Identification of the *Eucalyptus grandis* chitinase gene family and expression characterisation under different biotic stress challenges

Alternative title: Evidence for the importance of a single Class IA chitinase in response to fungal pathogen and gall wasp challenge in eucalypts

PERI A. TOBIAS*, NANETTE CHRISTIE, SANUSHKA NAIDOO, DAVID I. GUEST AND CARSTEN KÜLHEIM

*Author for correspondence: Peri A. Tobias (peri.tobias@sydney.edu.au) Sydney Institute of Agriculture, School of Life and Environmental Science, University of Sydney, Eveleigh, NSW 2015, Australia.

Nanette Christie (nanette.christie@gmail.com) Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute, University of Pretoria, Pretoria 0028, South Africa

Sanushka Naidoo (sanushka.naidoo@up.ac.za) Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute, University of Pretoria, Pretoria 0028, South Africa

David I. Guest (david.guest@sydney.edu.au) Sydney Institute of Agriculture, School of Life and Environmental Science, University of Sydney, Eveleigh, NSW 2015, Australia.

Carsten Külheim (carsten.kulheim@anu.edu.au) Research School of Biology, Australian National University, Canberra, ACT 2601, Australia

Key words: chitinase, *Eucalyptus grandis,* Class I chitinase, glycosyl hydrolase

Abstract

Eucalyptus grandis is an Australian Myrtaceae tree grown for timber in many parts of the world and for which the annotated genome sequence is available. Known to be susceptible to a number of pests and diseases, *E. grandis* is a useful study organism for investigating defence responses in woody plants. Chitinases are widespread in plants and cleave glycosidic bonds of chitin, the major structural component of fungal cell walls and arthropod exoskeletons. They are encoded by an important class of genes known to be up-regulated in plants in response to pathogens.

The current study identified sixty seven chitinase gene models from two families known as glycosyl hydrolase 18 and 19 (thirty-six GH18 and thirty-one GH19) within the *E. grandis* genome assembly (v1.1) indicating a recent gene expansion. Sequences were aligned and analyzed as conforming to currently recognized plant chitinase classes (I-V). Unlike other woody species investigated to date, *E. grandis* has a single gene encoding a putative vacuolar targeted Class I chitinase. In response to *Leptocybe invasa* (the eucalypt gall wasp) and *Chrysoporthe austroafricana* (causal agent of fungal stem canker), this *Class IA chitinase* is strongly up-regulated in both resistant and susceptible plants. Resistant plants however indicate greater constitutive expression and increased up-regulation than susceptible plants following fungal challenge. Up-regulation within fungal resistant clones was further confirmed with protein data. Clusters of putative chitinase genes, particularly on chromosome three and eight, are significantly up-regulated in response to fungal challenge, while a cluster on chromosome one is significantly down-regulated in response to gall wasp.

The results of this study show that the *E. grandis* genome has an expanded group of chitinase genes, compared with other plants. Despite this expansion, only a single *Class I chitinase*

is present and this gene is highly up-regulated within diverse biotic stress conditions. Our research provides insight into a major class of defence genes within *E. grandis* and indicates the importance of the *Class I chitinase*.

Background

The commercially important forestry species *Eucalyptus grandis* W. Hill ex Maiden (Flooded Gum, Rose Gum) originated in Australia and is now widely grown in plantations around the world (Boland et al. 2006). Eucalypts provide pulp and hardwood and are valued for their rapid growth and adaptability to a range of conditions (Myburg et al. 2014). As a listed biomass energy crop (Genomics Science Program Systems Biology for Bioenergy) and valuable forestry species, *E. grandis* was selected as one of the first woody plants for genome sequencing. The genome for *E. grandis*, from a 17 year old inbred tree clone, BRASUZ1 (genome size of 640Mbp, 11 haploid chromosomes), was published in 2014 (Myburg et al., 2014). Despite the great adaptability of eucalypts they are known to be susceptible to a range of pests and diseases (Whyte et al. 2011). The availability of genomic sequence data for *E. grandis* permits investigation into defence responses following biotic challenge.

The gall wasp, *Leptocybe invasa* Fisher & La Salle (Hymenoptera: Eulophidae), is a devastating pest that induces gall formation, stunting and dieback in *Eucalyptus* plantations (Nyeko et al. 2009, 2010). Resistant genotypes display evidence of oviposition holes on the midrib, shoot tip and petiole of young leaves without further gall development; however in susceptible genotypes, the larvae develop within protective galls which coalesce and, in extreme cases, cause loss of apical dominance of the tree (Dittrich-Schröder et al. 2012). In previous work (Oates et al. 2015), transcriptomic responses and changes in terpene profiles were investigated in a resistant and susceptible *E. grandis* clone seven days post oviposition. The significance of this time point

was that gall development had not yet initiated in the susceptible clone and thus facilitated the investigation of early transcriptional reprogramming induced by insect oviposition. Although galling insects are notorious manipulators of host defences (Tooker et al. 2008), in the absence of transcriptomic data of Eucalyptus – chewing insect interactions, the Eucalyptus- *Leptocybe invasa* interaction serves as a starting point to model Eucalyptus-insect interactions. *Chrysoporthe austroafricana*, a fungal pathogen, is able to cause stem canker on mature trees by entering sites of wounding, reducing wood quality and impeding growth. In artificial inoculation experiments, with *C. austroafricana*, the lesion lengths in resistant and susceptible year-old *E. grandis* clones were similar three days post inoculation (dpi) and the pathogen showed similar localization in xylem tissue at the microscopic level (Mangwanda et al. 2016). At seven dpi, the lesions are significantly different in the two genotypes. Hormone profiling of inoculated clones at these time points suggested that reduction in salicylic acid and gibberellic acid levels at three dpi in the resistant genotype was important for the resistant outcome. Transcriptome profiling at three days post inoculation supported the notion that hormone signaling pathways may contribute to resistance and the expression pattern of a selected number of genes further supported the hypothesis that a delayed defence response occurs in the susceptible interaction (Mangwanda et al. 2015). Determining the genetic basis for resistance is a research priority.

Pathogenesis-related (PR) proteins are categorized by their up-regulation in plants under pathogen and insect attack (Edreva 2005, Zhao et al. 2015). Numbered in the order of description and classified by amino acid sequence and enzymatic activity (van Loon and van Strien 1999), PR-proteins are important markers for systemic acquired resistance (SAR), whereby plants exhibit rapid and effective response to a broad range of pests and pathogens following an initial single organism exposure (Fu and Dong 2013). PR-3, 4, 8 and 11 proteins are identified as having

chitinase activity and include chitinases of type I, II, IV, V, VI, VII (PR-3), type I and II (PR-4), type III (PR-8), and type I, V (PR-11) (Edreva 2005). Chitinases (EC3.2.1.14) are known to actively interfere with or degrade chitin, a major structural component of fungal cell walls and arthropod exoskeletons, and are therefore good targets for defence response studies (Edreva 2005, Grover 2012). Most identified plant chitinases are endolytic, breaking up chitin polymers by acting on β-1,4 links of *N*-acetyl-D-glucosamine within the polymer chain. Chitinases in *Arabidopsis thaliana* have been identified as conforming to several isoforms with active sites (amino acid sequence, conformation and biochemical activity) defining the two major glycosyl hydrolase families (GH18 and GH19) and domain structure determining the five main classes (Class I – V) (Passarinho and DeVries 2002). Important domain components in identification and enzyme activity of chitinases are the chitin-binding and glycosyl hydrolase domains. GH19 chitinases have only been located in plants while GH18 are known to also occur within animals, fungi and bacteria (Kasprzewska 2003). A study investigating substrate specificities in tobacco chitinases showed that GH19 enzymes acted rapidly on long polymer chains while GH18 enzymes preferentially acted on chitin oligomers in a more sustained manner (Brunner et al. 1998) suggesting that the activity of complementary enzymes is required for effective defence.

There has, to date, been no scrutiny of the *E. grandis* genome to identify and characterize chitinases, though a review of PR proteins indicated that chitinase genes are more abundant in this species than in *Populus trichocarpa* and *Arabidopsis thaliana* (Naidoo et al. 2014). Previous studies have identified that *E. grandis* inoculated with the myrtle rust pathogen, *Puccinia psidii*, induces an increase in chitinase enzymes generally (Boava et al. 2009), and that a homologue for *Class I chitinase* (GH19) was up-regulated in resistant plants (Moon et al. 2007). Four putative basic chitinases were also identified as differentially expressed in resistant and susceptible *E.*

5

grandis following challenge with *Chrysoporthe austroafricana* (fungal stem canker) (Mangwanda et al. 2015). While there is evidence for up-regulation of chitinases in response to insect attack (Krishnaveni et al. 1999, Zhang et al. 2012), there is also evidence that mechanical wounding can induce even greater PR expression, particularly in susceptible plants (Reymond et al. 2000). It is suggested that targeted suppression of defence responses may occur in plants susceptible to insect damage (Reymond et al. 2000), while gall insects have been shown to systemically alter plant defences, such as chitinases and volatile compounds (Tooker et al. 2008, Oates et al. 2015). Nevertheless chitinases are well known for their accumulation in response to insect stressors and, though the functional basis is unclear, there have been correlations between the types of chitinases produced and resistance to infestation (Krishnaveni et al. 1999). It is therefore of great interest to determine the nature of resistance in challenged plants. Given the importance of chitinases in plant defence, the study of this family of genes is an important aspect of determining pest and pathogen responses in *E. grandis*.

The objective of the current research was to identify these genes within the *E. grandis* genome and classify them based on amino acid sequences of conserved domains to permit useful downstream investigations into tree responses. Our research therefore had two approaches; the identification and classification of the *chitinase* gene family and, the interrogation, and validation, of gene expression data pertaining to two well-designed biotic stress trials. We review gene expression based on previously published studies (Mangwanda et al. 2015, Oates et al. 2015) in clones moderately resistant, hereafter referred to as resistant, and susceptible to *Chrysoporthe austroafricana* and in clones resistant and susceptible to *Leptocybe invasa*, the eucalyptus gall wasp, and address the role of chitinases in pest and disease resistance.

Results
Identification of putative chitinase genes

Our search for chitinase genes identified 67 putative genes within two major glycosyl hydrolase families, GH18 and GH19. Within the five recognized plant chitinase classes we identified a single Class I, 16 Class II, 10 Class III, 14 Class IV and 26 Class V sequences. Figure 1 visualizes the locations of putative genes, from GH18 and GH19 families and Classes I -V, on the 11 *E. grandis* chromosomes and indicates that the classes are generally clustered on different chromosomes. In particular, all Class V genes are located on chromosome three and most Class III genes are located on chromosomes five and eleven. This phenomenon is also observed in *Populus trichocarpa* chitinases (Jiang et al. 2013).

Alignment and phylogenetic analysis

Sequences within the two GH families are highly conserved according to their evolutionary relatedness (Figure 2A and B). Within the phylogenetic analysis, both the GH19 and GH18 sequences separated out to the defined classes of chitinases (GH19 to Class IV and II, GH18 to Class III and V). The single Class I sequence aligned with the single *Arabidopsis thaliana* Class I sequence within the Class II group as expected, due to the full length GH19 domains known for these classes (Collinge et al. 1993).

The phylogenetic clustering of putative chitinase genes supports recently evolved and syntenic duplications on chromosomes, in particular for Eucgr.H00321 – H00328. Some unassigned putative genes (gene identifiers represented by Eucgr.L) were closely aligned with chromosome assigned genes, indicating potential allelic variants probably located within clusters, for example Eucgr.L00937 – L00938, within the sub-branch of Class II (cluster J-1) on chromosome ten (Figure 2B and S1 Table). Other examples were Eucgr.L00615 and
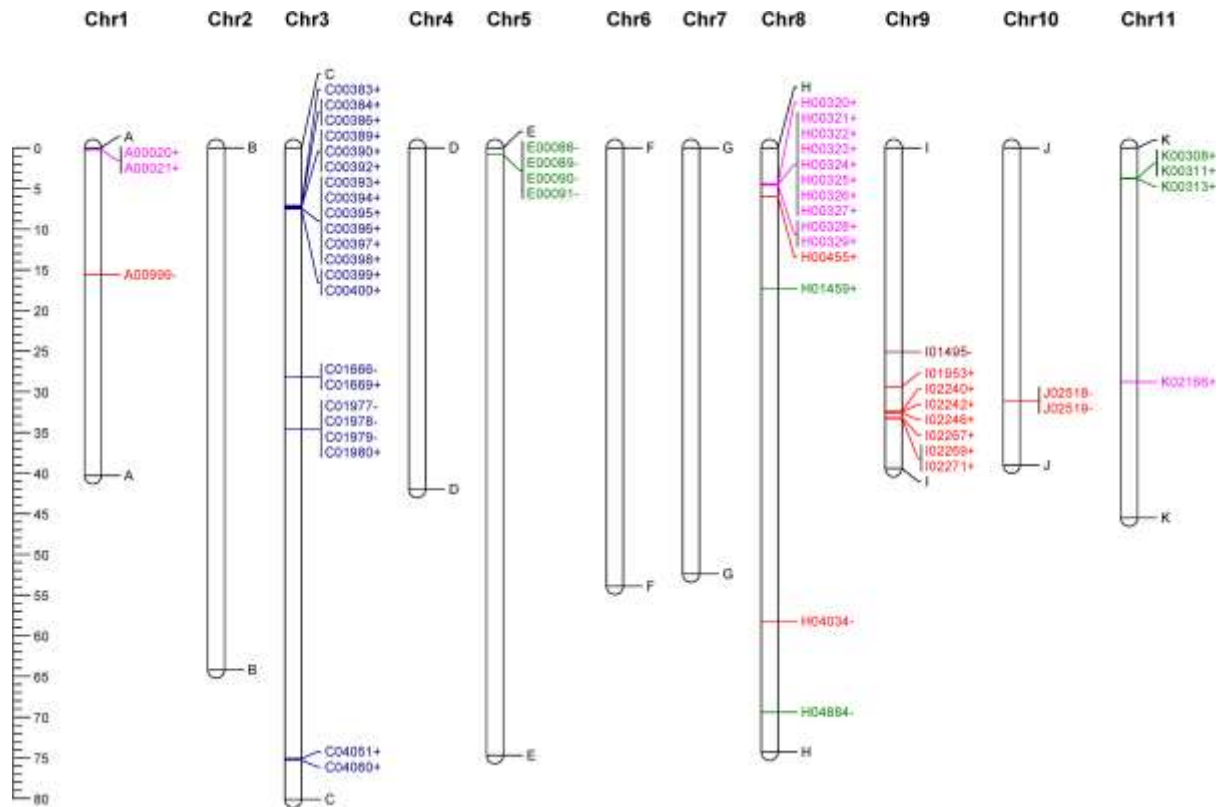
Fig 1. Physical mapping of all identified *Eucalyptus grandis* putative chitinases from the two glycosyl hydrolase families; GH18 (blue/green), and GH19 (red/pink/brown). All *E. grandis* gene ID's labelled minus the prefix Eucgr. Locations are mapped according to base pair start start positions with +/- indicating strand. The major classes are represented as follows; brown = Class I, red = Class II, green = Class III, pink = Class IV and blue = Class V. Scale bar represents Mb.
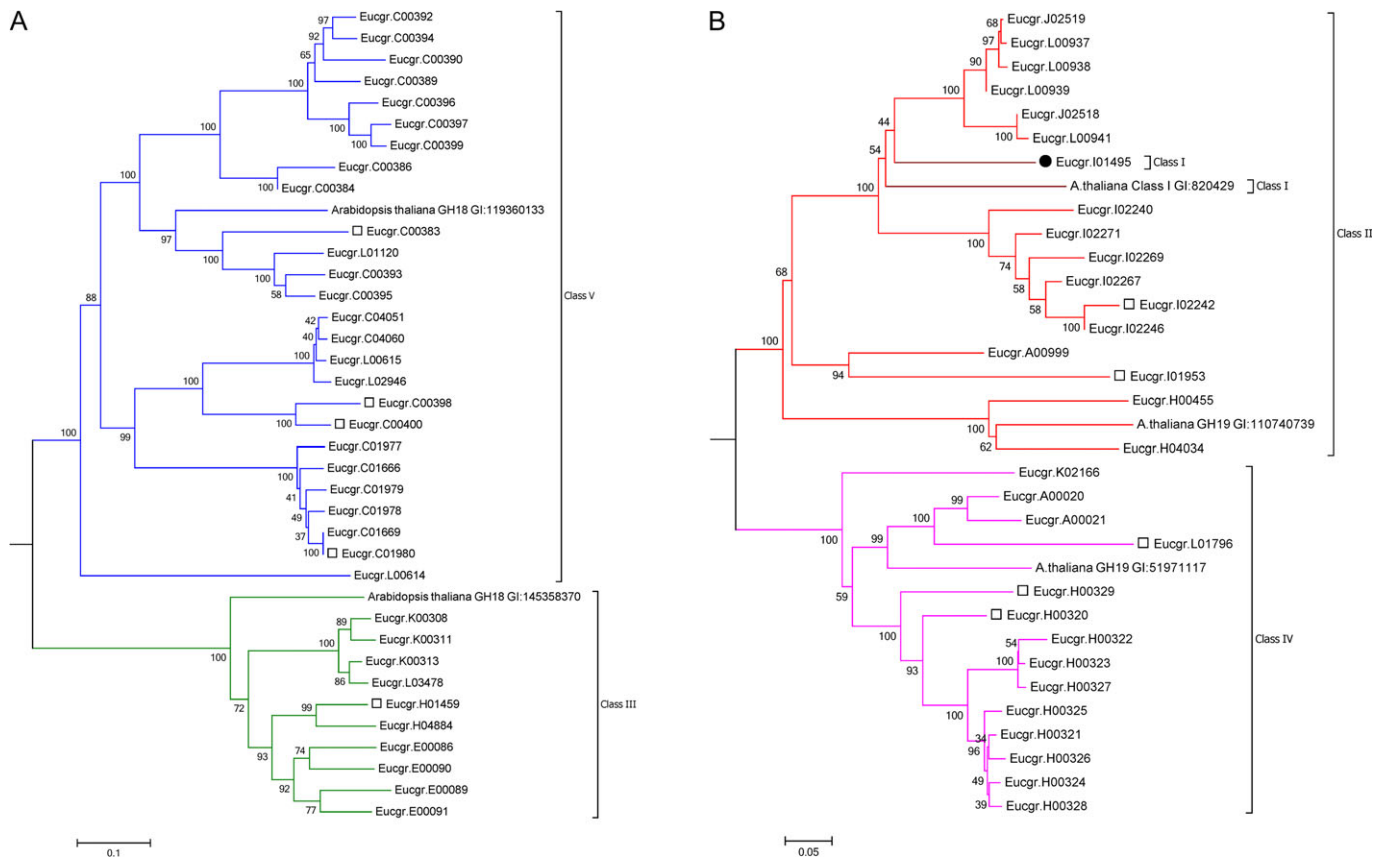
Figure 2. Evolutionary relationships of *Eucalyptus grandis* putative chitinase genes from two glycosyl hydrolase families: GH18 (A) Class III (green) and V (blue), and GH19 (B) Class IV (pink), Class II (red), Class I (brown) chitinase gene models. Eucgr.I01495 (black dot) was the only Class I chitinase. Putative genes that had no expression profiles are noted with square. The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. Scale: 0.05 or 0.1 substitutions per site. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

Eucgr.L02946 aligning with Class V genes (cluster C-4) on chromosome three and Eucgr.L03478 within Class III genes (cluster K-1) on chromosome eleven (Figure 2A and S1 Table).

Conserved domains and classification

Eucalyptus grandis chitinase sequences have conserved domain structures in accordance with previously described plant chitinases (Collinge et al. 1993, Passarinho and DeVries 2002). The alignment and MEME analysis of GH18 and GH19 sequences enabled the diagrammatic representation of these domains and facilitated classification of genes in accordance with the recognized classes of plant chitinases (Figure 3 and 4, S1 Fig.). A point of difference identified within *E. grandis* was the absence of a proline/threonine-rich hinge region in the single Class I chitinase (Eucgr.I01495). A glycine-rich (G) hinge is however evident as was also determined in *Populus trichocarpa* (Davis *et al.*, 1991).

Sequences homologous to the active sites of chitinases (Bishop et al. 2000) were identified within *E. grandis* sequences (Figure 3 and 4, boxed in red) and within the protein model for the Class I chitinase (Eugr.I01495) (S3 Fig.). Furthermore, homology to proline/threonine-rich hinge regions (Collinge et al. 1993), as well as N-terminal signal sequences, believed to target the gene products to the apoplast (Passarinho and DeVries 2002), were present in Class IV *E. grandis* sequences (Figure 3 and 4). The Eucgr.L00941 (Class II) sequence has a deletion from residues 160-210 which it shares with Eucgr.H00322 (Class IV). The deletion includes three identified active sites indicating that chitinase catalytic activity may be reduced or may not be the primary function of these proteins. Apparent insertions within the GH19 domain of Eucgr.I01495, Eucgr.J02518/9 and Eucgr.L00937/39/41 are in fact deletions in the other sequences, determined by referencing the domain structures of chitinase classes.
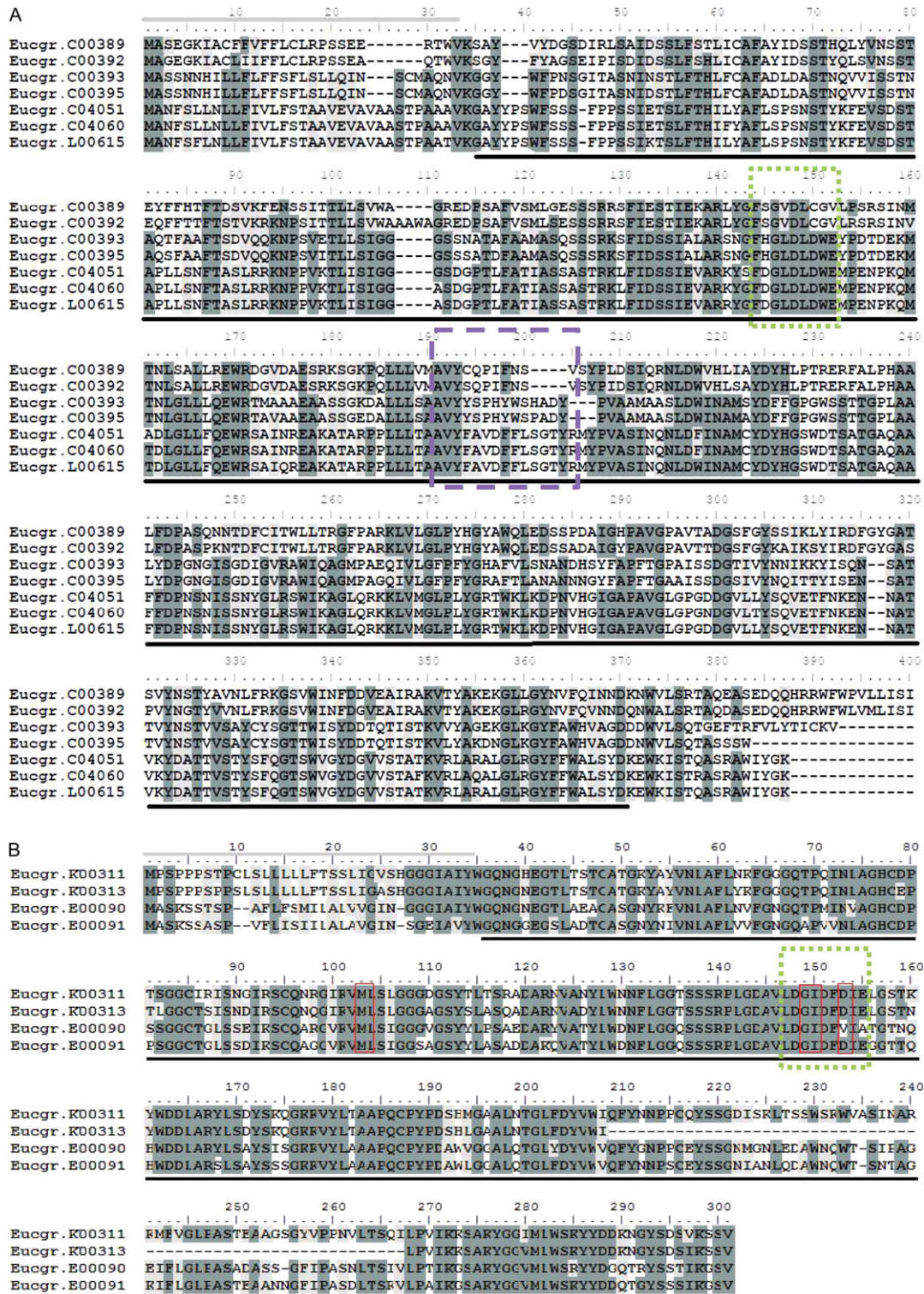
Figure 3. Conserved aligned sequences. Alignment, using ClustalW multiple alignment in BioEdit, of significantly differentially expressed *Eucalyptus grandis* GH18 Class V (A) and Class III (B) peptides, following treatment with *Chrysoporthe austroafricana* and *Leptocybe invasa*. Shaded amino acids are 70–100% homologous. Underlined regions indicate domain homology for glycosyl hydrolase family 18. Class V: Eucgr.C00389 and Eucgr. C00392 sequence lengths extend to 780 amino acids and are reduced for this alignment. Gray line over sequence indicates signal sequence. Red boxed residues are essential for catalytic activity. Purple box (dash line) = CRYSTALLYN_BETAGAMMA signature PS00225 ([LIVMFYWA]-{DEHRKSTP}-[FY]-[DEQHKY]-x(3)-[FY]-x-G-x(4)-[LIVMFC-ST]) and green box (dotted line) = Chitinase_18 signature PS01095 ([LIVMFY]-[DN]-G-[LIVMF]- [DN]-[LIVMF]-[DN]-x-E) (Passarinho and DeVries 2002).

Figure 4. Alignment using ClustalW multiple alignment in BioEdit, of significantly differentially expressed *Eucalyptus grandis* putative GH19 sequences following treatment with *Chrysoporthe austroafricana* and *Leptocybe invasa*. Shaded amino acids are 70–100% homologous. Underlined regions indicate domain homology for chitin-binding domain (amino acid residues 35–63) and GH19 (residues 89–320). Red lines over sequences indicate active sites as defined by residues within 0.6 nm of bound substrates (Bishop et al. 2000). Orange box indicates C-terminal extension for vacuolar targeting. Light gray line over sequence indicates signal sequence and dark line indicates (residues 65–80) proline/glycine-rich hinge region. Red boxed residues are essential for catalytic activity or enzyme function in Class I chitinases (Bishop et al. 2000). Eucgr.J02518, Eucgr.J02519, Eucgr.L00937, Eucgr. L00939 and Eucgr.L00941 do not have the chitin-binding domain, indicating that they are Class II. Two apparent insertions that the Class II genes share with Eucgr.I01495 from residues 163–178 and residues 245–265 are actually deletions in the other sequences, therefore determining these peptides as Class II and Class I, respectively. Blue box 1 (dot-dash line) = Chitinase 19_1 signature PS00773 (C-x(4,5)-F-Y-[ST]-x(3)-[FY]-[LIVMF]-x-A-x(3)-[YF]-x(2)-F-[GSA]) and blue box 2 (dot-dash line) = Chitinase 19_2 signature PS00774 ([LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W-[LIVM]) (Passarinho and DeVries 2002).

12

Class I and IV *E. grandis* sequences contain chitin binding and GH19 catalytic domains while Class II contains only the GH19 domain. Class III and V contain GH18 catalytic domains while Class V can also contain chitin binding domains (Passarinho and de Vries 2002), though not seen in *E. grandis*. *Eucalyptus grandis* Class IV chitinases contain deletions within the chitin binding and GH19 domains making the sequence shorter, as previously identified for plant chitinases (Grover 2012). Eucgr.H00329 and Eucgr.H00320 both contained GH19 domains with deletions matching Class IV chitinases, however they did not contain chitin-binding domains expected with this class. The characteristic C-terminal vacuolar targeting sequence was evident in the single *E. grandis* Class I chitinase (Eucgr.I01495), thereby determining it as Class IA (black dot, Figure 2b) (Nakamura and Matsuoka 1993, Kasprzewska 2003). The C-terminal extension sequence 'GLLVDTM' from amino acid residues 331-337 (Figure 4) matches the sequence determined in *Casuarina equisetfolia* (Veluthakkal and Dasgupta 2012) but varies slightly from the *A. thaliana* and *Arabis parishii* C-terminal extensions 'GLLEAAI' and 'GLLGAAI' respectively (Bishop et al. 2000, Passarinho and DeVries 2002).

Other domains

Nine putative Class V sequences, all located on chromosome three, incorporated predicted protein kinase domains, with a subset of these (Eucgr.C00396, Eucgr.C00386 and Eucgr.C00384) having homology with aminoglycoside phosphostransferase (APH) domains (PF01636), noted for antibiotic resistance to aminoglycosides. One of the nine sequences, Eucgr.C00386 incorporated a predicted transmembrane motif. Six other Class V sequences (Eucgr.C01666, Eucgr.C01669, Eucgr.C01977, Eucgr.C01978, Eucgr.C01979, and Eucgr.C01980) included both GH18 and GH85 domains, with four of these having predicted transmembrane motifs. One Class III peptide (Eucgr.K00313) had a predicted central domain for Cyclin-dependent kinase inhibitor 3. Of the

GH19 peptides, additional domain predictions included: Legume Lectin domain (Eucgr.I02271);

Ornatin domain (glycoprotein antagonist) (Eucgr.L00939) and Rifampin ADP-ribosyl transferase

(Eucgr.A00021 and Eucgr.H00320) (S1 Table).

Physical cluster and supercluster analysis

We identified eleven clusters and one supercluster of chitinase genes within the *Eucalyptus grandis*

genome assembly (Table 1). Only eight putative chitinases were singletons and ten were not

chromosome-assigned (Eucgr.L gene identifiers). A region of *E. grandis* chromosome eight

(Figure 5A) shows the clustering of ten predicted *Class IV chitinase*s (Eucgr.H00320 –

Eucgr.H00329) within 130 kb. All of these genes have high sequence homology. A chromosome

nine genomic region (Figure 5B) had three *Class II chitinases* (Eucgr.I00240, Eucgr.I02242,

Eucgr.I02246) alongside three predicted protein-kinase leucine-rich repeat genes (Eucgr.I02248,

Eucgr.I002251, Eucgr.I02256), known to be important in plant defence.

Predicted chitinase transmembrane regions and cleavage sites

Trans-membrane (TM) regions were predicted for twenty-four chitinase sequences including; 1

Class II, 7 Class III, 7 Class IV and 9 Class V (Table 1). Depending on the location of the predicted

TM motif the sequence presents a large intra- or extracellular peptide, perhaps suggesting a role

in chitin perception (Kaku et al. 2006). Predicted N-terminal cleavage sites, indicating secretion

of mature proteins, was identified for 26 of 36 of the GH18 peptides and 27 out of 31 GH19

peptides. Predicted localization for mature proteins indicated that 53 putative chitinases are

secreted (S1 Table).

Table 1. Putative (A) GH18 chitinase genes and (B) GH19 chitinase genes expressed following biotic challenge; insect (*Leptocybe invasa, L.i.*) and fun-gal (*Chrysoporthe austroafricana, C.a.*) in resistant (R) and susceptible (S) *Eucalyptus grandis* clones; fungal (*Calonectria pseudoreteaudii, C.p.*) resistant *Eucalyptus tereticornis x E. urophylla.*

Gene ID	Class[a]	TM[b]	Cluster[c]	L.i. R	L.i. S	C.a R	C.a S	C.p. R
(A) Glycoside hydrolase family 18								
Eucgr.C00383	V		C-1	noExp	noExp	noExp	noExp	
Eucgr.C00384	V		C-1					
Eucgr.C00386	V	Intra	C-1			down**		up***
Eucgr.C00389	V		C-1					
Eucgr.C00390	V		C-1					
Eucgr.C00392	V		C-1					
Eucgr.C00393	V		C-1			up***	up***	up***
Eucgr.C00394	V		C-1					up***
Eucgr.C00395	V		C-1	up***	up***	up***	up***	up***
Eucgr.C00396	V		C-1					
Eucgr.C00397	V		C-1	up***				
Eucgr.C00398	V		C-1					
Eucgr.C00399	V		C-1		up**			
Eucgr.C00400	V		C-1	noExp	noExp	noExp	noExp	
Eucgr.C01666	V		C-2					
Eucgr.C01669	V	Extra	C-2					
Eucgr.C01977	V		C-3					
Eucgr.C01978	V	Extra	C-3			up*		
Eucgr.C01979	V	Extra	C-3					
Eucgr.C01980	V	Extra	C-3	noExp	noExp	noExp	noExp	
Eucgr.C04051	V	Extra	C-4		down**	up***	up***	up***
Eucgr.C04060	V	Extra	C-4			up***	up***	up***
Eucgr.E00086	III	Extra	E-1					
Eucgr.E00089	III		E-1					
Eucgr.E00090	III	Intra	E-1				up**	
Eucgr.E00091	III	Intra	E-1			up*		
Eucgr.H01459	III			noExp	noExp	noExp	noExp	
Eucgr.H04884	III	Extra						
Eucgr.K00308	III		K-1					
Eucgr.K00311	III	Extra	K-1					up***
Eucgr.K00313	III	Intra	K-1					
Eucgr.L00614	V							
Eucgr.L00615	V	Extra			down*	up***	up***	up***
Eucgr.L01120	V			up**	up**	up***	up***	
Eucgr.L02946	V	Extra				up***	up***	
Eucgr.L03478	III	Extra				up**		
(B) Glycoside hydrolase family 19								
Eucgr.A00020	IV		A-1	down***	down***			
Eucgr.A00021	IV		A-1	down**	down**	up***		up***
Eucgr.A00999	II							
Eucgr.H00320	IV		H-1	noExp	noExp	noExp	noExp	
Eucgr.H00321	IV		H-1	up**	up**	up***	up***	up***
Eucgr.H00322	IV	Extra	H-1					
Eucgr.H00323	IV	Extra	H-1	noExp	noExp			up***
Eucgr.H00324	IV	Extra	H-1			up**	up**	
Eucgr.H00325	IV	Extra	H-1				up**	up***
Eucgr.H00326	IV	Extra	H-1			up***	up***	up***
Eucgr.H00327	IV	Extra	H-1					
Eucgr.H00328	IV	Extra	H-1			up**		
Eucgr.H00329	IV		H-1	noExp	noExp	noExp	noExp	
Eucgr.H00455	II	Extra						
Eucgr.H04034	II				down*			

Gene ID	Class[a]	TM[b]	Cluster[c]	L.i.		C.a		C.p.
				R	S	R	S	R
Eucgr.I01495	IA			up***	up***	up***	up***	up***
Eucgr.I01953	II			noExp	noExp	noExp	noExp	
Eucgr.I02240	II		I-1	noExp	noExp	noExp	noExp	down***
Eucgr.I02242	II		I-1	noExp	noExp	noExp	noExp	
Eucgr.I02246	II		I-1					
Eucgr.I02267	II		I-2					down***
Eucgr.I02269	II		I-2					
Eucgr.I02271	II		I-2	down*	down*			down***
Eucgr.J02518	II		J-1					
Eucgr.J02519	II		J-1		up**	up***	up***	up***
Eucgr.K02166	IV			down***	down***			down***
Eucgr.L00937	II			up***	up***	up***		up***
Eucgr.L00938	II							
Eucgr.L00939	II					up***	up***	up***
Eucgr.L00941	II					up***	up***	up***
Eucgr.L01796	IV			noExp	noExp	noExp	noExp	

Significance by treatment identified (inoculated vs control; *$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$) as well as $\log_2$ gene expression ratios greater than 1 (up) or smaller than −1 (down). noExp indicates no expression data.

[a]Chitinase class (I–V).

[b]Predicted TM region: Intra = predicted large intracellular region; Extra = predicted large extracellular region.

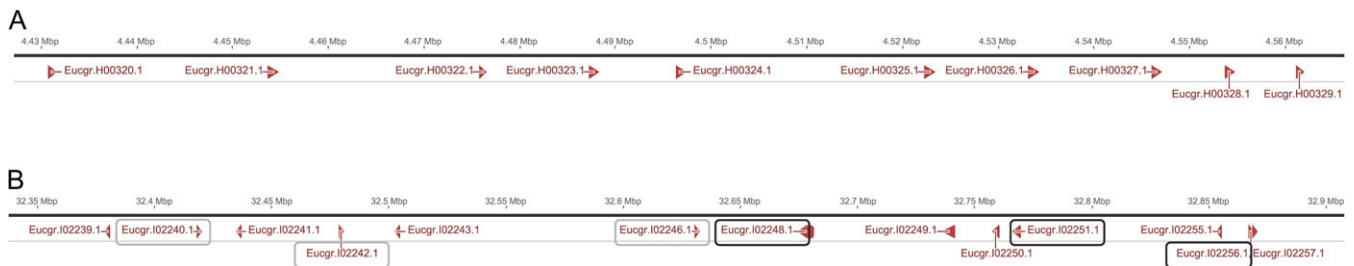[c]Physical cluster in *E. grandis:* 'scaffold' − 'cluster number on that scaffold'.

Figure 5. Regions of *Eucalyptus grandis* chromosome 8 (A) indicating the clustering of 10 putative chitinase genes (Eucgr.H00320–Eucgr.H00329) and chromosome 9 (B) (Eucgr.I02240, Eucgr.I02242 and Eucgr.I02246 = gray boxes). Eucgr.I02248, Eucgr.I002251 and Eucgr.I02256 (black boxes) are predicted protein-kinase LRR genes, important in plant defense. Triangles represent mRNA. Images extracted from GeneiousV6.1 (Kearse et al. 2012).

Chitinase expression in response to fungal pathogen and insect pest

Broad gene expression analysis identified 26 expressed GH19 genes of the 31 gene models and 31 expressed GH18 genes of the 36 gene models. Notably the single *Class IA chitinase* (Eucgr.I01495) constitutive expression is high in the fungal (*C. austroafricana*) treatment model, compared to other chitinases, and is also significantly up-regulated across both resistant and susceptible plants (Figure 6A). For example, in resistant control plants the mean normalized $\log_2$ of FPKM values was 107.7 (98.3 in susceptible) while in the inoculated plants the value was 409.1 (287.1 in susceptible). As a comparison, the mean normalized $\log_2$ of FPKM values for all other chitinase genes (excluding Eucgr.I01495) in resistant control plants were 8.9 while in the inoculated plants the values were 14.5. While the constitutive expression of *Class IA chitinase* within the insect (*L. invasa*) treatment model was not high, the up-regulation in comparison with other chitinases was significant in resistant plants (*P*-value < 0.01) (Figure 6B).

Other interesting expression changes include a single putative *Class IV chitinase* (Eucgr.A00021) significantly up-regulated only within the resistant plants in response to *C. austroafricana* (Table 1B, and Figure 7C). The same transcript was significantly down regulated in response to *L. invasa* in both resistant and susceptible plants (Figure 7D). A cluster of putative *Class IV chitinases* on chromosome eight were differentially expressed in resistant and susceptible plants in response to *C. austroafricana* but no significant expression was present with *L. invasa* challenge (Table 1B). Putative *Class V chitinases* within cluster C-1 were significantly up-regulated in both treatments (fungal and insect). Specifically Eucgr.C00395 was up-regulated across resistant and susceptible plants while the insect resistant plants also had high expression of Eucgr.C00397. Up-regulation for *C. austroafricana* challenge was determined for two putative *Class V chitinases* (Eucgr.C04051 and Eucgr.L00615) while these same transcripts were down-
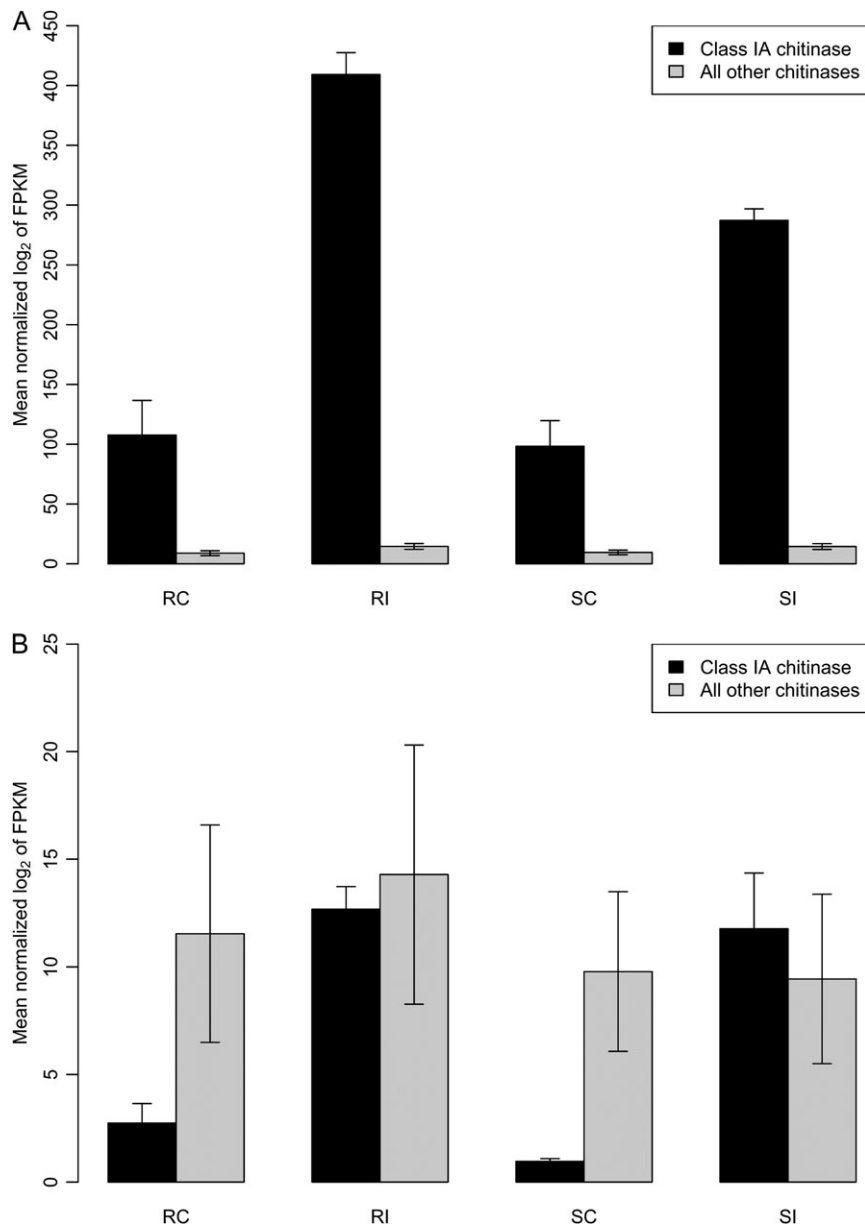
Figure 6. Comparison of the mean and standard error of normalized $\log_2$ of FPKM values for *Class IA chitinase* expression (Eucgr.I01495) (black bars) in *Eucalyptus grandis* against the mean normalized $\log_2$ of FPKM values for combined all other putative chitinases (gray bars) under (A) *Chrysoporthe austroafricana* challenge (3 days post-inoculation) from stem samples and (B) *Leptocybe invasa* challenge (7 days post-infestation) from leaf samples. RC = resistant, control; RI = resistant, inoculated; SC = susceptible, control; SI = susceptible, inoculated.
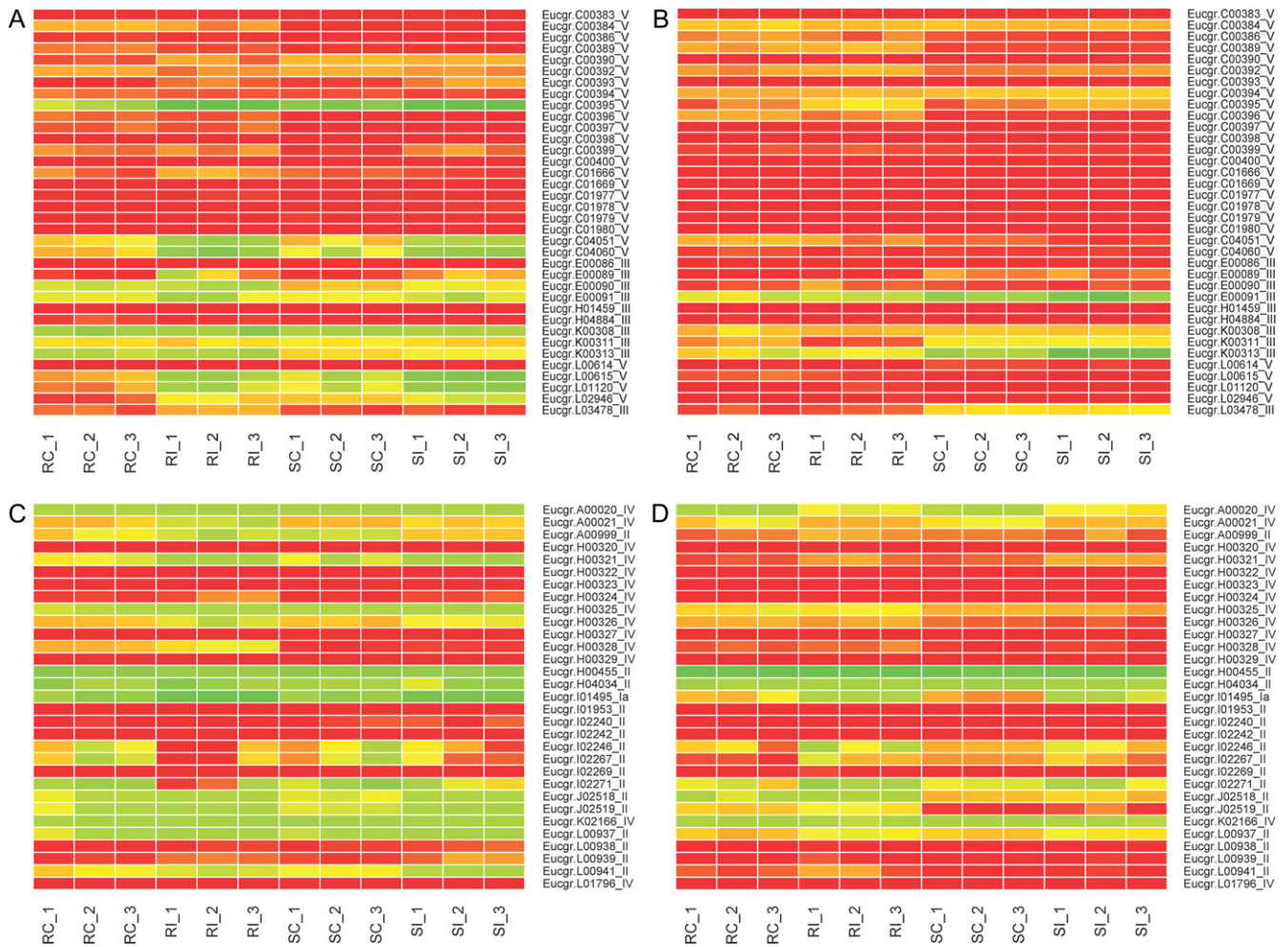
Figure 7. Heatmaps of putative chitinase gene expression for *Eucalyptus grandis* clones. (A) Glycosyl hydrolase 18 (GH18) chitinases, fungal treatment (*Chrysoporthe austroafricana*); (B) GH18 chitinases, insect treatment (*Leptocybe invasa*); (C) glycosyl hydrolase 19 (GH19) chitinases, fungal treat-ment; and (D) GH19 chitinases, insect treatment. Red: 0–1 (very low–low expression), yellow: 1–10 (low to medium expression), green: 10–220 (medium to very high expression). RC = resistant, control. RI = resistant, inoculated. SC = susceptible, control. SI = susceptible, inoculated.

regulated in response to *L. invasa*. Putative *Class III chitinases* were not significantly differentially expressed in either treatment, though expression was increased for Eucgr.E00090 (susceptible) and Eucgr.E00091 and Eucgr.L03478 (moderately resistant) in response to *C. austroafricana*.

While the ANOVA interaction effect ('treatment*genotype') was largely non-significant, indicating that significant 'treatment' *P*-values are due to true treatment response, there was a strong interaction effect (interaction *P*-value <0.0005) for one of our very significant resistant treatment (treatment *P*-value <0.0005) candidates: Eucgr.A00021 (*C. austroafricana*). A single significant resistant treatment (treatment *P*-value <0.0005): Eucgr.C00397 (*L. invasa*) had an interaction *P*-value <0.005. The treatment*genotype interaction effect of these two genes may influence the strong treatment effect that we observe. Of the following strong treatment effect candidates (*** Table 1), interaction effect *P*-values were <0.05: Eucgr.C04051, Eucgr.I01495, Eucgr.J02519, Eucgr.L00615, Eucgr.L01120 (*C. austroafricana*), and Eucgr.C00395 (*L. invasa*).

Validation of *Class IA chitinase* expression

The relative expression of *Class IA chitinase* (Eucgr.I01495) was higher in both susceptible and resistant pants, three days following *C. austroafricana* inoculation (S2 Fig. A). The ratio of gene expression change was determined with amplification efficiency calculations from serial dilutions of a reference gene (*Elongation factor S-II*) and *Class IA chitinase* (Pfaffl 2001), however efficiencies were not highly consistent. Nevertheless, results from quantitative reverse transcriptase PCR (qRT-PCR) were in accordance with expression data derived from RNAseq for the *Class IA chitinase* (Figure 7). Regression of FPKM fold change values against qRT-PCR fold change confirmed this ($R^2 = 0.89$, slope = 1.33). Of interest, the fold change for resistant plant two (R2) was not as high as for other plants however constitutive expression, in controls, was much higher in this plant (S2 Fig B).

Discussion

Expansion of chitinases in *Eucalyptus grandis*

Here we have identified and described the expression profiles of the putative chitinase gene family within the *E. grandis* genome assembly. We identified 67 chitinase gene models within two major families of endochitinases present in plants, glycosyl-hydrolase 18 and 19. Of the 67 putative chitinase genes we found expression data for 57 which, represents a gene expansion in *E. grandis* compared to other plants. Previous analyses of chitinases within woody plants have identified 39 (*Hevea brasiliensis*) and 37 (*Populus trichocarpa*) (Jiang et al. 2013, Misra 2015), while in non-woody plants 24 (*Arabidopsis thaliana*) and 37 (*Oryza sativa*) have been identified (Passarinho and DeVries, 2002). The tandem duplication of defence-related proteins has previously been reported for *E. grandis* in support of our findings (Naidoo et al. 2014, Külheim et al. 2015, Christie et al. 2016) while the identified clustering and phylogenetic analysis of putative chitinases within the current study further establishes this phenomenon (Figures1, 2A and 2B).

The *E. grandis* genome contains thirty-six predicted GH18 and thirty-one predicted GH19 gene models within five recognized classes; one Class I, 16 Class II, 10 Class III, 14 Class IV and 26 Class V. The comparative numbers in *H. brasiliensis* and *P. trichocarpa* for these classes were fewer, except for Class III (16 and 13 respectively) and Class I (7 and 11 respectively) (Jiang et al. 2013, Misra 2015). We note that, despite an expansion of chitinase genes, a single putative *Class IA chitinase* is identified for *E. grandis*. Known to be up-regulated in response to biotic stress we identified and validated the single *Class IA chitinase* transcript to be highly up-regulated in *E. grandis* under different biotic stresses (Table 1).

It has been speculated that Class I, II and IV chitinases are involved in the enzymatic hydrolysis of soluble chitin polymers in the apoplast, which results in oligomer fragments that may

be further cleaved by Class III chitinases (Collinge et al. 1993). Class IA and V, which are more effective against crystalline chitin, are believed to be activated once hyphae penetrate the cell and the vacuolar contents are released (Kasprzewska 2003). The sequences representing the active sites of Class IA chitinases are under diversifying selection, probably driven by pathogen evolution (unlike secreted chitinases, which have highly conserved active sites) (Bishop et al. 2000). The *E. grandis Class IA chitinase* gene model includes the C-terminal vacuolar targeting sequence characterizing Class IA chitinases (Nakamura and Matsuoka 1993; Kasprzewska A 2003). Though the presence of the C-terminal extension suggests vacuolar targeting, the N-terminal signal sequence indicates secretion. *Class I chitinase* genes within many species are present in two isoforms, Class IA and IB, with the presence of the vacuolar targeting sequence signifying Class IA and the absence of this motif signifying secretion (Renner and Specht 2012). The identification of only one *Class I chitinase* within *E. grandis,* and such high expression observed in response to pest and pathogen, suggests that the enzyme that may be variously ascribed (secreted or vacuolar) based on other cellular conditions such as phytohormonal induction. An *in vitro* study using *Nicotiana tabacum* (tobacco) cells found that vacuolar targeted Class I chitinases were secreted when the medium contained 2,4-dichlorophenoxyacetic acid (2,4-D), a synthetic analog of the phytohormone auxin (Kunze et al. 1998). We therefore speculate that phytohormonal induction of cleavage might explain the differential expression responses seen in fungal and insect interactions (Table 1). Identifying cellular locations for this Class I chitinase, following biotic challenge, would help to elucidate this while another important avenue of research involves chaperone proteins which are known to be involved in the lytic vacuolar pathway (daSilva et al. 2006). If chaperone protein expression is inhibited, under pathogenesis or phytohormonal induction, then one would expect the default secretion pathway to take precedence.

Recent evolution, clustering and co-expression of chitinases in *E. grandis*

Chitinases within the *E. grandis* assembly are highly clustered with all but eight putative genes from the 57 chromosome-assigned genes occurring within clusters. Ten putative *Class IV chitinases* are located within a 130 kb region on chromosome eight (cluster H-1) (Figure 5A). Furthermore, the high degree of sequence similarity suggests that they have arisen from recent gene duplication events. The translated sequences of these chitinases differ in some cases by only a few amino acids and yet significant differential expression is evident between susceptible and resistant plants in response to fungal challenge (Table 1B). The duplication and divergence of these gene sequences allows for accumulation of variation and perhaps new roles (Passarinho and DeVries 2002). A cluster of 14 *Class V chitinases* on chromosome three (cluster C-1) occur within 316 kb and include significantly up-regulated transcripts in both biological treatments.

Three putative *Class II chitinases* on chromosome nine are present within a region that is approximately 212 kb in size (Figure 5B). At least three protein-kinase leucine-rich repeat (LRR) genes are co-located with these chitinase genes. We did not determine significant expression within this cluster (cluster I-1), however the co-linear presence of two major defence gene family groupings, GH19 chitinases and kinase-LRR, is intriguing perhaps suggests that this region may provide an active transcriptional zone in response to pathogens. The genomic co-location of functionally related genes has been previously noted in plants, particularly among genes for defence and secondary metabolites, with the implication that they are co-transcribed in response to stimulus (Williams and Bowles 2004, Field and Osbourn 2012).

Differential expression patterns in resistant and susceptible *E. grandis* clones

We determined the significant up-regulation of putative chitinases in response to two biotic stressors. An additional transcriptome dataset was also investigated for comparative expression of

putative chitinases within fungal challenged eucalypts (Chen et al. 2015). The study reviewed expression within resistant *Eucalyptus tereticornis* x *E. urophylla* challenged with *Calonectria pseudoreteaudii* which causes leaf blight. Gene identifiers and accompanying data for putative chitinases were extracted from significantly differentially regulated gene lists at 24 hours post inoculation (*C.p.* Table 1).

Expression was significantly up-regulated across all biotic treatments for the single *Class IA*, a *Class IV* (Eucgr.H00321), and a *Class V chitinase* (Eucgr.C00395), suggesting that these genes are important in defence against both pests and pathogens. Expression of the *Class IA chitinase* was highly up-regulated across all biotic treatments (Table 1), and in comparison to all other chitinase genes, therefore identifying it as a candidate gene of importance (Figure 6A and B). Validation of *Class IA chitinase* expression, within the *C. austroafricana* (fungal) study, using qRT-PCR confirmed these transcriptome results (S2 Fig.A and B). In an experiment with Isobaric tags for relative and absolute quantitation (iTRAQ) conducted in our laboratory, the Class IA chitinase protein was up-regulated following inoculation with *C. austroafricana*, but not wounding, supporting a role in pathogen defence (Lizahn Zwart, unpublished). In accordance with these findings, *Casuarina equisetifolia*, *Class IA chitinase* expression was significantly up-regulated in leaves in response to pathogen but not to wounding (Veluthakkal and Dasgupta 2012).

Of additional interest was the very high constitutive expression of the *Class IA chitinase* within the fungal study (Figure 6A) in comparison with the *L. invasa* (insect) study (Figure 6B). As an example, baseline expression levels within resistant control plants from the insect study were relatively low (mean $\log_2$ of FPKM value = 2.7, Figure 6B - RC), compared with resistant plants in the fungal study (mean $\log_2$ of FPKM value = 107.7, Figure 6A - RC), indicating that these baseline levels play a role in fungal resistance. Up-regulation was also much greater within

the fungal study (mean $\log_2$ of FPKM value = 409.1, Figure 6A - RI) compared to the insect study (mean $\log_2$ of FPKM value = 12.7, Figure 6B - RI), in line with previous findings that indicate the role of gall insects in systemically suppressing defence (Tooker et al. 2008). Indeed, the selection of traits for fungal or insect resistance would necessitate variation in genotype, as appears evident here, and indicates that breeding for combined traits is an important consideration. Nevertheless, both insect and fungal inoculation clearly stimulates significant increase in expression. The lower induction of chitinases from *L. invasa* challenged clones provides further evidence for the proposal of host defence manipulation (Tooker et al. 2008).

We identified multiple chitinase genes within physical clusters that were up- or down-regulated following biotic challenge. While one would expect that physically clustered genes might simply indicate transcriptionally active sites, we note that *Class IV* genes in cluster H-1 under *C. austroafricana* challenge show different and specific genes up-regulated in resistant plants in comparison with susceptible plants (Table 1). Also interesting is the presence of a cluster (A-1) that is significantly down-regulated in response to *L. invasa* and yet one of these *Class IV* genes is significantly up-regulated in resistant plants in response to both fungal treatments. A similar response is seen at cluster C-4 where two *Class V* genes are significantly up-regulated in response to fungal challenge and yet, in susceptible plants, one of these genes is down-regulated in response to *L. invasa*.

While expression was significantly up-regulated across all biotic treatments for the single *Class IA*, a single *Class IV* (Eucgr.H00321), and *Class V chitinases* (Eucgr.C00395), suggesting that these genes are important in defence against both pests and pathogens, it should also be appreciated that samples were taken from stems three days post inoculation for *C. austroafricana*, from leaves 24 hours post inoculation for *C. pseudoreteaudii*, and from leaves seven days post

infestation for *L. invasa*. Comparisons between these biotic challenges are perhaps less meaningful than comparisons between resistant and susceptible plants within treatments although other studies have identified similar trends. For example, *Class IA chitinase* in fungal challenged *Casuarina equisetifolia* showed the highest levels of induction at 48 hours post inoculation (Veluthakkal and Dasgupta 2012). It has also been noted in *Panax ginseng* that *Class I chitinase* expression reaches highest up-regulation within 12 hours of biotic exposure (Pulla et al. 2011). In oil palm (*Elaeis guineensis*), expression responses continued to increase for a *Class III* and *Class V chitinases* up to 12 weeks post inoculation with pathogenic fungus (Yeoh et al. 2013), however there was no apparent difference to *Class I chitinase*, perhaps due to the timing of sample collections (at 3, 6 and 12 weeks). These studies suggest that future studies within *E. grandis* should monitor chitinase expression at daily intervals from 24 hours post-inoculation/infestation through to one week. For the single *Class IA chitinase*, all treatments were significantly up-regulated (Table 1), despite these time-scale variations.

Roles in chitin perception

Predicted trans-membrane regions in *E. grandis* chitinases were identified in Class IV (7), Class II (1), Class III (7) and Class V (9) sequences (Table 1). Class III (3) and Class V (1) gene models possess large intracellular catalytic regions, while the remainders have extracellular catalytic regions. The predicted chitinases with extracellular catalytic regions may have a role in binding and cleaving chitin on the cell surface. Alternatively they may be important for lysozomal activity within autophagic vesicles which are thought to fuse with the vacuole (Chung 2011). If these enzymes are important for lysozomal degradation of proteins (Brunner et al. 1998), it is conceivable that the TM motif acts as a means of targeting to the vacuole, through vesicle fusion, as vacuolar targeting sequences were not identified in these sequences. Many of these predicted

membrane-bound chitinases, however, possess catalytic active sites that are required for chitin cleavage, though in some this region is truncated. Perhaps the presence of these predicted membrane-bound chitinases in *E. grandis* indicates a role in perception and response initiation as determined for chitin oligosaccharide elicitor-binding protein in rice (Kaku et al. 2006). The presence of slightly truncated catalytic regions in some membrane-bound chitinases (for example, Eucgr.L03478) may indicate a role in ligand-binding, rather than cleavage, and further roles, such as potential signal transduction, would be interesting to investigate. One possibility is initiating receptor-mediated endocytosis by internalising chitin fragments and stimulating further defence transcription (Leborgne-Castel et al. 2010). Deletions within GH19 catalytic sites of some non-membrane bound chitinases, for example Eucgr.L00941 which is highly significantly up-regulated in *C. austroafricana* challenge, indicate that active sites may be absent or reduced. It is therefore not clear if they are involved in chitin degradation or whether they warrant scrutiny for alternative roles. The defensive up-regulation of chitinase-like proteins, with missing catalytic regions, has been previously reported and suggested to be the result of duplications leading to alternative functions (Kesari et al. 2015). A previous study identified the inhibitory nature of the chitin binding domain on fungal hyphal development without the presence of a catalytic domain (Garcia-Casado et al. 1998). Certainly chitinases are identified as regulating developmental stages of plant growth, particularly during embryogenesis and fruit ripening, though their specific role is not clear (Arjon *et al.*, 1998; Peumans *et al.*, 2002). A membrane-bound chitinase-like protein (*Brittle Culm15*) which contains a truncated Class II GH19 domain was identified in rice and found to play a role in cellulose biosynthesis (Wu et al. 2012).

Additional domains in *E. grandis* chitinases

Some putative *E. grandis* chitinases incorporate additional domains such as a fused Legume Lectin domain in the Class II chitinase, Eucgr.I02271 which was down-regulated in fungal interactions (Table 1). This domain, involved in binding to mannose sugars, is present as an extracellular domain within receptor-like kinases of rice blast resistance genes (Chen et al. 2006). An N-terminal cyclin dependent kinase inhibitor 3 (CdkN3) domain is associated with the Class III chitinase (Eucgr.K00313). As chitinases are indicated to have a role in programmed cell death (PCD) (Passarinho and DeVries 2002) and the induction of SAR (Zhang et al. 2012), it is conceivable that the presence of a domain involved in inhibition of cell division may be a consequence of the immune response to fungal infection. The CdkN3 protein has a role in regulating human cancers, although a role in plants has not been established (Xing et al. 2012). The ornatin domain (glycoprotein antagonist) (Eucgr.L00939) from Class II chitinase and Arr-ms domain (Rifampin ADP-ribosyl transferase) (Eucgr.A00021 and Eucgr.H00320) from Class IV chitinases also suggest that these proteins may possess hitherto uncharacterized roles in host immune responses. The *Class II chitinase* (Eucgr.L00939) was significantly up-regulated in all fungal interactions. The *Class IV chitinase* (Eucgr.A00021) was only significantly up-regulated within resistant clones in response to fungal interactions and significantly down-regulated in response to insect.

Conclusions

Chitinase enzymes in woody plants have been less well studied than in herbaceous plants. Here we provide evidence, within the *E. grandis* genome assembly, for 67 chitinase gene models within two families: glycosyl hydrolase 18 and 19. We note that, although the *E. grandis* genome has an expanded chitinase gene family, only a single putative *Class IA chitinase* is present unlike the

genomes of *Populus trichocarpa* and *Hevea brasiliensis*. Known to be up-regulated in response to biotic stress in other species, we identified and validated the single *Class IA chitinase* transcript to be highly up-regulated in *E. grandis* following fungal challenge, and there is further evidence that the protein levels are up under these conditions from recent unpublished work. We propose that the single *Class IA chitinase* is differentially secreted or vacuolar targeted based on the presence of biotic stressors. Further to this, we determined that many chitinase genes are closely physically clustered and that genes within clusters are co-expressed under biotic challenge. Our study has used the most recent functional genomics data available to identify key genes involved in defence. These results provide the basis for future gene characterization and the development of targeted responses to pests and diseases within an important forestry species.

Materials and Methods

Identification of putative chitinase genes

The *Eucalyptus grandis* annotation information file that was released as part of Phytozome v8.0 (Egrandis_201_annotation_info.txt, (JGI Phytozome: The Plant Genomics Resource)) was used to identify an initial list of 61 predicted chitinase genes (Goodstein et al. 2012). Separate *E. grandis* specific nucleotide Hidden Markov Models (HMM) were constructed with 26 x GH18 and 23 x GH19 aligned nucleotide sequences using HMMER3.0 (Eddy 2010). The *E. grandis* specific HMMs were used to search for additional GH18 and 19 chitinase gene models within the *E. grandis* transcript sequence data (Egrandis_201_transcript.fa) using NHMMER. All *E. grandis* translated sequences for gene identifiers above the NHMMER inclusion threshold were used in downstream analysis, including 36 putative GH18 and 31 putative GH19 sequences (67 in total). Two sequences identified within the annotation file (Eucgr.A02678 and Eucgr.A02680) did not fall within the inclusion threshold and were discarded from further analysis.

Alignment and phylogenetic analysis

The full GH18 and GH19 amino acidsequences were aligned using ClustalW (Thompson et al. 1994) with default parameters within Mega6 (Tamura et al., 2013). As the two families share no sequence similarity a single phylogeny was not employed. A neighbor joining tree with all default settings was conducted within Mega6. Outliers used for chitinase phylogeny from *Arabidopsis thaliana* were GI:820429 (Class I), GI:51971117 (Class IV) and GI:110740739 for GH19; and GI:145358370, GI:119360133 (Class V) for GH18.

Identification of conserved domains

GH18 and GH19 sequences were submitted to Dialign-Pfam (Al Ait et al. 2013) with all default settings to identify domains. Based on scrutiny of their aligned sequences, in comparison with previous chitinase characterization in *Arabidopsis thaliana* (Collinge et al. 1993, Passarinho and DeVries 2002), they were determined to fall within five recognized classes within the two glycosyl hydrolase catalytic families. MEME version 4.10.1 (Multiple Expectation Maximization for Motif Elicitation) (Bailey et al. 2009) analysis was used to identify conserved motifs based on *Arabidopsis thaliana* chitinases (Passarinho and DeVries 2002) using default parameters. The MEME results, together with multiple sequence alignments, aided identification of the conserved domains. A subset of the translated sequences, using genes that were significantly differentially regulated in two biotic treatments, were aligned using the Clustal algorithm in the BioEdit (Hall 1991) software package (Figures 4a and b). GH18 and GH19 predicted sequences were also searched, using BLAST, for vacuolar targeting sequences in accordance with described motifs (Nakamura and Matsuoka 1993, Neuhaus et al. 1994).

Physical cluster analysis

Clusters and superclusters of predicted chitinase genes were identified based on previously defined criteria (Christie et al. 2016). In brief: a gene cluster is a genomic region containing at least three chitinase genes, (i) with less than 9 other genes between neighboring chitinase genes and (ii) in which two neighboring chitinase genes are less than 250 kb apart; a gene supercluster is a genomic region containing at least one chitinase gene cluster and at least two additional chitinase genes, (i) with less than 99 other genes between neighboring chitinase genes and (ii) in which two neighboring chitinase genes are less than 2500 kb apart.

Visualization of chitinase genes on chromosomes

The positions of the GH18 and GH19 classes of putative chitinase genes were visualized by mapping to the eleven *E. grandis* chromosomes using base pair start positions in Mapchart2.2 (Voorrips 1994) (Figure 1).

Chitinase predicted protein structure

The peptide sequence of the single identified Class IA chitinase (Eucgr.I01495) was submitted to the I-Tasser server (I-TASSER: Protein structure and function predictions, Roy et al. 2010) to determine predicted protein structures. The resulting protein model was compared to the closest structural homolog, a crystal structure for Class I chitinase from *Oryza sativa* L. japonica (Kezuka et al. 2010) and catalytic sites were identified based on previously identified residues (Bishop et al. 2000).

Chitinase transmembrane motifs and cleavage sites identified

The sequences of translated putative GH18 and GH19 chitinase genes were run through a protein trans-membrane and cleavage site prediction server to identify motifs (TMHMM Server v. 2.0 Prediction of transmembrane helices in proteins, Emanuelsson et al. 2000) (Table 1).

Biotic stress trials

Resistant and susceptible clones of *E. grandis* were challenged with the fungal stem canker pathogen (*Chrysoporthe austroafricana*) and leaf gall wasp (*Leptocybe invasa*) as previously described (Mangwanda et al. 2015, Oates et al. 2015). Briefly, stem samples were harvested around the inoculation site of three biological replicates (three ramets of each) from moderately resistant and susceptible one year old *E. grandis* clones three days post inoculation (*C. austroafricana*). At this time point, no difference in lesion lengths were observed between the moderately resistant and susceptible genoptypes. Control plants of the biological replicates (three ramets of each), were mock inoculated (Mangwanda et al. 2015). All plants were grown under controlled conditions in a randomized block trial. Gall wasp trials were conducted on two-year old resistant and susceptible *E. grandis* clones (three ramets of each) which had been coppiced and allowed to regrow for four months. Control plants were grown in wasp-proof mesh while treatment plants were exposed to natural infestation. Infested and uninfested leaf samples were collected seven days post oviposition with (*L. invasa*). At this time-point, oviposition holes were evident in both genotypes however galls had not started to develop in the susceptible genotype as of yet. Total RNA was extracted and sent to the Beijing Genome Institute (BGI) for RNA-Sequencing using the Illumina Genome Analyser with a 50 bp paired end module (Illumina, San Diego, CA) (described by Mangwanda *et al.*, 2015 and Oates *et al.*, 2015).

RNA-Seq data analysis

RNA-Seq data, generated in Mangwanda et al. 2015; Oates et al. 2015, was analyzed using the Galaxy workspace as previously described (Christie et al. 2016). In brief, FASTQC v0.52 was used to verify RNA-Seq data quality. Reads were mapped to the E. grandis v1.1 genome assembly using Bowtie (Langmead et al. 2009a) (Langmead et al. 2009b) and Tophat2 v2.0.9 (Trapnell et al. 2009) (Trapnell et al., 2009). Unique reads as well as reads that mapped to <20 locations were used, but fragment bias correction and multi read mapping correction was applied in Cuffdiff (Roberts et al. 2011) (Trapnell et al. 2010)(Trapnell et al., 2010; Roberts et al., 2011). Mapped reads were assembled into transcripts and fragments per kilobase of exon per million fragments mapped (FPKM) values were calculated with Cufflinks v2.1.1 (Trapnell et al., 2010). Quartile normalization was conducted in Cufflinks. The Eucalyptus data sets supporting these results are available in the NCBI Gene Expression Omnibus repository for *Chrysoporthe austroafricana* challenge (GSE67554: Mangwanda *et al.*, 2015) and NCBI BioProject ID PRJNA305347 for *Leptocybe invasa* challenge (Oates et al. 2015).

Chitinase transcript expression analysis

Expression profiles for the putative chitinase genes were extracted from a transcriptome-wide expression matrix using a custom Python script. Expression data were only available for genes in *E. grandis* v1.1. Analysis of variance (ANOVA) for putative *chitinase* genes from treatment, control, resistant and susceptible groups was performed in GenStat (v. 16.2.0.11713, VSN International, Hemel Hempstead, UK). False discovery rate calculations on ANOVA *P*-values, using a less stringent cut-off, produced similar results. We therefore proceeded with analysis on ANOVA *P*-values using the 'treatment' effect only. Interaction effect ('treatment*genotype') was largely non-significant indicating that significant 'treatment' *P*-values are due to treatment

response.

The *E. grandis* datasets from both *C. austroafricana* and *L. invasa* were analyzed separately. Expression analysis was based on the $\log_2$ fold change of inoculated versus control samples. Genes in resistant and susceptible plants were considered up or down-regulated if their $\log_2$ gene expression ratios were greater than 1 or less than -1. Differential expression was determined by taking significant *P*-values (<0.05 =*, <0.005 =**, <0.0005=***) from the ANOVA analysis by treatment and comparing this data with fold change values. Heatmaps that depict gene expression (as $\log_2$ of the normalized read count (FPKM)) of both gene expression and treatments, were drawn in R studio (v. 0.98.981, (RStudio Team 2015)(RStudio Team, 2015)) using the gplots and RColorBrewer packages. Color breaks were non-linear at 1, 10 and the maximum $\log_2$ FPKM value, using red: 0-1 (very low - low expression), yellow: 1-10 (low to medium expression), green: 10-220 (medium to very high expression).

Chitinase primer design

The putative *E. grandis Class IA* gene and a housekeeping gene (*Elongation factor S-II*, Eucgr.A00774) were used to design primers using Primer3 version 0.4.0 (http://frodo.wi.mit.edu/). Forward and reverse primers derived from these were individually run through BLAST against the *E. grandis* genome within Phytozome (v.1) for target specificity and then checked in MWG Operon oligo analysis tool (Eurofins Scientific, Luxembourg) for sequence complementarity . Primers were ordered from Macrogen, tested on cDNA using Kapa SYBR® FAST mastermix with the following thermocycle conditions: once cycle of 95 ºC for 2 minutes then 40 cycles of 95 ºC for 30 seconds, 60 ºC for 30 seconds and 72 ºC for 1 minute followed by cycle of 72 ºC for 5 minutes. Resulting amplicons were visualized using 2% agarose gel electrophoresis and sequences validated with Macrogen EZI-seq.

Quantitative reverse-transcriptase PCR (qRT-PCR)

Complementary DNA (cDNA), synthesized using the Improm-II^{TM} Reverse Transcription System (Promega, Wisconsin, USA), with RNA isolated from moderately resistant and susceptible clones of *E. grandis* inoculated and mock-inoculated with *C. austroafricana*, was generated in the study by Mangwanda *et al*. (2015) and kindly provided as a gift. Insufficient cDNA was extracted from one biological replicate (R3), made from a pool of six ramets of the moderately resistant clone, and therefore was not included in qRT-PCR. The *Class IA chitinase* (F: ACGTTAGTGGCCTCATCTCG, R: CGTTTCATGGGAAGTCTGTG) and house-keeping gene (Eucgr.A00774), *Elongation factor S-II*, (F: TCCAATCCGAGTCGCTGTCATTGT, R: TGATGAGCCTCTCTGGTTTGACCT) primer pairs were run separately on 96 well plates (20 µl) using BioLine SensiFAST™ SYBR No-ROX Kit with resistant and susceptible samples of both inoculated and control cDNA. Serial dilutions of cDNA (neat (~100ng), 1 in 5, 1 in 25, and 1 in 125) were loaded in triplicate (1 µl per 20 µl sample volume). PCR thermocycle conditions were as per primer assay (annealing temperatures for both primer pairs were 60 ℃). Melt curve analysis was performed; 65 - 95 ℃ with 0.5 ℃ increments every 15 seconds.

Relative chitinase gene expression in inoculated plants in relation to control plants was calculated, with calibration to the reference gene, *EFs-II*, using efficiency corrected calculation models based on multiple samples according to the following equation (Pfaffl 2001):

$$ratio = \frac{(E_{target})^{\Delta Ct_{target}(MEAN\ control-MEAN\ sample)}}{(E_{reference})^{\Delta Ct_{reference}(MEAN\ control-MEAN\ sample)}}$$

where E is efficiency of amplification and delta Ct is the change in threshold cycles of amplification. Efficiency (E) was calculated using linear regression slopes of mean Ct values

against the logarithmic value of cDNA concentrations using the equation below (Pfaffl 2001). A minimum of three data points were used for regression equations.

$$E = 10^{(-1/slope)}$$

Availability of data

RNAseq data is available at NCBI Gene Expression Omnibus repository for *Chrysoporthe austroafricana* challenge (GSE67554: Mangwanda *et al.*, 2015) and NCBI BioProject ID PRJNA305347 for *Leptocybe invasa* challenge. All additional data is available in the supplementary files.

Competing interests

The authors declare no conflict of interest in the reporting of these results.

Author contributions

PT and NC shared the lead author duties for this manuscript. PT conducted some laboratory work, data and gene expression analysis and much of the manuscript writing. NC carried out much of the data analysis and aspects of the manuscript drafting. DG, SN and CK were involved in the design and co-ordination of the study and conducted some of the expression analysis and writing. All authors read and approved the final manuscript.

Acknowledgements

References

Al Ait L, Yamak Z, Morgenstern B (2013) DIALIGN at GOBICS--multiple sequence alignment using various sources of external information. Nucleic Acids Res 41:W3-7.

Arjon J van H, Guzzo F, Kammen A van, Vries SC de (1998) Expression Pattern of the Carrot EP3 Endochitinase Genes in Suspension Cultures and in Developing Seeds. Plant Physiol 117:43–53.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME Suite: Tools for motif discovery and searching. Nucleic Acids Res 37:202–208.

Bishop JG, Dean AM, Mitchell-Olds T (2000) Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. Proc Natl Acad Sci U S A 97:5322–5327.

Boava LP, Kuhn OJ, Pascholati SF, Di Piero RM, Furtado EL (2009) Effect of acibenzolar-S-methyl and Saccharomyces cerevisiae on the activation of Eucalyptus defences against rust. Australas Plant Pathol 38:594–602.

Boland DJ, Brooker MIH, Chippendale GM, Hall N, Hyland BPM, Johnson RD, Kleinig DA, McDonald MW (2006) Forest Trees of Australia. CSIRO Publishing, Collingwood.

Brunner F, Stintzi A, Fritig B, Legrand M (1998) Substrate specificities of tobacco chitinases. Plant J 14:225–34.

Chen Q, Guo W, Feng L, Ye X, Xie W, Huang X, Liu J (2015) Data for transcriptome and proteome analysis of Eucalyptus infected with *Calonectria pseudoreteaudii*. J Proteomics 3:24–28. http://dx.doi.org/10.1016/j.jprot.2014.12.008

Chen X, Shang J, Chen D, Lei C, Zou Y, Zhai W, Liu G, Xu J, Ling Z, Cao G, Ma B, Wang Y, Zhao X, Li S, Zhu L (2006) A B-lectin receptor kinase gene conferring rice blast resistance. Plant J 46:794–804. http://dx.doi.org/10.1111/j.1365-313X.2006.02739.x (2 May 2014, date last accessed).

Christie N, Tobias PA, Naidoo S, Külheim C (2016) The *Eucalyptus grandis NBS-LRR* gene family: physical clustering and expression hotspots. Front Plant Sci 6:1–16.

Chung T (2011) See How I Eat My Greens—Autophagy in Plant Cells. J Plant Biol 54:339–350. http://dx.doi.org/10.1007/s12374-011-9176-5 (2 May 2014, date last accessed).

Collinge DB, Kragh KM, Mikkelsen JD, Nielsen KK, Rasmussen U, Vad K (1993) Plant chitinases. Plant J 3:31–40. http://www.ncbi.nlm.nih.gov/pubmed/8401605

daSilva LLP, Foresti O, Denecke J (2006) Targeting of the plant vacuolar sorting receptor BP80 is dependent on multiple sorting signals in the cytosolic tail. Plant Cell 18:1477–1497.

Davis JM, Clarke HR, Bradshaw HD, Gordon MP (1991) Populus chitinase genes: structure, organization, and similarity of translated sequences to herbaceous plant chitinases. Plant Mol Biol 17:631–9.

http://www.ncbi.nlm.nih.gov/pubmed/1912489

Dittrich-Schröder G, Wingfield MJ, Hurley BP, Slippers B (2012) Diversity in Eucalyptus susceptibility to the gall-forming wasp *Leptocybe invasa*. Agric For Entomol 14:419–427.

Eddy SR (2010) HMMER user's guide. HMMER websiteftp//selab.janelia.org/pub/software/hmmer/CURRENT/Userguide.pdf (accessed 10 December 2014)

Edreva A (2005) Pathogenesis-related proteins:research progress in the last 15 years. Gen Appl Plant Physiol 31:105–124.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016.

Field B, Osbourn A (2012) Order in the playground. Formation of plant gene clusters in dynamic chromosomal regions. Mob Genet Elements 2:46–50.

Fu ZQ, Dong X (2013) Systemic acquired resistance: turning local infection into global defense. Annu Rev Plant Biol 64:839–63. http://www.ncbi.nlm.nih.gov/pubmed/23373699

Garcia-Casado G, Collada C, Allona I, Casado R, Pacios LF, Aragoncillo C, Gomez L (1998) Site-directed mutagenesis of active site residues in a class I endochitinase from chesnut seeds. Glycobiology 8:1021–1028.

Genomics Science Program Systems Biology for Bioenergy. http://genomicsgtl.energy.gov/biofuels/ (17 June 2016, date last accessed).

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178-86. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3245001&tool=pmcentrez&rendertype=abstract (29 April 2014, date last accessed).

Grover A (2012) Plant Chitinases: Genetic Diversity and Physiological Roles. CRC Crit Rev Plant Sci 31:57–73. http://www.tandfonline.com/doi/abs/10.1080/07352689.2011.616043 (2 May 2014, date last accessed).

Hall TA (1991) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98.

I-TASSER: Protein structure and function predictions. http://zhanglab.ccmb.med.umich.edu/I-TASSER/

JGI Phytozome: The Plant Genomics Resource. http://phytozome.jgi.doe.gov/pz/portal.html

Jiang C, Huang RF, Song JL, Huang MR, Xu LA (2013) Genome-wide analysis of the chitinase gene family in Populus trichocarpa. J Genet 92:121–125.

Kaku H, Nishizawa Y, Ishii-Minami N, Akimoto-Tomiyama C, Dohmae N, Takio K, Minami E, Shibuya N (2006) Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. Proc Natl Acad Sci 103:11086–11091. http://www.ncbi.nlm.nih.gov/pubmed/16829581

Kasprzewska A (2003) Plant chitinases--regulation and function. Cell Mol Biol Lett 8:809–824.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kesari P, Patil DN, Kumar P, Tomar S, Sharma AK, Kumar P (2015) Structural and functional evolution of chitinase-like proteins from plants. Proteomics 15:1693–1705.

Kezuka Y, Kojima M, Mizuno R, Suzuki K, Watanabe T, Nonaka T (2010) Structure of full-length class I chitinase from rice revealed by X-ray crystallography and small-angle X-ray scattering. Proteins:2295–2305. http://www.rcsb.org/pdb/explore/explore.do?structureId=2dkv (22 February 2016, date last accessed).

Krishnaveni S, Muthukrishnan S, Liang G., Wilde G, Manickam a (1999) Induction of chitinases and β-1,3-glucanases in resistant and susceptible cultivars of sorghum in response to insect attack, fungal infection and wounding. Plant Sci 144:9–16. http://linkinghub.elsevier.com/retrieve/pii/S0168945299000497

Külheim C, Padovan A, Hefer C, Krause ST, Köllner TG, Myburg A a, Degenhardt J, Foley WJ (2015) The Eucalyptus terpene synthase gene family. BMC Genomics 16:450. http://www.biomedcentral.com/1471-2164/16/450

Kunze I, Kunze G, Bröker M, Manteuffel R, Meins F, Müntz K, Broker M, Muntz K (1998) Evidence for secretion of vacuolar alpha-mannosidase, class I chitinase, and class I beta-1,3-glucanase in suspension cultures of tobacco cells. Planta 205:92–99. http://www.ncbi.nlm.nih.gov/pubmed/9599806

Langmead B, Trapnell C, Pop M, Salzberg SL (2009a) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009b) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19261174

Leborgne-Castel N, Adam T, Bouhidel K (2010) Endocytosis in plant-microbe interactions. Protoplasma 247:177–193. http://www.ncbi.nlm.nih.gov/pubmed/20814704 (2 May 2014, date last accessed).

van Loon LC, van Strien EA (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Pathol 55:85–97. http://igitur-archive.library.uu.nl/bio/2002-0103-100912/UUindex.html

Mangwanda R, Myburg AA, Naidoo S (2015) Transcriptome and hormone profiling reveals *Eucalyptus grandis* defence responses against *Chrysoporthe austroafricana*. BMC Genomics 16:1–13. http://www.biomedcentral.com/1471-2164/16/319

Mangwanda R, Zwart L, van der Merwe N, Moleleki L, Berger D, Myburg A, Naidoo S (2016) Localization and transcriptional responses of *Chrysoporthe austroafricana* in *Eucalyptus grandis* identify putative pathogenicity factors. Front Microbiol 7:1953. http://journal.frontiersin.org/article/10.3389/fmicb.2016.01953

Misra BB (2015) Molecular Evolution and Functional Divergence of Chitinase Gene Family in *Hevea brasiliensis* Genome. The Winnower

Moon DH, Salvatierra GR, Caldas DGG, Gallo de Carvalho MCC, Carneiro RT, Franceschini LM, Oda S, Labate C a. (2007) Comparison of the expression profiles of susceptible and resistant *Eucalyptus grandis* exposed to *Puccinia psidii* Winter using SAGE. Funct Plant Biol 34:1010. http://www.publish.csiro.au/?paper=FP07094

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, Goodstein DM, Dubchak I, Poliakov A, Mizrachi E, Kullan ARK, Hussey SG, Pinard D, van der Merwe K, Singh P, van Jaarsveld I, Silva-Junior OB, Togawa RC, Pappas MR, Faria D a.,

Sansaloni CP, Petroli CD, Yang X, Ranjan P, Tschaplinski TJ, Ye C-Y, Li T, Sterck L, Vanneste K, Murat F, Soler M, Clemente HS, Saidi N, Cassan-Wang H, Dunand C, Hefer C a., Bornberg-Bauer E, Kersting AR, Vining K, Amarasinghe V, Ranik M, Naithani S, Elser J, Boyd AE, Liston A, Spatafora JW, Dharmwardhana P, Raja R, Sullivan C, Romanel E, Alves-Ferreira M, Külheim C, Foley W, Carocha V, Paiva J, Kudrna D, Brommonschenkel SH, Pasquali G, Byrne M, Rigault P, Tibbits J, Spokevicius A, Jones RC, Steane D a., Vaillancourt RE, Potts BM, Joubert F, Barry K, Pappas GJ, Strauss SH, Jaiswal P, Grima-Pettenati J, Salse J, Van de Peer Y, Rokhsar DS, Schmutz J (2014) The genome of *Eucalyptus grandis*. Nature 510:356–362. http://www.nature.com/doifinder/10.1038/nature13308 (11 June 2014, date last accessed).

Naidoo S, Külheim C, Zwart L, Mangwanda R, Oates CN, Visser EA, Wilken FE, Mamni TB, Myburg AA (2014) Uncovering the defence responses of Eucalyptus to pests and pathogens in the genomics age. Tree Physiol 34:931–43. http://www.ncbi.nlm.nih.gov/pubmed/25261123 (30 September 2014, date last accessed).

Nakamura K, Matsuoka K (1993) Protein targeting to the vacuole in plant cells. Plant Physiol 101:1–5. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=344539&tool=pmcentrez&rendertype=abstract

Neuhaus JM, Pietrzak M, Boller T (1994) Mutation analysis of the C-terminal vacuolar targeting peptide of tobacco chitinase: low specificity of the sorting system, and gradual transition between intracellular retention and secretion into the extracellular space. Plant J 5:45–54. http://www.ncbi.nlm.nih.gov/pubmed/8130797

Nyeko P, Mutitu EK, Day RK (2009) Eucalyptus infestation by *Leptocybe invasa* in Uganda. Afr J Ecol 47:299–307.

Nyeko P, Mutitu KE, Otieno BO, Ngae GN, Day RK (2010) Variations in *Leptocybe invasa* (Hymenoptera: Eulophidae) population intensity and infestation on eucalyptus germplasms in Uganda and Kenya. Int J Pest Manag 56:137–144.

Oates CN, Külheim C, Myburg AA, Slippers B, Naidoo S (2015) The Transcriptome and Terpene Profile of *Eucalyptus grandis* Reveals Mechanisms of Defense Against the Insect Pest, *Leptocybe invasa*. Plant Cell Physiol 56:1–11. http://pcp.oxfordjournals.org/cgi/doi/10.1093/pcp/pcv064

Passarinho PA, DeVries SC (2002) Arabidopsis Chitinases: a Genomic Survey. Arabidopsis Book 1:1–25. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3243303&tool=pmcentrez&rendertype=abstract (2 May 2014, date last accessed).

Peumans WJ, Proost P, Swennen RL, Damme EJM Van (2002) The Abundant Class III Chitinase Homolog in Young Developing Banana Fruits Behaves as a Transient Vegetative Storage Protein and Most Probably Serves as an Important Supply of Amino Acids for the Synthesis of Ripening-Associated Proteins. Plant Physiol 130:1063–1072. http://www.jstor.org/stable/4280733

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=55695&tool=pmcentrez&rendertype=abstract

Pulla RK, Lee OR, In JG, Parvin S, Kim YJ, Shim JS, Sun H, Kim YJ, Senthil K, Yang DC (2011) Identification and characterization of class i chitinase in Panax ginseng C. A. Meyer. Mol Biol Rep 38:95–102.

Renner T, Specht CD (2012) Molecular and functional evolution of class I chitinases for plant carnivory in the Caryophyllales. Mol Biol Evol 29:2971–2985.

Reymond P, Weber H, Damond M, Farmer EE (2000) Differential Gene Expression in Response to Mechanical Wounding and Insect Feeding in Arabidopsis. Plant Cell 12:707–720. http://www.plantcell.org/cgi/doi/10.1105/tpc.12.5.707

Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol 12:R22. http://genomebiology.com/2011/12/3/R22

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. http://www.nature.com/doifinder/10.1038/nprot.2010.5

RStudio Team (2015) RStudio: Integrated Development for R. RStudio, Inc. Boston, MA.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680.

TMHMM Server v. 2.0 Prediction of transmembrane helices in proteins. http://www.cbs.dtu.dk/services/TMHMM/ (17 June 2016, date last accessed).

Tooker JF, Rohr JR, Abrahamson WG, De Moraes CM (2008) Gall insects can avoid and alter indirect plant defenses. New Phytol 178:657–671.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–11. http://www.ncbi.nlm.nih.gov/pubmed/19289445

Trapnell C, Williams B a, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. Nat Biotechnol 28:511–515. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3146043&tool=pmcentrez&rendertype=abstract

Veluthakkal R, Dasgupta MG (2012) Isolation and characterization of pathogen defence-related class I chitinase from the actinorhizal tree *Casuarina equisetifolia*. For Pathol 42:467–480.

Voorrips RE (1994) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78.

Whyte G, Howard K, Hardy G, Burgess T (2011) Foliar pests and pathogens of *Eucalyptus dunnii* plantations in southern Queensland. Aust For 74:161–169.

Williams EJB, Bowles DJ (2004) Coexpression of Neighboring Genes in the Genome of *Arabidopsis thaliana*. Genome Res 14:1060–1067.

Wu B, Zhang B, Dai Y, Zhang L, Shang-Guan K, Peng Y, Zhou Y, Zhu Z (2012) Brittle Culm15 Encodes a Membrane-Associated Chitinase-Like Protein Required for Cellulose Biosynthesis in Rice. Plant Physiol 159:1440–1452. http://www.plantphysiol.org/content/159/4/1440.abstract (2 May 2014, date last accessed).

Xing C, Xie H, Zhou L, Zhou W, Zhang W, Ding S, Wei B, Yu X, Su R, Zheng S (2012) Cyclin-dependent kinase inhibitor 3 is overexpressed in hepatocellular carcinoma and promotes tumor cell proliferation. Biochem Biophys Res Commun 420:29–35.

Yeoh KA, Othman A, Meon S, Abdullah F, Ho C-LL (2013) Sequence analysis and gene expression of putative oil palm chitinase and chitinase-like proteins in response to colonization of *Ganoderma boninense* and *Trichoderma harzianum*. Mol Biol Rep 40:147–158.

Zhang J-Y, Guo Z-R, Qu S-C, Zhang Z (2012) Identification and Molecular Characterization of a Class I Chitinase Gene (*Mhchit*) from *Malus hupehensis*. Plant Mol Biol Report 30:760–767. http://dx.doi.org/10.1007/s11105-011-0387-1 (2 May 2014, date last accessed).

Zhao H, Zhang X, Xue M, Zhang X (2015) Feeding of Whitefly on Tobacco Decreases Aphid Performance via Increased Salicylate Signaling. PLoS One 10:e0138584. http://dx.plos.org/10.1371/journal.pone.0138584