *ZIK2* and encodes a serine/threonine kinase<sup>117</sup>. The function of AT5G58350/*WNK4* is not yet known, although its expression is influenced by circadian rhythms and other members of the *WNK* gene family in *A. thaliana* are involved in the regulation of flowering time<sup>117</sup>. It has also been suggested that plant *WNK* genes are involved in enhancing the oxidative defence system<sup>117</sup>. GO terms have been associated with AT5G58350/*WNK4* via identification of functional gene modules<sup>118</sup>, namely: CELLULAR RESPONSE TO PHOSPHATE STARVATION, GALACTOLIPID BIOSYNTHETIC PROCESS, CELLULAR RESPONSE TO WATER DEPRIVATION and ORGAN SENESCENCE (see [http://bioinformatics.psb.ugent.be/cig\\_data/plant\\_modules/createResultsHTML2.php?geneID=AT5G58350&moduleTypes=repModules](http://bioinformatics.psb.ugent.be/cig_data/plant_modules/createResultsHTML2.php?geneID=AT5G58350&moduleTypes=repModules)). The orthologue of FRAEX38873\_v2\_000122730 in *S. lycopersicum* appears to be Solyc06g082470 (Supplementary Fig. 3d) and is also known as SIMAPKKK42<sup>119</sup>. The function of Solyc06g082470/SIMAPKKK42 has not been fully characterised, but it has been suggested that SIMAPKKK genes in general might have a role in plant hormone signalling in relation to development and defence response<sup>119</sup>.

**OG11720 (FRAEX38873\_v2\_000155730).** Grand-conv detected evidence of a convergent valine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an isoleucine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *NITRATE TRANSPORTER 1.5* (AT1G32450), which is also known as *NRT1.5* or *NPF7.3*. AT1G32450/*NRT1.5* is a transmembrane nitrate transporter and has been shown to be involved in transport of nitrate from the root to shoot<sup>32</sup>. In addition to nitrate, some members of the NRT1 family (also known as the PTR or NPF family) are also involved in transport of some phytohormones and defence compounds<sup>120</sup>. Indeed, many of the 53 members of the NRT1 family in *A. thaliana* have multiple

substrates and may be able transport jamonate, abscisic and/or gibberellins as well as nitrates, or other substrates<sup>121</sup>.

**OG13887 (FRAEX38873\_v2\_000015960).** Grand-conv detected evidence of a convergent lysine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an arginine (Supplementary Table 5). The top match for this gene in *A. thaliana* is 54 CHLOROPLAST PROTEIN (AT5G03940), or 54CP, and is also known as cpSRP54. AT5G03940/54CP/cpSRP54 encodes the 54kDa subunit of the chloroplast signal recognition particle and an *A. thaliana* mutant (*cbd*) with a T-DNA insertion in the gene has lower chlorophyll and carotenoid content compared with wild type plants, shows defects in chloroplast development and is deficient in abscisic acid<sup>122</sup>. The *cbd* mutant also shows evidence of cell death in leaves; it has been suggested that a reduced ability to quench reactive oxygen species as a result the lower levels of carotenoids in the mutant may contribute to the observed cell death<sup>122</sup>.

**OG15551 (FRAEX38873\_v2\_000261700).** Grand-conv detected evidence of four convergent amino acids in *F. mandshurica* and *F. platypoda*, valine, isoleucine, isoleucine and methionine, where the other *Fraxinus* species in the analysis have an isoleucine at the first site and leucine at the other three (Supplementary Table 5). The top match for this gene in *A. thaliana* is the CYP98A3 gene (AT2G40890; also known as REF8). In *A. thaliana* this gene is at a critical bottleneck in the phenylpropanoid pathway<sup>18</sup>. When CYP98A3/REF8 is mutated in *A. thaliana*, it results in reduced lignin content, changes in lignin composition, a lack of soluble sinapoyl esters, and accumulation of flavonol glycosides<sup>18</sup>. Another study found a more than 90% reduction in the level of scopoletin and scopolin in the roots of *A. thaliana* mutants with T-DNA insertions in this gene compared with that found in wild type plants<sup>123</sup>. Phylogenetic analysis of the OrthoMCL cluster containing AT2G40890 and FRAEX38873\_v2\_000261700 indicates that they are close paralogues, rather than orthologues and suggests that the *Fraxinus* gene lacks a direct orthologue in *A. thaliana* (Supplementary Fig. 3a). There appears to be a small family of CYP98A3-like genes in *F. excelsior*, including FRAEX38873\_v2\_000095940 and FRAEX38873\_v2\_000127720 as well as FRAEX38873\_v2\_000261700 (Supplementary Fig. 3a). In a predicted protein model for OG15551/FRAEX38873\_v2\_000261700 all four variable sites identified by grand-conv are in close proximity to the active site (Fig. 3).

**OG16673 (FRAEX38873\_v2\_000150510).** Grand-conv detected evidence of a convergent isoleucine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a valine (Supplementary Table 5). However, the isoleucine is also found in the alternate allele for susceptible *F. ornus* (Supplementary Table 8). The top match for this gene in *A. thaliana* is *B-GLUCOSIDASE 41* (AT5G54570; also known as *BGLU41*). *BGLU41* belongs to the glycoside hydrolase family 1 (GH1) in *A. thaliana*, members of which function in processes such as chemical defence against herbivory, lignification, cell wall modification and control of phytohormone levels<sup>23,124</sup>. A number of *cis*-elements have been found in the region upstream of AT5G54570/*BGLU41*, including those that are responsive to abscisic acid and gibberellins<sup>124</sup>. The putative orthologue of OG16673/FRAEX38873\_v2\_000150510 in tomato is Solyc07g063880; this gene is differentially expressed in the *ovate* mutant and it had been suggested that its downregulation could be responsible for decreased glucose content in the fruit of *ovate* plants<sup>125</sup>. There has been debate regarding whether glucose should be considered as part of defence response, because it can reduce herbivore performance<sup>126</sup>.

**OG17252 (FRAEX38873\_v2\_000111170).** Grand-conv detected evidence of a convergent asparagine and methionine in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have aspartic acid and threonine (Supplementary Table 5). The top match for this gene in *A. thaliana* is AT1G50460 (*HEXOKINASE-LIKE 1/HKL1*), which has a role in abiotic stress response in addition to its other functions (growth, root hair development<sup>127</sup>). AT4G29130 (*HXK1*) belongs to the same OrthoMCL cluster as AT1G50460, along with three other *A. thaliana* genes, and has response to oxidative stress and pathogen resistance among its functions<sup>127</sup>. Phylogenetic analysis of the OrthoMCL cluster indicates that FRAEX38873\_v2\_000111170 is orthologous to both AT1G50460/*HKL1* and another *A. thaliana* gene, AT3G20040 (*HEXOKINASE-LIKE 2/HKL2*; Supplementary Fig. 3e). AT3G20040/*HKL2* is also known as *ATHXK4*, but its function is apparently unknown<sup>127</sup>.

**OG19104 (FRAEX38873\_v2\_000248940).** Grand-conv detected evidence of a convergent leucine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a serine (Supplementary Table 5). The top match

for this gene in *A. thaliana* is AT4G01240, a gene encoding a S-adenosyl-L-methionine-dependent methyltransferases superfamily protein. Phylogenetic analysis of the OrthoMCL cluster containing FRAEX38873\_v2\_000248940 and AT4G01240 indicates they are likely to be orthologous (Supplementary Fig. 3f). The function of AT4G01240 has not been characterised but the gene was found to be within a region related to aphid feeding performance from a GWAS conducted in *A. thaliana*<sup>128</sup>. Plant S-adenosyl-L-methionine-dependent methyltransferases are also key enzymes in the phenylpropanoid, flavonoid and other metabolic pathways<sup>129</sup>. Methyltransferases in maize may function in generating volatile methyl esters (which function in plant defence) in response to herbivore (African cotton leaf worm) damage<sup>130</sup>. The serine variant in *Fraxinus* is a putative phosphorylation site (Supplementary Table 5); it was not possible to generate a predicted model for OG19104/FRAEX38873\_v2\_000248940 and therefore the position of the putative phosphorylation site within the structure of the protein is unknown.

**OG20252 (FRAEX38873\_v2\_000397500).** Grand-conv detected evidence of a convergent valine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an isoleucine (Supplementary Table 5). However, the gene-tree inferred for OG20252 with the codon for the valine/isoleucine variant excluded groups sect. *Fraxinus* (all of which have the valine; Supplementary Table 8) and *F. platypoda* together (PP=0.98; Supplementary Fig. 1a) in conflict with the species-tree (Fig. 2), indicating that the evidence of convergence found by grand-conv is likely due to introgression or incomplete lineage sorting. The top match for this gene in *A. thaliana* is *BROMODOMAIN 4* (AT1G61215) which encodes a bromodomain protein with a DNA binding motif. The exact function of the bromodomain is unclear, but it may be involved in protein-protein interactions (<http://www.ebi.ac.uk/interpro/entry/IPR001487>). The putative orthologue of OG20252/FRAEX38873\_v2\_000397500 in *S. lycopersicum* is Solyc02g078810.2 and is annotated with a MYB-like DNA binding domain in PhytoMine (<https://phytozome.jgi.doe.gov/phytoMine/report.do?id=285298920&trail=%7c285298920>).

**OG20859 (FRAEX38873\_v2\_000138760).** Grand-conv detected evidence of a convergent alanine and glutamic acid in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have threonine or serine and aspartic acid

(Supplementary Table 5). The top match for this gene in *A. thaliana* is AT5G57830, which encodes a zein-binding protein of unknown function.

**OG23214 (FRAEX38873\_v2\_000080400).** Grand-conv detected evidence of a convergent valine and isoleucine in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have a methionine and threonine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding an eukaryotic aspartyl protease family protein (AT1G64830). Aspartic proteases are involved in many biological processes, including senescence, stress responses, and programmed cell death<sup>131</sup>. Phylogenetic analysis of the OrthoMCL cluster that contains FRAEX38873\_v2\_000080400 indicates it is orthologous to both AT1G64830 and another *A. thaliana* gene, AT5G33340 (*CDRI/CONSTITUTIVE DISEASE RESISTANCE I*; see Supplementary Fig. 3g). AT5G33340/*CDRI* is an extracellular aspartic protease that is involved in disease signalling via salicylic acid-dependent inducible resistance<sup>132</sup>; overexpression of AT5G33340/*CDRI* leads increased expression of other defence-related genes and enhanced resistance to a bacterial pathogen<sup>132</sup>. Salicylic acid is more often linked with defense against biotrophic pathogens, but is involved in response to oviposition by insect herbivores<sup>17</sup> and plays a key role in pathogen-triggered programmed cell death<sup>33</sup>.

**OG24614 (FRAEX38873\_v2\_000093260).** Grand-conv detected evidence of a convergent isoleucine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a threonine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a zinc finger protein (AT5G54630). However, phylogenetic analysis of the relevant OrthoMCL cluster indicates another *A. thaliana* gene, AT4G27240, is equally closely related to FRAEX38873\_v2\_000093260, with both appearing to be paralogues of the *F. excelsior* gene (Supplementary Fig. 3h). AT4G27240 also encodes a zinc finger protein; both AT4G27240 and AT5G54630 are transcription factors<sup>133</sup>, but their functions are not fully characterised. AT5G54630 is expressed in response to brassinolide<sup>134</sup>, a brassinosteroid, which have a crucial role in regulating the growth-immunity trade-off<sup>135</sup>; transcriptions factors act both up and downstream of phytohormone signalling to regulate defence responses<sup>27</sup>. One of the susceptible taxa included in our convergence analyses, *F. ornus*, may lack a functional copy of this gene (Supplementary Table 6).

**OG24969 (FRAEX38873\_v2\_000395930).** Grand-conv detected evidence of a convergent serine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a threonine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *ROHI* (AT1G63930). AT1G63930 has been suggested to play a role in seed coat development, but it is possible that it does not have this function in wild-type *A. thaliana*<sup>136</sup>. One of the susceptible taxa included in our convergence analyses, *F. pennsylvanica*, may lack a functional copy of this gene (Supplementary Table 6).

**OG26964 (FRAEX38873\_v2\_000042410).** Grand-conv detected evidence of a convergent histidine and methionine in *F. mandshurica*, at sites where the other *Fraxinus* species in the analysis have an arginine and threonine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *C5orf35* (AT5G23200); the function of this gene has not been characterised.

**OG27080 (FRAEX38873\_v2\_000000770).** Grand-conv detected evidence of a convergent aspartic acid and methionine in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have an asparagine and leucine (Supplementary Table 5). The top match for this gene in *A. thaliana* is the uncharacterised gene AT5G67020. However, phylogenetic analysis of the OrthoMCL cluster containing FRAEX38873\_v2\_000000770 indicates it is orthologous to both AT5G67020 and AT3G50340 (Supplementary Fig. 3i). Although the function of AT5G67020 is unknown it is suggested to be a possible target of members of the *R2R3-MYB* gene family, which appear to control accumulation of flavonols<sup>137</sup>. AT5G67020 has reduced expression in a mutant where three *R2R3-MYB* genes have been knocked out; other genes that also show a reduction in expression include those encoding known flavonoid biosynthesis enzymes, as well as the *4CL3* gene (AT1G65060) that functions in the general phenylpropanoid pathway<sup>137</sup>. AT3G50340 is also not fully characterised but has been identified as an auxin-inducible gene<sup>138</sup> and is also suggested to be regulated by drought, abscisic acid and jasmonic acid<sup>139</sup>.

**OG28712 (FRAEX38873\_v2\_000047370).** Grand-conv detected evidence of a convergent histidine and methionine in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have alanine and threonine (Supplementary

Table 5). The top match for this gene in *A. thaliana* is *PROTOCHLOROPHYLLIDE OXIDOREDUCTASE A* (AT5G54190), also known as *PORA*. AT5G54190/*PORA* encodes light-dependent NADPH:protochlorophyllide oxidoreductase A and is involved in chlorophyll biosynthesis ([www.arabidopsis.org/servlets/TairObject?type=locus&name=AT5G54190](http://www.arabidopsis.org/servlets/TairObject?type=locus&name=AT5G54190)). The putative orthologue of OG28712/FRAEX38873\_v2\_000047370 in *S. lycopersicum* is Solyc12g013710; this gene has been identified as a mycorrhiza-regulated gene<sup>140</sup>, and is upregulated in plants colonised with a mycorrhizal fungus along with a number of other genes involved in photosynthesis<sup>140</sup>. Expression of Solyc12g013710 is repressed by overexpression of *SIRBZ* (Solyc03g033560<sup>141</sup>) which is the putative orthologue of another of the candidate loci, OG32176. One of the susceptible taxa included in our convergence analyses, *F. pennsylvanica*, may lack a functional copy of this gene (see Supplementary Table 6).

**OG30208 (FRAEX38873\_v2\_000228080).** Grand-conv detected evidence of a convergent histidine and valine in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have a glutamine and alanine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoded an uncharacterised transmembrane protein, DUF677 (AT1G20180). AT1G20180 is upregulated in two *A. thaliana* autophagy mutant lines in response to differing amounts of nitrogen<sup>142</sup>; other genes upregulated at the same time were predominantly involved in response to biotic stress, chemical and abiotic stress and salicylic acid<sup>142</sup>. The FRAEX38873\_v2\_000228080 reference gene model from *F. excelsior* is truncated relative to the other sequences in the OG, which may be due to a misassembly introducing a premature stop codon or might indicate that the *F. excelsior* reference individual does not have a full-length copy of the gene (Supplementary Table 5). One of the susceptible taxa included in our convergence analyses, *F. ornus*, may lack a functional copy of this gene (Supplementary Table 6).

**OG32176 (FRAEX38873\_v2\_000268710).** Grand-conv detected evidence of a convergent histidine and leucine in *F. mandshurica* and *F. platypoda*, at sites where the other *Fraxinus* species in the analysis have an glutamine and serine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *STRESS ASSOCIATED RNA-BINDING PROTEIN 1* (AT2G17975), also known as *SRP1*;



FRAEX38873\_v2\_000268710 and AT2G17975/*SRP1* appear orthologous on the basis of phylogenetic analysis of the OrthoMCL cluster to which they belong (Supplementary Fig. 3j). AT2G17975/*SRP1* is an RNA-binding protein involved in the post-transcriptional regulation of abscisic acid (ABA) signalling via the modulation of *ABI* genes<sup>143</sup>; expression of AT2G17975 is down regulated in response to ABA and abiotic stress<sup>143</sup>. The apparent orthologue of OG32176/FRAEX38873\_v2\_000268710 in *S. lycopersicum* is Solyc03g033560 (Supplementary Fig. 3j); this gene is also known as *SIRBZ* and is a RanBP2-type zinc finger protein gene<sup>141</sup>. Mutants that overexpress *SIRBZ* have a dwarf phenotype and are chlorotic<sup>141</sup>. Overexpression of *SIRBZ* also leads to decreased expression of photosynthesis genes, including Solyc12g013710 which is the putative orthologue of another of the candidate loci, OG28712; gibberellic acid (GA) biosynthesis genes were also repressed<sup>141</sup>. It is suggested that *SIRBZ* regulates the formation of chloroplasts, and as a result also controls chlorophyll, carotenoid and GA biosynthesis, which take place in the chloroplast<sup>141</sup>. GA may play a role in regulating anti-herbivore defense<sup>27</sup>.

**OG33348 (FRAEX38873\_v2\_000106520).** Grand-conv detected evidence of a convergent serine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an asparagine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *S-ADENOSYLMETHIONINE DECARBOXYLASE 4* (AT5G18930), or *SAMDC4*. AT5G18930/*SAMDC4* is also known as *BUSHY AND DWARF 2 (BUD2)* and is involved in the synthesis of S-adenosylmethioninamine from S-adenosyl-L-methionine within the S-adenosylmethioninamine biosynthesis pathway (<https://www.uniprot.org/uniprot/Q3E9D5>). Phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000106520 and AT5G18930 indicates they are likely to be orthologous, albeit the relevant node is not well supported (PP<0.95; Supplementary Fig. 3k). AT5G18930/*SAMDC4/BUD2* is essential for biosynthesis of the polyamines spermidine and spermine<sup>144</sup>; polyamines in higher plants can be involved in mediating biotic and abiotic stress responses, such as pathogen infection, osmotic stress and wounding<sup>144</sup>. Spermine in particular is suggested to play a role in defence response signalling<sup>36</sup>. Loss of function of AT5G18930/*SAMDC4/BUD2* alters growth and development<sup>144</sup> and analysis of its promoter region found elements that are associated with response to auxin, dehydration, drought, salt, cold, wounding, defence related gene expression and wound response, amongst other physiological responses<sup>145</sup>.

**OG37560 (FRAEX38873\_v2\_000076410).** Grand-conv detected evidence of a convergent alanine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a proline (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a heavy metal transport/detoxification superfamily protein (AT5G27690). AT5G27690 has also been called *AthHIPP36* or *AthHIP36* and is a member of a heavy metal-associated isoprenylated plant protein (HIPP) family<sup>146</sup>. HIPPs may have roles in heavy metal homeostasis and detoxification, transcriptional responses to cold and drought, and plant-pathogen interactions<sup>146</sup>. However, FRAEX38873\_v2\_000076410 does not belong to the same OrthoMCL cluster as AT5G27690, and appears not to have an orthologue in *A. thaliana*. FRAEX38873\_v2\_000076410 does belong to the same OrthoMCL cluster as, and appears orthologous to (Supplementary Fig. 3I), the *P. trichocarpa* gene Potri.010G024700 (also called POPTR\_0010s02550), which encodes a putative copper (Cu) transport protein (<https://phytozome.jgi.doe.gov/phytozome/report.do?id=49365037&trail=%7c49365037>) and belongs to the HIPP family (*PtrHIP43*<sup>146</sup>). It has been suggested that mechanisms for signalling and responding to Cu stress could overlap with those involved in biotic stress, possibly relating to both types of stress causing the production of reactive oxygen species<sup>147</sup>; exposure to high Cu can prime maize plants for increased production of volatile organic compounds and faster phytohormone signalling upon subsequent herbivory by an insect herbivore<sup>147</sup>. Moreover, copper transport genes have been implicated in the regulation of defence response to whitefly (*Bemisia tabaci*) in cotton<sup>148</sup>. One of the susceptible taxa included in our convergence analyses, *F. latifolia*, may lack a functional copy of this gene (see Supplementary Table 6).

**OG37870 (FRAEX38873\_v2\_000034500).** Grand-conv detected evidence of a convergent glutamine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an arginine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *METACASPASE 9* (AT5G04200), or *AtMC9*. Results of phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000034500 and AT5G04200/*AtMC9* are compatible with the inference that these genes are orthologues (although the topology does not fit exactly

with expected species relationships, suggesting the *A. thaliana* and *F. excelsior* sequences may be from different paralogues, this is not well supported; Supplementary Fig. 3m). AT5G04200/*AtMC9* plays a role in controlling autophagy in tracheary elements, ensuring that cell death does not spread to surrounding non-target cells<sup>149</sup>. The site with evidence of convergence in *Fraxinus* is located within a linker region between p20 and p10-like domains of the *AtMC9* protein<sup>150</sup>; the linker region appears to be important to the function of *AtMC9*, as its removal causes a significant reduction in the enzyme's activity<sup>150</sup>. Two poplar genes, Potri.006G026500.1 and Potri.016G024500.1, appear orthologous to FRAEX38873\_v2\_000034500 (Supplementary Fig. 3m) and are known as *PtMC13* and *PtMC14*, respectively<sup>42</sup>. Both poplar genes are suggested to play a role in controlling cell death in xylem elements<sup>42</sup>.

**OG38407 (FRAEX38873\_v2\_000324730).** Grand-conv detected evidence of a convergent aspartic acid in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an asparagine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *SNC1-INFLUENCING PLANT E3 LIGASE REVERSE GENETIC SCREEN 4* (AT3G48880), or *SNIPER4*, which encodes an F-box protein. Phylogenetic analysis of their OrthoMCL cluster indicates that FRAEX38873\_v2\_000324730 and AT3G48880/*SNIPER4* are likely to be orthologues (Supplementary Fig. 3n). AT3G48880/*SNIPER4* was found to be a positive regulator of effector triggered immunity (ETI) in *A. thaliana* and has a role in maintaining the balance of immune response by forming part of a complex directing the MUSE13 and MUSE14 proteins for degradation<sup>28</sup>. The targeted degradation of MUSE13/14 ensures sufficient levels of SNC1, an immune sensor, are available to trigger defence response against a bacterial pathogen<sup>28</sup>.

**OG38543 (FRAEX38873\_v2\_000362960).** Grand-conv detected evidence of a convergent aspartic acid in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have an asparagine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a cyclin-like protein (AT3G19650). Specific functional information is lacking for AT3G19650, however, cyclins regulate activity of cyclin-dependent kinases, which play critical roles in the control of cell cycle progression<sup>151</sup>.

**OG39275 (FRAEX38873\_v2\_000319270).** Grand-conv detected evidence of a convergent glutamic acid in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a glutamine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *ISOPENTENYLTRANSFERASE 6* (AT1G25410), also called *IPT6*. Phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000319270 indicates that, in addition to AT1G25410/*IPT6*, it is equally closely related to two other *A. thaliana* genes (AT1G68460/*IPT1* and AT3G19160/*IPT8*) and may be orthologous to all three (Supplementary Fig. 3o). AT1G68460/*IPT1*, AT1G25410/*IPT6* and AT3G19160/*IPT8* belong to a plant-specific ATP/ADP IPT clade of genes, which synthesize iP- and tZ-type cytokinins<sup>152</sup>. Cytokinins play a role in plant growth, defence response and immunity; they help to mediate trade-offs between growth and defence<sup>31,153</sup>. The FRAEX38873\_v2\_000319270 reference gene model from *F. excelsior* appears to be incorrect, possibly due to a misassembly causing an internal inversion (Supplementary Table 5).

**OG40061 (FRAEX38873\_v2\_000309460).** Grand-conv detected evidence of a convergent asparagine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a serine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a NAD(P)-binding Rossmann-fold superfamily protein (AT2G23910), which is reported to be involved in lignin biosynthetic process and response to karrikin and to cinnamoyl-CoA reductase activity and oxidoreductase activity (<https://www.arabidopsis.org/servlets/TairObject?name=AT2G23910&type=locus>). However, phylogenetic analysis indicates that FRAEX38873\_v2\_000309460 is paralogous to AT2G23910, along with AT4G30470 (another NAD(P)-binding Rossmann-fold superfamily protein), and may lack a direct orthologue in *A. thaliana* (Supplementary Fig. 3p). AT2G23910 is also known as *CCR(Cinnamoyl-CoA reductase)6*<sup>154</sup>, *CCR-like8*<sup>155</sup> or *CCRL9*<sup>156</sup> and is co-expressed with genes involved in flavonoid metabolism<sup>156</sup>. AT4G30470 is also known as *CCR-like6*<sup>155</sup> and, in common with the best matching *A. thaliana* gene for another of our candidates (OG27080, see above), is a possible target of transcription factors in the *R2R3-MYB* gene family that appear to control accumulation of flavonols<sup>137</sup>. The apparent orthologue of OG40061/FRAEX38873\_v2\_000309460 in *S. lycopersicum* is Solyc03g097170 (Supplementary Fig. 3p); Solyc03g097170 was found to be upregulated in response to

gibberellin and has been defined as a DELLA-dependent gene<sup>157</sup>. The serine variant in *Fraxinus* is a putative phosphorylation site (Supplementary Table 5), and is found on the exterior of the protein in the predicted model for OG40061/FRAEX38873\_v2\_000309460 in close proximity to the putative active site (Supplementary Fig. 2b). The FRAEX38873\_v2\_000309460 reference gene model from *F. excelsior* appears to be truncated on the 3' end due to the genome assembly being incomplete (Supplementary Table 5).

**OG41448 (FRAEX38873\_v2\_000207320).** Grand-conv detected evidence of a convergent alanine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have glycine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *JASSY* (AT1G70480); phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000207320 and AT1G70480/*JASSY* indicates they are likely to be orthologous (Supplementary Fig. 3q). AT1G70480/*JASSY* encodes a chloroplast outer membrane protein that is involved in the transport of the JA precursor 2-oxophytodienoic acid (OPDA) from the chloroplast<sup>158</sup>. AT1G70480/*JASSY* loss-of-function of mutants show increased susceptibility to a fungal pathogen compared with wild-type plants and also lacked activation of JA-responsive genes following wounding, suggesting that the gene is essential for initiation of JA-signalling pathways<sup>158</sup>. Further analysis revealed that the defects observed in the mutants result from a failure of JA accumulation in the absence of AT1G70480/*JASSY* expression<sup>158</sup>. *JASSY* belongs to the Bet v1-like protein superfamily<sup>158</sup>; a related protein has been proposed to play a role in resistance to EAB in *F. mandshurica*<sup>75</sup>.

**OG41488 (FRAEX38873\_v2\_000363060).** Grand-conv detected evidence of a convergent threonine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have alanine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *AT-HOOK MOTIF NUCLEAR LOCALIZED PROTEIN 17* (AT5G49700), or *AHL17*. Phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000363060 indicates that it is likely to be orthologous to both AT5G49700/*AHL17* and AT1G14490/*AHL28* (Supplementary Fig. 3r). AT5G49700/*AHL17* and belong AT1G14490/*AHL28* to the AHL gene family in *A. thaliana*, members of which are involved in modulating plant growth and development<sup>159</sup>. *AHL* genes may also play a role in regulating homeostasis of

gibberellins, jasmonic acid and cytokinins, and some *AHL* genes may be involved in regulating plant defence response (reviewed by Zhao et al.,<sup>160</sup>). The apparent orthologue of OG41488/FRAEX38873\_v2\_000363060.1 in *S. lycopersicum* is Solyc04g076220 and is also known as *AHL17a*<sup>161</sup> (Howden et al., 2017); the potential role of Solyc04g076220/*AHL17a* in immunity to a fungal pathogen in tomato has been investigated, but over-expression this gene did not visibly impact immunity<sup>161</sup>.

**OG46977 (FRAEX38873\_v2\_000304400).** Grand-conv detected evidence of a convergent arginine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a serine (two taxa with a glutamine instead were found to have errors in their predicted gene models; Supplementary Table 5). The top match for this gene in *A. thaliana* is a transmembrane proteins 14C gene (AT3G43520). AT3G43520 is also known as *FATTY ACID EXPORT 2 (AtFAX2)* and may be involved in free fatty acids export from the plastids (<https://www.uniprot.org/uniprot/Q94A32>). The FRAEX38873\_v2\_000304400 reference gene model from *F. excelsior* appears to be incorrect on the 3' end, possibly due to a misassembly (Supplementary Table 5).

**OG47629 (FRAEX38873\_v2\_000173180).** Grand-conv detected evidence of a convergent aspartic acid in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a glycine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a MIZU-KUSSEI-like protein of unknown function (AT2G37880). AT2G37880 has been identified as an early high-light responsive gene, showing increased expression in an *A. thaliana* cell culture 30 min after exposure to high light stress conditions<sup>162</sup>. The putative orthologue of OG47629/FRAEX38873\_v2\_000173180 in *S. lycopersicum* is Solyc09g011350 and has been identified as a "disease-specific gene" that is differentially regulated in tomato leaves at 24 hours post inoculation with three pathogens<sup>163</sup>. Solyc09g011350 was upregulated in the infected leaves compared with the control, specifically in response to *Botrytis cinerea*<sup>163</sup>.

**OG49074 (FRAEX38873\_v2\_000159740).** Grand-conv detected evidence of a convergent valine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a glycine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *MEMBRANE-ASSOCIATED PROGESTERONE BINDING*

*PROTEIN 4* (AT4G14965), or *MAPR4*. Although the function of AT4G14965/*MAPR4* has not been fully characterised, GO terms have been associated with it via identification of functional gene modules<sup>118</sup>, which are: NEGATIVE REGULATION OF CELLULAR PROCESS and POSTREPLICATION REPAIR (see [http://bioinformatics.psb.ugent.be/cig\\_data/plant\\_modules/createResultsHTML.php?geneID=AT4G14965](http://bioinformatics.psb.ugent.be/cig_data/plant_modules/createResultsHTML.php?geneID=AT4G14965)).

**OG50989 (FRAEX38873\_v2\_000237270)**. Grand-conv detected evidence of a convergent serine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a proline (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a calcium-binding EF-hand family protein (AT4G26470), which functions in calcium ion binding and calcium-mediated signaling ([www.arabidopsis.org/servlets/TairObject?type=locus&name=At4g26470](http://www.arabidopsis.org/servlets/TairObject?type=locus&name=At4g26470)) and appears orthologous to FRAEX38873\_v2\_000237270 on the basis of phylogenetic analysis of their OrthoMCL cluster (Supplementary Fig. 3s). AT4G26470 is also known as *CML21* (short for *Calmodulin-like protein 21*)<sup>164</sup> and is transcriptionally upregulated during pollen tube growth<sup>165</sup>. AT4G26470/*CML21* also shows increased expression in response to melatonin<sup>166</sup>; many of the other genes that also showed altered expression in response to melatonin levels were involved in stress defence response, indicating an important role for melatonin triggering defence response to both abiotic and biotic stresses<sup>166</sup>. Calcium signalling plays a key role in plant defence response<sup>167</sup> including in pathogen-triggered programmed cell death<sup>33</sup> and Ca<sup>2+</sup> influx is one of the early signals of feeding insects and is involved in triggering defense response<sup>17</sup>.

**OG59564 (FRAEX38873\_v2\_000174000)**. Grand-conv detected evidence of a convergent aspartic acid in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a glutamic acid (Supplementary Table 5). The top match for this gene in *A. thaliana* is *Na-translocating NADH-quinone reductase subunit A* (AT5G55640). AT5G55640 is also known as *MDF20.8* ([www.uniprot.org/uniprot/Q9FM74](http://www.uniprot.org/uniprot/Q9FM74)). Information on the precise function of AT5G55640/*MDF20.8* is lacking, however, quinone reductases are considered to be detoxifying enzymes and have a role in protecting organisms from oxidative stress<sup>168</sup>. One of the susceptible taxa included in our convergence analyses, *F. latifolia*, may lack a functional copy of this gene (Supplementary Table 6).

**OG60899 (FRAEX38873\_v2\_000255120).** Grand-conv detected evidence of a convergent serine in *F. mandshurica* and *F. platypoda*, at a site where the other *Fraxinus* species in the analysis have a glycine. However, the serine is also found in the alternate allele for susceptible *F. ornus* (Supplementary Table 5). The convergent site is predicted to be within a signal peptide, but the amino acid variant is not suggested to alter the localisation of the protein (Supplementary Table 5). The top match for this gene in *A. thaliana* is *EARLY NODULIN-LIKE PROTEIN 7* (AT1G79800), or *ENODL7*. Although the function of AT1G79800 has not been fully characterised, GO terms have been associated with it via identification of functional gene modules<sup>118</sup>, which are: COPPER ION BINDING and ELECTRON CARRIER ACTIVITY ([http://bioinformatics.psb.ugent.be/cig\\_data/plant\\_modules/createResultsHTML.php?geneID=AT1G79800](http://bioinformatics.psb.ugent.be/cig_data/plant_modules/createResultsHTML.php?geneID=AT1G79800)). Two *P. trichocarpa* genes belong to the same OrthoMCL cluster as FRAEX38873\_v2\_000255120 (Sollars et al.<sup>11</sup>) both of which (Potri.001G187700 and Potri.003G050500) have been categorised as AGP genes (highly glycosylated arabinogalactan-proteins), and are known as *PtPAG36* and *PtPAG37*<sup>169</sup>. However, there is no specific information on the function of these genes. One of the resistant taxa included in our convergence analyses, *F. floribunda*, may lack a functional copy of this gene (Supplementary Table 6). The FRAEX38873\_v2\_000255120 reference gene model from *F. excelsior* appears to be incorrect, possibly due to a misassembly (Supplementary Table 5).

#### **4.4. Candidates found in *F. mandshurica* and resistant taxa in section *Ornus* (grand-conv analysis 2)**

**OG21449 (FRAEX38873\_v2\_000266260).** Grand-conv detected evidence of a convergent threonine in *F. mandshurica*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at a site where the other *Fraxinus* species in the analysis have an isoleucine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *PHOTOLYASE/BLUE-LIGHT RECEPTOR 2* (AT2G47590), also known as *PHR2*. Photolyases are light activated and repair damage caused to DNA by UV radiation; *PHR2* photolyases are specific to plants<sup>170</sup>.

**OG23284 (FRAEX38873\_v2\_000124490).** Grand-conv detected evidence of a convergent phenylalanine and serine in *F. mandshurica*, *F. baroniana*, *F. floribunda* and



*Fraxinus* sp. D2006-0159, at sites where the other *Fraxinus* species in the analysis have a cysteine and a leucine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding an alpha/beta-Hydrolases superfamily member (AT3G19970); phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000124490 and AT3G19970 indicates they are likely to be orthologous (Supplementary Fig. 3t). Although the function of AT3G19970 has not been fully characterised, GO terms have been associated with it via identification of functional gene modules<sup>118</sup>, including: CALLOSE DEPOSITION DURING DEFENSE RESPONSE, SALICYLIC ACID MEDIATED SIGNALING PATHWAY, JASMONIC ACID MEDIATED SIGNALING PATHWAY, REGULATION OF PLANT-TYPE HYPERSENSITIVE RESPONSE, REGULATION OF IMMUNE RESPONSE, RESPONSE TO CHITIN (see [http://bioinformatics.psb.ugent.be/cig\\_data/plant\\_modules/createResultsHTML.php?geneID=AT3G19970](http://bioinformatics.psb.ugent.be/cig_data/plant_modules/createResultsHTML.php?geneID=AT3G19970)) indicating that this gene may potentially have a role related to plant defence response. The apparent orthologue of FRAEX38873\_v2\_000382300 in *S. lycopersicum* is Solyc02g020890 (Supplementary Fig. 3t); Solyc02g020890 was among 4774 genes in tomato whose expression was significantly affected by application of methyl jasmonate<sup>171</sup>.

**OG47560 (FRAEX38873\_v2\_000262780).** Grand-conv detected evidence of a convergent lysine and phenylalanine in *F. mandshurica*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at sites where the other *Fraxinus* species in the analysis have a glutamic acid and tyrosine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *TAPETUM 1* (AT3G42960), or *TAI*. AT3G42960/*TAI* is also known as *ASD* or *ATAI* and is expressed in tapetal cells; it has been reported to be involved in flower development and to have oxidoreductase activity ([www.arabidopsis.org/servlets/TairObject?name=AT3G42960&type=locus](http://www.arabidopsis.org/servlets/TairObject?name=AT3G42960&type=locus)). Phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000262780 and AT3G42960/*TAI* suggests they may be paralogues, although this result is not conclusive (Supplementary Fig. 3u). The apparent orthologue of OG47560/FRAEX38873\_v2\_000262780 in *S. lycopersicum* is Solyc11g018600 (Supplementary Fig. 3u) and has decreased expression in a male-sterile tomato mutant<sup>172</sup>. The PhytoMine entry for Solyc11g018600 predicts that this gene is involved in the abscisic acid (ABA) biosynthesis pathway

(<https://phytozome.jgi.doe.gov/phytozome/report.do?id=286588936&trail=%7c286588936>). AT3G42960/*TAI* is also predicted to be involved in the ABA biosynthesis pathway, although this has not yet been confirmed with experimental evidence (<https://pmn.plantcyc.org/ARA/NEW-IMAGE?type=PATHWAY&object=PWY-695>). A major function of ABA is response to abiotic stress<sup>31</sup>; ABA also plays an important role in triggering stomatal closure, which represents a plant defence mechanism<sup>31</sup>. Evidence suggests that ABA interacts with other phytohormones to influence response to herbivory and plant defence against pathogens<sup>30,31</sup>.

**OG56563 (FRAEX38873\_v2\_000135960).** Grand-conv detected evidence of a convergent valine in *F. mandshurica*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at sites where the other *Fraxinus* species in the analysis have an alanine or threonine (Supplementary Table 5). However, the valine is also found in the alternate allele for susceptible *F. ornus* (Supplementary Table 8). The top match for this gene in *A. thaliana* is *CASP-LIKE PROTEIN 4D1* (AT2G39530), *CASPL4D1*. The function of AT2G39530/*CASPL4D1* is not well characterised, but it has been shown to have reduced expression in response to drought stress<sup>173</sup>. AT2G39530/*CASPL4D1* also showed upregulation in *A. thaliana* plants infiltrated with an avirulent strain of *Pseudomonas syringae* pv. tomato but downregulation after ABA induced susceptibility followed by pathogen inoculation<sup>174</sup>. However, FRAEX38873\_v2\_000135960 does not belong to the same OrthoMCL cluster as AT2G39530 (Sollars et al.<sup>11</sup>), and appears not to have an orthologue in *A. thaliana*. The putative orthologue of OG56563/FRAEX38873\_v2\_000135960 in *S. lycopersicum* is Solyc03g098090, but information on the function of this gene is apparently lacking.

**OG64545 (FRAEX38873\_v2\_000393500).** Grand-conv detected evidence of a convergent serine in *F. mandshurica*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159 in a position where the other *Fraxinus* species in the test have an asparagine (Supplementary Table 5). The top match for this gene in *A. thaliana* is a gene encoding a single hybrid motif superfamily protein (AT2G35120) and is involved in glycine decarboxylation via the glycine cleavage system ([www.arabidopsis.org/servlets/TairObject?name=AT2G35120&type=locus](http://www.arabidopsis.org/servlets/TairObject?name=AT2G35120&type=locus)). AT2G35120 is also known as *GDCH2*<sup>175</sup> or *GDH2* (short for *Glycine cleavage system H protein 2*; [www.uniprot.org/uniprot/O82179](http://www.uniprot.org/uniprot/O82179)). The glycine decarboxylase (GDC)

multienzyme system (also known as the glycine cleavage system; [www.uniprot.org/uniprot/O82179](http://www.uniprot.org/uniprot/O82179)) consists of four enzymes (P, H, T and L proteins) and catalyses the destruction of glycine molecules emitted from the peroxisomes during photorespiration<sup>176</sup>. The H protein shuttles the methylamine group of glycine from the P protein to the T protein ([www.uniprot.org/uniprot/O82179](http://www.uniprot.org/uniprot/O82179)). GDC might be involved in plant stress-response, via cross-talk between nitric oxide and reactive oxygen species<sup>177</sup>. Two *P. trichocarpa* genes belong to the same OrthoMCL cluster as FRAEX38873\_v2\_000393500<sup>11</sup> (Supplementary Fig. 3v) both of which (Potri.012G123700 and Potri.015G122500, also known as *gdcH1* and *gdcH2* respectively; <https://phytozome.jgi.doe.gov/>) have correlated expression with several genes involved in the phenylpropanoid pathway, including a *P. trichocarpa* *CYP98A3* putative orthologue (Potri.006G033300; <https://phytozome.jgi.doe.gov/phytozome/report.do?id=48451623>) which belongs to the same OrthoMCL cluster as another of our candidate loci, OG15551/FRAEX38873\_v2\_000261700 (Supplementary Fig. 3a). Potri.012G123700/*gdcH1* was found to be most highly expressed in xylem tissues and Potri.015G122500/*gdcH2* also shows evidence of higher expression in wood tissues than in numerous other tissue types examined<sup>178</sup> (see <http://popgenie.org/gene?id=Potri.T122200>). Both poplar genes contain the AC element within their promoters<sup>179</sup>. This element also occurs within many phenylpropanoid biosynthesis genes in *P. trichocarpa* (including *PAL*, *C4H*, *C3'H*, *CCR* and *CAD*), suggesting that the expression of Potri.012G123700/*gdcH1* and Potri.015G122500/*gdcH2* is coordinated with that of phenylpropanoid genes to support lignin and flavonoid biosynthesis<sup>179</sup>. This role for GDC H protein genes may be a tree-specific adaptation in response to high demands for one-carbon units during lignification, which is distinct from the role of *GDC* genes in *A. thaliana*<sup>179</sup>.

#### 4.5. Candidates found in *F. platypoda* and resistant taxa in section *Ornus* (grand-conv analysis 3)

**OG16739 (FRAEX38873\_v2\_000338500)**. Grand-conv detected evidence of a convergent alanine in *F. platypoda*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at a site where the other *Fraxinus* species in the analysis have a threonine (Supplementary Table 5). However, the alanine is also found in the alternate allele for susceptible *F. ornus* (Supplementary Table 8). The top match for this gene in *A. thaliana*

is a gene encoding a RNA-binding (RRM/RBD/RNP motifs) protein family member (AT2G41060), which is also known as *UBA2B*. Phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000338500 and AT2G41060 indicates they are paralogues rather than orthologues (Supplementary Fig. 3w) and that FRAEX38873\_v2\_000338500 is equally closely related to AT3G56860, also known as *UBA2A*. Overexpression of both *UBA2A* and *UBA2B* leads to increased expression of a number of senescence-associated and defence related genes, as well as to increased ethylene biosynthesis, hypersensitive-like patterns of cell death and callose deposition<sup>41</sup>. Expression of the *UBA2* genes was not found to increase in relation to natural, age-related, senescence, but instead appears connected to wound-induced cell-death and senescence pathways, supporting the suggestion that these genes play a role in regulating senescence in response to wounding and other environmental cues<sup>41</sup>.

**OG20520 (FRAEX38873\_v2\_000382300)**. Grand-conv detected evidence of a convergent isoleucine in *F. platypoda*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at a site where the other *Fraxinus* species in the analysis have a serine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *SHOOT GRAVITROPISM 7* (AT4G37650), which is also known as *SHORT ROOT (SHR)*. AT4G37650/*SHR* is essential for root ground tissue patterning and interacts with *SCARECROW/SCR*<sup>180</sup>; they are thought to function as transcriptional regulators<sup>180</sup>. The putative orthologue of OG20520/FRAEX38873\_v2\_000382300 in *S. lycopersicum* is Solyc02g092370; this gene is also known as *SlSHRa* and is involved in root radial patterning<sup>181</sup>. The putative orthologue from *Populus trichocarpa*, Potri.007G063300 (being the only *P. trichocarpa* gene belonging to the same OrthoMCL cluster as FRAEX38873\_v2\_000382300<sup>11</sup>), is also known as *PtSHR1*<sup>182</sup> and down-regulation of this gene in transgenic hybrid poplar leads to increased height and girth, indicating that *SHR* is involved in regulation of cell division and meristem activity in shoots as well as roots<sup>183</sup>.

**OG21033 (FRAEX38873\_v2\_000092000)**. Grand-conv detected evidence of a convergent lysine in *F. platypoda*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at a site where the other *Fraxinus* species in the analysis have a glutamine (Supplementary Table 5). The top match for this gene in *A. thaliana* is an O-Glycosyl hydrolases family 17 protein encoding gene (AT3G55430); proteins in this family

include those with a number of  $\beta$ -glucosidase related activities ([www.cazy.org/GH17.html](http://www.cazy.org/GH17.html)) and are involved in hydrolysis of glycosidic bonds in carbohydrates (<http://www.ebi.ac.uk/interpro/entry/IPR000490>). Phylogenetic analysis of the OrthoMCL cluster including FRAEX38873\_v2\_000092000 indicates it is equally closely related to AT3G55430 and another O-Glycosyl hydrolases family 17 protein encoding gene from *A. thaliana*, AT2G39640, and appears orthologous to both (Supplementary Fig. 3x). Expression of AT3G55430 seems to be at least in part dependent on nitric oxide, and *A. thaliana* mutants where AT3G55430 function is impaired have increased susceptibility to the fungal pathogen *Botrytis cinerea*<sup>184</sup>. It is suggested that the control of AT3G55430 expression via nitric oxide might form part of the mechanism for basal resistance to *B. cinerea* in *A. thaliana*<sup>184</sup>. In the predicted protein model for OG21033/FRAEX38873\_v2\_000092000 the lysine/glutamine variant is close to the active site (Supplementary Fig. 2d).

**OG27693 (FRAEX38873\_v2\_000056890).** Grand-conv detected evidence of a convergent isoleucine in *F. platypoda*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at a site where the other *Fraxinus* species in the analysis have a threonine or alanine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *HESPERIN* (AT1G31500); this gene is a transcriptional regulator of circadian rhythms but apparently also has a role in response to oxidative stress<sup>185</sup>.

**OG35707 (FRAEX38873\_v2\_000303580).** Grand-conv detected evidence of a convergent leucine in *F. platypoda*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at a site where the other *Fraxinus* species in the analysis have a phenylalanine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *ZRT/IRT-LIKE PROTEIN 2* (AT5G59520), or *ZIP2*. AT5G59520/*ZIP2* probably plays a role in Mn (and possibly Zn) transport into the root vasculature for translocation to the shoot<sup>186</sup>. In plants, Mn is a cofactor in processes such as photosynthesis, lipid biosynthesis and oxidative stress<sup>187</sup>. Three *P. trichocarpa* genes belong to the same OrthoMCL cluster as FRAEX38873\_v2\_000303580 (Sollars et al.<sup>11</sup>) one of which, Potri.009G034600, is also known as *ZIP2* and is involved in Cd<sup>2+</sup> uptake<sup>188</sup>.

**OG36502 (FRAEX38873\_v2\_000266620).** Grand-conv detected evidence of a convergent serine in *F. platypoda*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-

0159, at a site where the other *Fraxinus* species in the analysis have an asparagine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *TRNA METHYLTRANSFERASE 140B* (AT1G54650), which encodes a methyltransferase family protein. AT1G54650/*TRM140B* is proposed to have a role in tRNA nucleoside methylation, generating m<sup>3</sup>C (3-methylcytidine) modifications<sup>189</sup>. The serine variant in *Fraxinus* is a putative phosphorylation site (Supplementary Table 5), and is found on the exterior of the protein in the predicted model for OG36502/FRAEX38873\_v2\_000266620 (Supplementary Fig. 2a).

#### **OG43828 (FRAEX38873\_v2\_000227880).**

Grand-conv detected evidence of a convergent leucine in *F. platypoda*, *F. baroniana*, *F. floribunda* and *Fraxinus* sp. D2006-0159, at a site where the other *Fraxinus* species in the analysis have a serine (Supplementary Table 5). The top match for this gene in *A. thaliana* is *ATDET2* (AT2G38050), a gene involved in the brassinolide (brassinosteroid) biosynthetic pathway. Phylogenetic analysis indicates that FRAEX38873\_v2\_000227880 and AT2G38050/*ATDET2* are likely to be orthologues (Supplementary Fig. 3y). AT2G38050/*ATDET2* is also known as *DWARF 6* (*DWF6*) and is involved in multiple reactions within the brassinosteroid biosynthesis pathway<sup>190</sup>. Brassinosteroids are hormones whose functions in plants include regulating growth and development<sup>191</sup> and which also play an important part in modulating growth–defence trade-offs<sup>135</sup>. Brassinosteroids in plants have potential role in mediating response to stresses such as freezing, drought, salinity, disease, heat and nutrient deficiency<sup>191</sup>. There is also evidence for a role for brassinosteroids in regulating glucosinolate profiles, which function in defence response against insects in *A. thaliana*<sup>192</sup>. The *det2-1* mutant in *A. thaliana* has impaired stomatal opening in response to blue light, suggesting a further possible role for brassinosteroids in stomatal opening<sup>193</sup>. One of the apparent orthologues of OG43828/FRAEX38873\_v2\_000227880 in *P. trichocarpa*, Potri.T122200 (Supplementary Fig. 3y), was found to be preferentially expressed in beetle damaged leaves<sup>178</sup> (see <http://popgenie.org/gene?id=Potri.T122200>). The FRAEX38873\_v2\_000227880 reference gene model from *F. excelsior* appears to be incomplete on the 5' end, apparently due to a frameshift induced by the assembly of a chimeric haplotype combining variants from both alleles (Supplementary Table 5).

#### **Supplementary Note 5**

## 5. Significant enrichment of GO terms

Three GO terms are significantly enriched among the refined list of 53 candidates ( $p$ -value  $< 0.01$ ) relative to the set of genes included in the grand-conv analyses, according to Fisher's exact test run with the weight algorithm, all within the biological process (BP) domain and all related to hormone metabolic processes: estrogen metabolic process, C21-steroid hormone metabolic process and androgen metabolic process (Supplementary Table 7). These terms are associated with two of the 53 candidates, OG40061 (FRAEX38873\_v2\_000309460) and OG43828 (FRAEX38873\_v2\_000227880), the latter of which appears to be involved in brassinosteroid biosynthesis (see Supplementary Note 4.5). The same terms were found to be significantly enriched when running Fisher's exact test run with the elim algorithm. However, two additional terms within the BP domain were also significantly enriched: hormone biosynthetic process and organic cyclic compound biosynthetic process (Supplementary Table 7). Three of the candidate genes are associated with the term "hormone biosynthetic process", OG16739 (FRAEX38873\_v2\_000338500.1), OG39275 (FRAEX38873\_v2\_000319270.1) and OG43828 (FRAEX38873\_v2\_000227880). Nine candidate genes are associated with the term "organic cyclic compound biosynthetic process" including some that appear to be involved in phytohormone biosynthesis (Supplementary Table 7 and Supplementary Note 4). Also, two terms within the molecular function (MP) domain were found to be significantly enriched when using the elim algorithm: steroid dehydrogenase activity and carboxy-lyase activity (Supplementary Table 7). Both of these terms are associated with OG40061 (FRAEX38873\_v2\_000309460) and OG43828 (FRAEX38873\_v2\_000227880). Some of the genes among the top 53 candidates did not have any GO terms annotated by the analysis performed by (Sollars et al.<sup>11</sup>), which may explain why some of the candidates that appear to be involved in phytohormone biosynthesis (Supplementary Note 4) are not associated with the significantly enriched GO terms.

## Supplementary Note 6

### 6. Differential expression in response to EAB-feeding

For 29 the 53 loci in our refined set of candidates we were able to identify the likely orthologue from the published *F. pennsylvanica* transcriptome<sup>25</sup>; of these 29 *F. pennsylvanica* genes, 11 showed evidence for differential expression subsequent to EAB-feeding (four with decreased expression, and seven with increased expression;

Supplementary Table 10)<sup>25</sup>. However, none of these genes showed a difference in expression pattern between susceptible and resistant *F. pennsylvanica* individuals, suggesting that the majority of these loci may not play a critical role in governing the degree to which this species can defend itself against EAB. Nevertheless, loci that show a response to EAB, but where expression patterns are not differentiated between individuals with contrasting levels of susceptibility, may form part of a suite of genes that are required for defence, with trees that succumb fully to the insect lacking further essential components. Alternatively, these differentially expressed genes may reflect other changes that impacted the seedlings during the eight week period between collection of the “pre” and “post” EAB-feeding samples<sup>25</sup>.



## References

1. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356–369 (2008).
2. Doerge, R. W. Multifactorial genetics: Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* **3**, 43 (2002).
3. Xu, S. *et al.* Genome-Wide Convergence during Evolution of Mangroves from Woody Plants. *Mol. Biol. Evol.* **34**, 1008–1015 (2017).
4. Castoe, T. A. *et al.* Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8986–8991 (2009).
5. Orlova-Bienkowskaja, M. J. & Volkovitsh, M. G. Are native ranges of the most destructive invasive pests well known? A case study of the native range of the emerald ash borer, *Agrilus planipennis* (Coleoptera: Buprestidae). *Biol. Invasions* **20**, 1275–1286 (2018).
6. Herms, D. A. & McCullough, D. G. Emerald ash borer invasion of North America: history, biology, ecology, impacts, and management. *Annu. Rev. Entomol.* **59**, 13–30 (2014).
7. Orlova-Bienkowskaja, M. J. Ashes in Europe are in danger: the invasive range of *Agrilus planipennis* in European Russia is expanding. *Biol. Invasions* **16**, 1345–1349 (2014).
8. Semizer-Cuming, D., Krutovsky, K. V., Baranchikov, Y. N., Kjær, E. D. & Williams, C. G. Saving the world's ash forests calls for international cooperation now. *Nature Ecology & Evolution* **3**, 141–144 (2019).
9. Koch, J. L., Carey, D. W., Mason, M. E., Poland, T. M. & Knight, K. S. Intraspecific variation in *Fraxinus pennsylvanica* responses to emerald ash borer (*Agrilus planipennis*). *New Forests* **46**, 995–1011 (2015).
10. Wallander, E. Systematics and floral evolution in *Fraxinus* (Oleaceae). *Belgische Dendrologie Belge* **2013**, 39–58 (2012).
11. Sollars, E. S. A. *et al.* Genome sequence and genetic diversity of European ash trees. *Nature* **541**, 212–216 (2017).
12. Cruz, F. *et al.* Genome sequence of the olive tree, *Olea europaea*. *Gigascience* **5**, 29 (2016).
13. Hellsten, U. *et al.* Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19478–

- 19482 (2013).
14. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641 (2012).
  15. Withers, J. & Dong, X. Post-translational regulation of plant immunity. *Curr. Opin. Plant Biol.* **38**, 124–132 (2017).
  16. Bernards, M. A. & Båstrup-Spohr, L. Phenylpropanoid Metabolism Induced by Wounding and Insect Herbivory. in *Induced Plant Resistance to Herbivory* (ed. Schaller, A.) 189–211 (Springer Netherlands, 2008).
  17. Stahl, E., Hilfiker, O. & Reymond, P. Plant-arthropod interactions: who is the winner? *Plant J.* **93**, 703–728 (2018).
  18. Abdulrazzak, N. *et al.* A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant meta-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth. *Plant Physiol.* **140**, 30–48 (2006).
  19. Rupasinghe, S., Baudry, J. & Schuler, M. A. Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. *Protein Eng.* **16**, 721–731 (2003).
  20. Dolan, W. L. & Chapple, C. Conservation and Divergence of Mediator Structure and Function: Insights from Plants. *Plant Cell Physiol.* **58**, 4–21 (2017).
  21. Bonawitz, N. D. *et al.* Disruption of Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant. *Nature* **509**, 376–380 (2014).
  22. Dolan, W. L. & Chapple, C. Transcriptome Analysis of Four *Arabidopsis thaliana* Mediator Tail Mutants Reveals Overlapping and Unique Functions in Gene Regulation. *G3* **8**, 3093–3108 (2018).
  23. Xu, Z. *et al.* Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol. Biol.* **55**, 343–367 (2004).
  24. Rigsby, C. M., Herms, D. A., Bonello, P. & Cipollini, D. Higher Activities of Defense-Associated Enzymes may Contribute to Greater Resistance of Manchurian Ash to Emerald Ash Borer Than A closely Related and Susceptible Congener. *J. Chem. Ecol.* **42**, 782–792 (2016).
  25. Lane, T. *et al.* The green ash transcriptome and identification of genes responding to abiotic and biotic stresses. *BMC Genomics* **17**, 702 (2016).
  26. Villari, C., Herms, D. A., Whitehill, J. G. A., Cipollini, D. & Bonello, P. Progress and gaps in understanding mechanisms of ash tree resistance to emerald ash borer, a

- model for wood-boring insects that kill angiosperms. *New Phytol.* **209**, 63–79 (2016).
27. Erb, M. & Reymond, P. Molecular Interactions Between Plants and Insect Herbivores. *Annu. Rev. Plant Biol.* **70**, 527–557 (2019).
  28. Huang, J., Zhu, C. & Li, X. SCFSNIPER4 controls the turnover of two redundant TRAF proteins in plant immunity. *Plant J.* **95**, 504–515 (2018).
  29. Hua, Z. & Vierstra, R. D. The cullin-RING ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* **62**, 299–334 (2011).
  30. Erb, M., Meldau, S. & Howe, G. A. Role of phytohormones in insect-specific plant reactions. *Trends Plant Sci.* **17**, 250–259 (2012).
  31. Berens, M. L., Berry, H. M., Mine, A., Argueso, C. T. & Tsuda, K. Evolution of Hormone Signaling Networks in Plant Defense. *Annu. Rev. Phytopathol.* **55**, 401–425 (2017).
  32. Lin, S.-H. *et al.* Mutation of the Arabidopsis NRT1.5 nitrate transporter causes defective root-to-shoot nitrate transport. *Plant Cell* **20**, 2514–2528 (2008).
  33. Huysmans, M., Lema A, S., Coll, N. S. & Nowack, M. K. Dying two deaths - programmed cell death regulation in development and disease. *Curr. Opin. Plant Biol.* **35**, 37–44 (2017).
  34. Bellin, D., Asai, S., Delledonne, M. & Yoshioka, H. Nitric oxide as a mediator for defense responses. *Mol. Plant. Microbe. Interact.* **26**, 271–277 (2013).
  35. Zebelo, S. A. & Maffei, M. E. Role of early signalling events in plant-insect interactions. *J. Exp. Bot.* **66**, 435–448 (2015).
  36. Seifi, H. S. & Shelp, B. J. Spermine Differentially Refines Plant Defense Responses Against Biotic and Abiotic Stresses. *Front. Plant Sci.* **10**, 117 (2019).
  37. Whitehill, J. G. A., Rigsby, C., Cipollini, D., Herms, D. A. & Bonello, P. Decreased emergence of emerald ash borer from ash treated with methyl jasmonate is associated with induction of general defense traits and the toxic phenolic compound verbascoside. *Oecologia* **176**, 1047–1059 (2014).
  38. Nelson, R., Wiesner-Hanks, T., Wisser, R. & Balint-Kurti, P. Navigating complexity to breed disease-resistant crops. *Nat. Rev. Genet.* **19**, 21–33 (2018).
  39. Radville, L., Chaves, A. & Preisser, E. L. Variation in plant defense against invasive herbivores: evidence for a hypersensitive response in eastern hemlocks (*Tsuga canadensis*). *J. Chem. Ecol.* **37**, 592–597 (2011).
  40. Hilker, M. & Fatouros, N. E. Resisting the onset of herbivore attack: plants perceive

- and respond to insect eggs. *Curr. Opin. Plant Biol.* **32**, 9–16 (2016).
41. Kim, C. Y., Bove, J. & Assmann, S. M. Overexpression of wound-responsive RNA-binding proteins induces leaf senescence and hypersensitive-like cell death. *New Phytol.* **180**, 57–70 (2008).
  42. Bollhöner, B. *et al.* The function of two type II metacaspases in woody tissues of *Populus* trees. *New Phytol.* **217**, 1551–1565 (2018).
  43. Altmann, S. *et al.* Transcriptomic basis for reinforcement of elm antiherbivore defence mediated by insect egg deposition. *Mol. Ecol.* **27**, 4901–4915 (2018).
  44. Duan, J. J., Larson, K., Watt, T., Gould, J. & Lelito, J. P. Effects of host plant and larval density on intraspecific competition in larvae of the emerald ash borer (Coleoptera: Buprestidae). *Environ. Entomol.* **42**, 1193–1200 (2013).
  45. Cappaert, D., McCullough, D. G., Poland, T. M. & Siegert, N. W. Emerald ash borer in North America: a research and regulatory challenge. *American Entomologist*. *51* (3): 152-165. **51**, (2005).
  46. Chamorro, M. L., Volkovitsh, M. G., Poland, T. M., Haack, R. A. & Lingafelter, S. W. Preimaginal stages of the emerald ash borer, *Agrilus planipennis* Fairmaire (Coleoptera: Buprestidae): an invasive pest on ash trees (*Fraxinus*). *PLoS One* **7**, e33185 (2012).
  47. Loureiro, J., Rodriguez, E., Doležel, J. & Santos, C. Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Ann. Bot.* **100**, 875–888 (2007).
  48. Doležel, J., Binarová, P. & Lucretti, S. Analysis of Nuclear DNA content in plant cells by flow cytometry. *Biol. Plant.* **31**, 113–120 (1989).
  49. Bennett Michael David & Smith J. B. Nuclear DNA amounts in angiosperms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **334**, 309–345 (1991).
  50. Whittemore, A. T. & Xia, Z.-L. Genome Size Variation in Elms (*Ulmus* spp.) and Related Genera. *HortScience* **52**, 547–553 (2017).
  51. Doležel, J. *et al.* Plant Genome Size Estimation by Flow Cytometry: Inter-laboratory Comparison. *Ann. Bot.* **82**, 17–26 (1998).
  52. Greilhuber, J. & Obermayer, R. Genome size and maturity group in *Glycine max* (soybean). *Heredity* **78**, 547–551 (1997).
  53. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11–15 (1987).
  54. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing

- reads. *EMBnet.journal* **17**, 10–12 (2011).
55. Joshi, N. A. & Fass, J. N. *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files*. (2011).
  56. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
  57. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011).
  58. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
  59. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
  60. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
  61. Keilwagen, J. *et al.* Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **44**, e89 (2016).
  62. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
  63. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
  64. Altenhoff, A. M., Gil, M., Gonnet, G. H. & Dessimoz, C. Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS One* **8**, e53786 (2013).
  65. Altenhoff, A. M. *et al.* The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. *Nucleic Acids Res.* **43**, D240–9 (2015).
  66. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–86 (2012).
  67. Wallander, E. Systematics of Fraxinus (Oleaceae) and evolution of dioecy. *Plant Syst. Evol.* **273**, 25–49 (2008).
  68. Hinsinger, D. D. *et al.* The phylogeny and biogeographic history of ashes (Fraxinus, Oleaceae) highlight the roles of migration and vicariance in the diversification of temperate trees. *PLoS One* **8**, e80431 (2013).
  69. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high

- throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
70. Sela, I., Ashkenazy, H., Katoh, K. & Pupko, T. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* **43**, W7–14 (2015).
  71. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
  72. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
  73. Ané, C., Larget, B., Baum, D. A., Smith, S. D. & Rokas, A. Bayesian estimation of concordance among gene trees. *Mol. Biol. Evol.* **24**, 412–426 (2007).
  74. Larget, B. R., Kotha, S. K., Dewey, C. N. & Ané, C. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* **26**, 2910–2911 (2010).
  75. Whitehill, J. G. A. *et al.* Interspecific proteomic comparisons reveal ash phloem genes potentially involved in constitutive resistance to the emerald ash borer. *PLoS One* **6**, e24863 (2011).
  76. Rigsby, C. M., McCartney, N. B., Herms, D. A., Tumlinson, J. H. & Cipollini, D. Variation in the Volatile Profiles of Black and Manchurian Ash in Relation to Emerald Ash Borer Oviposition Preferences. *J. Chem. Ecol.* **43**, 831–842 (2017).
  77. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
  78. Otto, T. D. *et al.* Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. *Nat Microbiol* **3**, 687–697 (2018).
  79. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
  80. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
  81. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
  82. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
  83. Martin, M. *et al.* WhatsHap: fast and accurate read-based phasing. *bioRxiv* 085050

- (2016). doi:10.1101/085050
84. Milne, I. *et al.* Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14**, 193–202 (2013).
  85. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
  86. Pond, S. L. K., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
  87. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. W. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* **23**, 1891–1901 (2006).
  88. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–42 (2013).
  89. Liu, Z. *et al.* Evolutionary interplay between sister cytochrome P450 genes shapes plasticity in plant metabolism. *Nat. Commun.* **7**, 13026 (2016).
  90. Altenhoff, A. M. *et al.* The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* **46**, D477–D485 (2018).
  91. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
  92. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).
  93. Alexa, A. & Rahnenführer, J. *topGO: enrichment analysis for gene ontology.* (2016).
  94. Team, R. C. & Others. R: A language and environment for statistical computing. (2013).
  95. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).
  96. Käll, L., Krogh, A. & Sonnhammer, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.* **35**, W429–32 (2007).
  97. Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).
  98. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures

- and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
99. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
  100. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
  101. Dallakyan, S. & Olson, A. J. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* **1263**, 243–250 (2015).
  102. Liu, Y. & Imai, R. Function of Plant DExD/H-Box RNA Helicases Associated with Ribosomal RNA Biogenesis. *Front. Plant Sci.* **9**, 125 (2018).
  103. Katori, T. *et al.* Dissecting the genetic control of natural variation in salt tolerance of *Arabidopsis thaliana* accessions. *J. Exp. Bot.* **61**, 1125–1138 (2010).
  104. Sharma, M. & Pandey, G. K. Expansion and Function of Repeat Domain Proteins During Stress and Development in Plants. *Front. Plant Sci.* **6**, 609 (2016).
  105. Godoy Herz, M. A. & Kornblihtt, A. R. Alternative Splicing and Transcription Elongation in Plants. *Front. Plant Sci.* **10**, 309 (2019).
  106. Dolan, W. L., Dilkes, B. P., Stout, J. M., Bonawitz, N. D. & Chapple, C. Mediator Complex Subunits MED2, MED5, MED16, and MED23 Genetically Interact in the Regulation of Phenylpropanoid Biosynthesis. *Plant Cell* **29**, 3269–3285 (2017).
  107. Wang, J. *et al.* Re-analysis of long non-coding RNAs and prediction of circRNAs reveal their novel roles in susceptible tomato following TYLCV infection. *BMC Plant Biol.* **18**, 104 (2018).
  108. García-Andrade, J., Ramírez, V., López, A. & Vera, P. Mediated plastid RNA editing in plant immunity. *PLoS Pathog.* **9**, e1003713 (2013).
  109. Ramírez, V. *et al.* Drought tolerance in *Arabidopsis* is controlled by the OCP3 disease resistance regulator. *Plant J.* **58**, 578–591 (2009).
  110. Collart, M. A., Panasenko, O. O. & Nikolaev, S. I. The Not3/5 subunit of the Ccr4-Not complex: a central regulator of gene expression that integrates signals between the cytoplasm and the nucleus in eukaryotic cells. *Cell. Signal.* **25**, 743–751 (2013).
  111. Teixeira, M. A. *et al.* Classification and phylogenetic analyses of the *Arabidopsis* and tomato G-type lectin receptor kinases. *BMC Genomics* **19**, 239 (2018).
  112. Vaid, N., Macovei, A. & Tuteja, N. Knights in action: lectin receptor-like kinases in plant development and stress responses. *Mol. Plant* **6**, 1405–1418 (2013).



113. Gilardoni, P. A., Hettenhausen, C., Baldwin, I. T. & Bonaventure, G. *Nicotiana attenuata* LECTIN RECEPTOR KINASE1 suppresses the insect-mediated inhibition of induced defense responses during *Manduca sexta* herbivory. *Plant Cell* **23**, 3512–3532 (2011).
114. Liu, Y. *et al.* A gene cluster encoding lectin receptor kinases confers broad-spectrum and durable insect resistance in rice. *Nat. Biotechnol.* **33**, 301–305 (2015).
115. Nguyen, C. T. *et al.* Characterization of the Chloride Channel-Like, AtCLCg, Involved in Chloride Tolerance in *Arabidopsis thaliana*. *Plant Cell Physiol.* **57**, 764–775 (2016).
116. Bryant, N., Lloyd, J., Sweeney, C., Myouga, F. & Meinke, D. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in *Arabidopsis*. *Plant Physiol.* **155**, 1678–1689 (2011).
117. Cao-Pham, A. H., Urano, D., Ross-Elliott, T. J. & Jones, A. M. Nudge-nudge, WNK-WNK (kinases), say no more? *New Phytol.* **220**, 35–48 (2018).
118. Heyndrickx, K. S. & Vandepoele, K. Systematic identification of functional plant modules through the integration of complementary data sources. *Plant Physiol.* **159**, 884–901 (2012).
119. Wu, J. *et al.* Genome-wide identification of MAPKK and MAPKKK gene families in tomato and transcriptional profiling analysis during development and stress response. *PLoS One* **9**, e103032 (2014).
120. O'Brien, J. A. *et al.* Nitrate Transport, Sensing, and Responses in Plants. *Mol. Plant* **9**, 837–856 (2016).
121. Corratgé-Faillie, C. & Lacombe, B. Substrate (un)specificity of *Arabidopsis* NRT1/PTR FAMILY (NPF) proteins. *J. Exp. Bot.* **68**, 3107–3113 (2017).
122. Yu, B. *et al.* *Arabidopsis* cpSRP54 regulates carotenoid accumulation in *Arabidopsis* and *Brassica napus*. *J. Exp. Bot.* **63**, 5189–5202 (2012).
123. Kai, K., Shimizu, B.-I., Mizutani, M., Watanabe, K. & Sakata, K. Accumulation of coumarins in *Arabidopsis thaliana*. *Phytochemistry* **67**, 379–386 (2006).
124. Cao, Y.-Y. *et al.* A Phylogenetically Informed Comparison of GH1 Hydrolases between *Arabidopsis* and Rice Response to Stressors. *Front. Plant Sci.* **8**, 350 (2017).
125. Liu, J. *et al.* MuMADS1 and MaOFP1 regulate fruit quality in a tomato ovate mutant. *Plant Biotechnol. J.* **16**, 989–1001 (2018).
126. Erb, M. Plant Defenses against Herbivory: Closing the Fitness Gap. *Trends Plant*

- Sci.* **23**, 187–194 (2018).
127. Granot, D., David-Schwartz, R. & Kelly, G. Hexose Kinases and Their Role in Sugar-Sensing and Plant Development. *Front. Plant Sci.* **4**, 44 (2013).
128. Kloth, K. J. *et al.* AtWRKY22 promotes susceptibility to aphids and modulates salicylic acid and jasmonic acid signalling. *J. Exp. Bot.* **67**, 3383–3396 (2016).
129. Joshi, C. P. & Chiang, V. L. Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. *Plant Mol. Biol.* **37**, 663–674 (1998).
130. Köllner, T. G. *et al.* Herbivore-induced SABATH methyltransferases of maize that methylate anthranilic acid using s-adenosyl-L-methionine. *Plant Physiol.* **153**, 1795–1807 (2010).
131. Guo, R. *et al.* Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape. *BMC Genomics* **14**, 554 (2013).
132. Xia, Y. *et al.* An extracellular aspartic protease functions in Arabidopsis disease resistance signaling. *EMBO J.* **23**, 980–988 (2004).
133. Riechmann, J. L. *et al.* Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* **290**, 2105–2110 (2000).
134. Guo, H. *et al.* Three related receptor-like kinases are required for optimal cell elongation in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7648–7653 (2009).
135. Lozano-Durán, R. & Zipfel, C. Trade-off between growth and immunity: role of brassinosteroids. *Trends Plant Sci.* **20**, 12–19 (2015).
136. Kulich, I. *et al.* Arabidopsis exocyst subunits SEC8 and EXO70A1 and exocyst interactor ROH1 are involved in the localized deposition of seed coat pectin. *New Phytol.* **188**, 615–625 (2010).
137. Stracke, R. *et al.* Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling: TFs of A. thaliana R2R3-MYB subgroup 7. *Plant J.* **50**, 660–677 (2007).
138. Zhao, Y., Dai, X., Blackwell, H. E., Schreiber, S. L. & Chory, J. SIR1, an upstream component in auxin signaling identified by chemical genetics. *Science* **301**, 1107–1110 (2003).
139. Huang, D., Wu, W., Abrams, S. R. & Cutler, A. J. The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors. *J. Exp. Bot.* **59**, 2991–3007 (2008).
140. Zouari, I. *et al.* From root to fruit: RNA-Seq analysis shows that arbuscular

- mycorrhizal symbiosis may affect tomato fruit metabolism. *BMC Genomics* **15**, 221 (2014).
141. Fan, M. *et al.* Overexpression of SIRBZ Results in Chlorosis and Dwarfism through Impairing Chlorophyll, Carotenoid, and Gibberellin Biosynthesis in Tomato. *Front. Plant Sci.* **7**, 907 (2016).
142. Masclaux-Daubresse, C. *et al.* Stitching together the Multiple Dimensions of Autophagy Using Metabolomics and Transcriptomics Reveals Impacts on Metabolism, Development, and Plant Responses to the Environment in Arabidopsis. *Plant Cell* **26**, 1857–1877 (2014).
143. Xu, J. *et al.* A Novel RNA-Binding Protein Involves ABA Signaling by Post-transcriptionally Repressing ABI2. *Front. Plant Sci.* **8**, 24 (2017).
144. Ge, C. *et al.* BUD2, encoding an S-adenosylmethionine decarboxylase, is required for Arabidopsis growth and development. *Cell Res.* **16**, 446–456 (2006).
145. Majumdar, R., Shao, L., Turlapati, S. A. & Minocha, S. C. Polyamines in the life of Arabidopsis: profiling the expression of S-adenosylmethionine decarboxylase (SAMDC) gene family during its life cycle. *BMC Plant Biol.* **17**, 264 (2017).
146. de Abreu-Neto, J. B., Turchetto-Zolet, A. C., de Oliveira, L. F. V., Zanettini, M. H. B. & Margis-Pinheiro, M. Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. *FEBS J.* **280**, 1604–1616 (2013).
147. Winter, T. R., Borkowski, L., Zeier, J. & Rostás, M. Heavy metal stress can prime for herbivore-induced plant volatile emission. *Plant Cell Environ.* **35**, 1287–1298 (2012).
148. Li, J. *et al.* Transcriptome analysis reveals a comprehensive insect resistance response mechanism in cotton to infestation by the phloem feeding insect Bemisia tabaci (whitefly). *Plant Biotechnol. J.* **14**, 1956–1975 (2016).
149. Escamez, S. *et al.* METACASPASE9 modulates autophagy to confine cell death to the target cells during Arabidopsis vascular xylem differentiation. *Biol. Open* **5**, 122–129 (2016).
150. Fortin, J. & Lam, E. Domain swap between two type-II metacaspases defines key elements for their biochemical properties. *Plant J.* **96**, 921–936 (2018).
151. Wang, G. *et al.* Genome-wide analysis of the cyclin family in Arabidopsis and comparative phylogenetic analysis of plant cyclin-like proteins. *Plant Physiol.* **135**, 1084–1099 (2004).

152. Miyawaki, K. *et al.* Roles of Arabidopsis ATP/ADP isopentenyltransferases and tRNA isopentenyltransferases in cytokinin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16598–16603 (2006).
153. Albrecht, T. & Argueso, C. T. Should I fight or should I grow now? The role of cytokinins in plant growth and immunity and in the growth-defence trade-off. *Ann. Bot.* **119**, 725–735 (2017).
154. Capron, A. *et al.* Identification of quantitative trait loci controlling fibre length and lignin content in Arabidopsis thaliana stems. *J. Exp. Bot.* **64**, 185–197 (2013).
155. Rogers, L. A. *et al.* Light, the circadian clock, and sugar perception in the control of lignin biosynthesis. *J. Exp. Bot.* **56**, 1651–1663 (2005).
156. Yonekura-Sakakibara, K. *et al.* Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *Plant Cell* **20**, 2160–2176 (2008).
157. Livne, S. *et al.* Uncovering DELLA-Independent Gibberellin Responses by Characterizing New Tomato procerca Mutants. *Plant Cell* **27**, 1579–1594 (2015).
158. Guan, L. *et al.* JASSY, a chloroplast outer membrane protein required for jasmonate biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10568–10575 (2019).
159. Zhao, J., Favero, D. S., Peng, H. & Neff, M. M. Arabidopsis thaliana AHL family modulates hypocotyl growth redundantly by interacting with each other via the PPC/DUF296 domain. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E4688–97 (2013).
160. Zhao, J., Favero, D. S., Qiu, J., Roalson, E. H. & Neff, M. M. Insights into the evolution and diversification of the AT-hook Motif Nuclear Localized gene family in land plants. *BMC Plant Biol.* **14**, 266 (2014).
161. Howden, A. J. M. *et al.* Quantitative analysis of the tomato nuclear proteome during *Phytophthora capsici* infection unveils regulators of immunity. *New Phytol.* **215**, 309–322 (2017).
162. González-Pérez, S. *et al.* Early transcriptional defense responses in Arabidopsis cell suspension culture under high-light conditions. *Plant Physiol.* **156**, 1439–1456 (2011).
163. Rezzonico, F., Rupp, O. & Fahrenttrapp, J. Pathogen recognition in compatible plant-microbe interactions. *Sci. Rep.* **7**, 6383 (2017).
164. McCormack, E. & Braam, J. Calmodulins and related potential calcium sensors of Arabidopsis. *New Phytol.* **159**, 585–598 (2003).
165. Wang, Y. *et al.* Transcriptome analyses show changes in gene expression to

- accompany pollen germination and tube growth in Arabidopsis. *Plant Physiol.* **148**, 1201–1211 (2008).
166. Weeda, S. *et al.* Arabidopsis transcriptome analysis reveals key roles of melatonin in plant defense systems. *PLoS One* **9**, e93462 (2014).
167. Lecourieux, D., Ranjeva, R. & Pugin, A. Calcium in plant defence-signalling pathways. *New Phytol.* **171**, 249–269 (2006).
168. Heyno, E., Alkan, N. & Fluhr, R. A dual role for plant quinone reductases in host-fungus interaction. *Physiol. Plant.* **149**, 340–353 (2013).
169. Showalter, A. M., Keppler, B. D., Liu, X., Lichtenberg, J. & Welch, L. R. Bioinformatic Identification and Analysis of Hydroxyproline-Rich Glycoproteins in *Populus trichocarpa*. *BMC Plant Biol.* **16**, 229 (2016).
170. Mei, Q. & Dvornyk, V. Evolutionary History of the Photolyase/Cryptochrome Superfamily in Eukaryotes. *PLoS One* **10**, e0135940 (2015).
171. Du, M. *et al.* MYC2 Orchestrates a Hierarchical Transcriptional Cascade That Regulates Jasmonate-Mediated Plant Immunity in Tomato. *Plant Cell* **29**, 1883–1906 (2017).
172. Omidvar, V. *et al.* Transcriptional regulation of male-sterility in 7B-1 male-sterile tomato mutant. *PLoS One* **12**, e0170715 (2017).
173. Li, Y. *et al.* Genome-wide identification of osmotic stress response gene in *Arabidopsis thaliana*. *Genomics* **92**, 488–493 (2008).
174. Mohr, P. G. & Cahill, D. M. Suppression by ABA of salicylic acid and lignin accumulation and the expression of multiple genes, in *Arabidopsis* infected with *Pseudomonas syringae* pv. tomato. *Funct. Integr. Genomics* **7**, 181–191 (2007).
175. Timm, S. *et al.* Mitochondrial Dihydrolipoyl Dehydrogenase Activity Shapes Photosynthesis and Photorespiration of *Arabidopsis thaliana*. *Plant Cell* **27**, 1968–1984 (2015).
176. Douce, R., Bourguignon, J., Neuburger, M. & Rébeillé, F. The glycine decarboxylase system: a fascinating complex. *Trends Plant Sci.* **6**, 167–176 (2001).
177. Palmieri, M. C., Lindermayr, C., Bauwe, H., Steinhauser, C. & Durner, J. Regulation of plant glycine decarboxylase by s-nitrosylation and glutathionylation. *Plant Physiol.* **152**, 1514–1528 (2010).
178. Sundell, D. *et al.* The Plant Genome Integrative Explorer Resource: PlantGenIE.org. *New Phytol.* **208**, 1149–1156 (2015).
179. Rajinikanth, M., Harding, S. A. & Tsai, C.-J. The glycine decarboxylase complex

- multienzyme family in *Populus*. *J. Exp. Bot.* **58**, 1761–1770 (2007).
180. Zhang, X. *et al.* Mediator subunit MED31 is required for radial patterning of *Arabidopsis* roots. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5624–E5633 (2018).
181. Li, P. *et al.* Spatial Expression and Functional Analysis of Casparian Strip Regulatory Genes in Endodermis Reveals the Conserved Mechanism in Tomato. *Front. Plant Sci.* **9**, 832 (2018).
182. Miguel, A., Milhinhos, A., Novák, O., Jones, B. & Miguel, C. M. The SHORT-ROOT-like gene PtSHR2B is involved in *Populus* phellogen activity. *J. Exp. Bot.* **67**, 1545–1555 (2016).
183. Wang, J. *et al.* Reduced expression of the SHORT-ROOT gene increases the rates of growth and development in hybrid poplar and *Arabidopsis*. *PLoS One* **6**, e28878 (2011).
184. Rasul, S. *et al.* Nitric oxide production mediates oligogalacturonide-triggered immunity and resistance to *Botrytis cinerea* in *Arabidopsis thaliana*. *Plant Cell Environ.* **35**, 1483–1499 (2012).
185. Delis, C. *et al.* AtHESPERIN: a novel regulator of circadian rhythms with poly(A)-degrading activity in plants. *RNA Biol.* **13**, 68–82 (2016).
186. Milner, M. J., Seamon, J., Craft, E. & Kochian, L. V. Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *J. Exp. Bot.* **64**, 369–381 (2013).
187. Socha, A. L. & Guerinot, M. L. Mn-euvering manganese: the role of transporter gene family members in manganese uptake and mobilization in plants. *Front. Plant Sci.* **5**, 106 (2014).
188. Ma, Y. *et al.* Ectomycorrhizas with *Paxillus involutus* enhance cadmium uptake and tolerance in *Populus × canescens*. *Plant Cell Environ.* **37**, 627–642 (2014).
189. Wang, Y. *et al.* Identification of tRNA nucleoside modification genes critical for stress response and development in rice and *Arabidopsis*. *BMC Plant Biol.* **17**, 261 (2017).
190. Chung, Y. & Choe, S. The Regulation of Brassinosteroid Biosynthesis in *Arabidopsis*. *CRC Crit. Rev. Plant Sci.* **32**, 396–410 (2013).
191. Tang, J., Han, Z. & Chai, J. Q&A: what are brassinosteroids and how do they act in plants? *BMC Biol.* **14**, 113 (2016).
192. Lee, J. H. *et al.* Brassinosteroids regulate glucosinolate biosynthesis in *Arabidopsis thaliana*. *Physiol. Plant.* (2018). doi:10.1111/ppl.12691

193. Inoue, S.-I. *et al.* Brassinosteroid Involvement in *Arabidopsis thaliana* Stomatal Opening. *Plant Cell Physiol.* **58**, 1048–1058 (2017).