

# Molecular defense responses to natural enemies determine seedling survival in a subtropical forest

Hui Shang<sup>1</sup>, Yunquan Wang<sup>1</sup>, Baocai Han<sup>1</sup>, Franca Bongers<sup>1</sup>, Xiang-Cheng Mi<sup>2</sup>, Lei Chen<sup>3</sup>, Yu Liang<sup>1</sup>, and Keping Ma<sup>1</sup>

<sup>1</sup>Institute of Botany, Chinese Academy of Sciences

<sup>2</sup>Institute of Botany Chinese Academy of Sciences

<sup>3</sup>Affiliation not available

November 25, 2021

## Abstract

Negative density dependence (NDD) has been accepted as a key mechanism for biodiversity maintenance in natural forests and different lineages of natural enemies (fungus, bacterium, insect and virus) may be involved. Previous NDD related studies usually correlated seedling survival to the density of host-specific pests or pathogens along the distance to conspecific neighbors, and molecular defense responses of focal seedlings to natural enemies were seldomly concerned. By employing community functional genomics strategy, we extracted copy numbers of homologous genes in defense responses from transcriptomic data of 99 tree species and their inherent impacts on seedling survival were evaluated using partial linear regression analysis and general linear mixed-effects models. The community-level transcriptomic gene copy number of defense responses to fungus, insect and virus showed significant negative correlations with survival rates of the seedling community and the species-level gene copy number of defense response to insect significantly correlated with survival rates of top-twenty common seedling species. Moreover, presence of adult neighbors with distinct defense response to bacterial and viral pathogens survival of focal seedlings as predicted by NDD, while presence of seedling neighbors with similar defense response to insect tended to promote survival of focal seedlings which may be driven by insect-mediated biotic filtering or competitive exclusion. We conclude that both gene copy number and dissimilarities to adult and seedling neighbors in defense response to natural enemies determined seedling survival, indicating the critical contributions of molecular defense responses of plants to species coexistence and diversity maintenance in subtropical forests.

## Molecular defense responses to natural enemies determine seedling survival in a subtropical forest

Running title: Seedling molecular defenses drive survival

Hui Shang<sup>1,2,3</sup>, Yunquan Wang<sup>1,4</sup>, Baocai Han<sup>1,5</sup>, Franca J. Bongers<sup>1</sup>, Xiangcheng Mi<sup>1</sup>, Lei Chen<sup>1</sup>, Yu Liang<sup>1,\*</sup>, Keping Ma<sup>1,\*</sup>

1. State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

3. Shanghai Chenshan Plant Science Research Center, Chinese Academy of Sciences; Shanghai Chenshan Botanical Garden, Shanghai 201602, China

4. College of Chemistry and Life Sciences, Zhejiang Normal University, 688 Yingbin Rd., Jinhua 321004, China

5. State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China

\*Correspondence authors at: State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Beijing 100093, China. Email address: coolrain@ibcas.ac.cn (Yu Liang), kpma@ibcas.ac.cn (Keping Ma).

## Abstract

Negative density dependence (NDD) has been accepted as a key mechanism for biodiversity maintenance in natural forests and different lineages of natural enemies (fungus, bacterium, insect and virus) may be involved. Previous NDD related studies usually correlated seedling survival to the density of host-specific pests or pathogens along the distance to conspecific neighbors, and molecular defense responses of focal seedlings to natural enemies were seldomly concerned. By employing community functional genomics strategy, we extracted copy numbers of homologous genes in defense responses from transcriptomic data of 99 tree species and their inherent impacts on seedling survival were evaluated using partial linear regression analysis and general linear mixed-effects models. The community-level transcriptomic gene copy number of defense responses to fungus, insect and virus showed significant negative correlations with survival rates of the seedling community and the species-level gene copy number of defense response to insect significantly correlated with survival rates of top-twenty common seedling species. Moreover, presence of adult neighbors with distinct defense response to bacterial and viral pathogens survival of focal seedlings as predicted by NDD, while presence of seedling neighbors with similar defense response to insect tended to promote survival of focal seedlings which may be driven by insect-mediated biotic filtering or competitive exclusion.

We conclude that both gene copy number and dissimilarities to adult and seedling neighbors in defense response to natural enemies determined seedling survival, indicating the critical contributions of molecular defense responses of plants to species coexistence and diversity maintenance in subtropical forests.

**Keywords:** community genomics; fitness difference; gene annotation; niche differentiation; pathogen and pest; RNA-seq

## Introduction

One of the most enduring challenges for community ecology is how species diversity maintains in natural communities (Sutherland et al. 2013). Many mechanisms have been proposed to explain species coexistence and community assembly, such as negative density dependence (NDD), environmental filtering, competitive exclusion, and facilitation (Chesson & Peter 2000, Webb et al. 2006, Adler et al. 2007, HilleRisLambers et al. 2012, Fichtner et al. 2017, Adler et al. 2018).

Among these various species coexistence mechanisms, NDD is a widely accepted explanation for the maintenance of forest biodiversity where seedlings suffer negative effects from host specialized natural enemies (pathogen and pest) accumulated by local conspecific or phylogenetically related neighbors (Janzen 1970, Connell 1971, Webb et al. 2006, Gripenberg et al. 2014, Zambrano et al. 2017, Forrister et al. 2019, Song et al. 2021). Multiple lineages of natural enemies, such as fungal, bacterial and viral pathogens together with herbivorous mammals and insects, may contribute to NDD (Hazelwood et al. 2021, Song & Corlett 2021). The contributions of fungal pathogens and herbivores to NDD have been documented in many studies, while the contributions of bacterial pathogens and plant virus were seldom evaluated (Bagchi et al. 2014, Gripenberg et al. 2014, Hazelwood et al. 2021, Song & Corlett 2021). Moreover, most previous researches focused mainly on single type of natural enemy, and studies simultaneously demonstrating the contribution of multiple natural enemies are still lacking (Hazelwood et al. 2021).

A recently developed community functional genomics strategy may help us to take an insight into the intrinsic mechanism of NDD from the plant side and consider multiple natural enemies simultaneously. Plants can develop multiple molecular responses to microbial pathogens and herbivores (Wiesner-Hanks & Nelson 2016, Aljibory & Chen 2018, Ramirez-Prado et al. 2018). The genomic variations involved in these defense responses of plant species may be related to plant-pathogen or plant-pest co-evolutionary history, reflect susceptibility

of plants to diseases and pests, and thereby determine growth and survival of individual plant species as well as the whole plant community (Żmieńko et al. 2014, Dolatabadian et al. 2017). The community functional genomics strategy seeks to quantify the similarity of non-model species via transcriptomic data using Gene Ontology (GO) annotation of functional genes (Swenson 2012, Swenson et al. 2017). This strategy can provide information of hundreds of defense related genes and also has the potential to provide insights into how different lineages of natural enemies affect plant community structure separately and simultaneously.

Copy number variations (CNVs) and phylogenetic similarity of homologous genes are two principal methods in community functional genomics to describe the gene expressional similarity of plant species in response to biotic and abiotic environments (Han et al. 2017, Zambrano et al. 2017). CNV is defined as copy numbers of homologous genes or size of gene family (Liu 2014). CNV has been discovered for decades (Bridges 1936, Ohno 1970), which is resulting from gene duplication, gain or loss, and recently recognized as a common type of genomic variations (Żmieńko et al. 2014, Dolatabadian et al. 2017). Gene duplication may induce functional changes of genes such as subfunctionalization, pseudogenization, and neofunctionalization under rapid evolutionary processes, or keep their identical function by purifying selection or concerted evolution (Zhang 2003, Magadum et al 2013). Moreover, the functional conservation of genes with high sequence similarity, can be even across distinct phylogenetic lineages (Kater et al 2006, Liu et al 2014). For example, the expression of *Arabidopsis* ortholog AtWRKY40 in barley also compromised basal resistance to powdery mildew, indicating functional conservation of related proteins across dicots and monocots (Liu et al 2014). The function-conserved gene copies can be easily detected based on their high levels of sequence similarity (Magadum et al 2013), and therefore CNVs based on sequence similarity may act as a useful tool in researches on molecular functional diversity of species and natural communities. CNVs seem likely to be involved in many biological processes and can be associated with some important phenotypic variations (Manolio et al. 2009, Bai et al. 2016). Based on a functional genetics research on rice, Bai et al. (2016) suggested that CNVs might be involved in plant resistance to insects. CNVs may also differ among species and has been implicated in ecological adaptation. Besides model species, we can also identify CNVs from non-model species and it had been used in several evolutionary and ecological studies on microbes (Schirmer et al. 2012, Roller et al. 2016, Nunan et al. 2020), butterflies (Seixas et al. 2021), polar bears (Rinker et al. 2019), and plant species (Zhang et al. 2019, Han et al. 2017). For example, Han et al. (2017) used CNVs of light-related GOs to test how differences of tree species in light utilization strategies affect seedling regeneration in a subtropical forest. The CNVs extracted from transcriptomic data, defined as functional CNVs, could be useful in functional ecological researches because functional CNVs could reflect not only the evolutionary history of species but also the ecological responses to biotic and abiotic environments.

There are also debates on how CNVs are involved in defense responses and plant performances (Katju & Bergthorsson 2013, Żmieńko et al. 2014, Dolatabadian et al. 2017). It is expected that high copy number of defense genes is advantageous for plants due to the enhanced resistance against natural enemies, while there might be a trade-off between the expression of defense genes and plant performance and the high cost in transcription and translation will in turn limit the copy number of resistance genes (Lin et al. 2013). Moreover, the copy number of defense genes might also be the co-evolutionary consequence with exposure chance to natural enemies, which may reflect pathogen profiles or herbivore range of host plants (Zhai et al. 2011).

In this study, we extracted CNVs of homologous genes annotated as defense response to four lineages of natural enemies (fungus, bacterium, insect and virus). We used partial linear regression analysis to reveal the effect of gene copy number on seedling survival at species- and community- level. And we also used generalized linear mixed-effects models (GLMM) to assess how the neighborhood dissimilarities in functional CNVs impact seedling survival. For these aims, we hypothesized that the high copy number of defense response to natural enemies is associated with high seedling survival (Hypothesis I), because increased gene copy number may enhance biotic resistance. As the functional CNVs of defense responses to natural enemies may be strongly linked to their pathogen profiles or herbivore range, we also hypothesized that seedlings surround by adult and seedling neighbors with high dissimilarity in defense-related CNVs should share fewer natural enemies and therefore have a high survival possibility (Hypothesis II).

## Materials and Methods

### Study site and seedling censuses

The study site is located in the Gutianshan 24-ha subtropical Forest Dynamic Plot (GTS FDP; 29.25°N, 118.12°E) with altitude varying from 446.3 to 714.9 m above sea level, in Zhejiang province, China. The climate there is seasonal with wet season from March to June and two dry seasons from July to August and from October to February (Lai et al. 2009). The plot is censused every five years for all trees with diameter at breast height (d.b.h.)  $\geq 1$  cm. The seedling survival dataset was from the seedling monitoring network which initiated in May 2006 in GTS FDP (Chen et al. 2010, Wang et al. 2020). The plot contained 169 census stations with three 1 m<sup>2</sup> seedling plots in each station. The first 130 stations were arranged by a stratified random design and the subsequent 39 stations were placed in naturally occurring tree fall gaps. For each plot of the 169 stations, all the seedlings (< 1 cm d.b.h.) were tagged, identified to species, and measured for height. The statuses (alive or dead) of all seedlings recorded in May from 2009 to 2015 were used in this study. This resulted in six one-year intervals for seedling survival. After matching the tree species with transcriptomic data (see next section below), we obtained 4009 seedlings from the 2009-2010 interval, 3802 seedlings from the 2010-2011 interval, 3660 seedlings from the 2011-2012 interval, 3239 seedlings from the year 2012-2013 interval, 3202 seedlings from the year 2013-2014 interval and 2576 seedlings from the year 2014-2015 interval, and a total of 7718 unique seedlings belonging to 81 woody species across the 2009–2015 period were used in our analysis. We also determined the survival rates of each station and the top 20 dominant seedling species with more than 100 recorded seedling individuals in GTS FDP.

### Functional CNVs extraction from transcriptome

The transcriptome data of 99 tree species (listed in Table S1) was from a research program of Han et al. (2017). Plant mRNAs were extracted from seedling leaves collected from GTS FDP with the sampling strategy as follow. Fully expanded leaves from three seedling individuals per species were sampled from each of the five main habitats (low valley, low ridge, mid-slope, high slope and high ridge) (Chen et al. 2010). While some rare species were only sampled from three individuals or three leaves from the only seedling individual. The samples were immediately frozen in liquid nitrogen in the field and then stored in a -80degC freezer before sequencing. The transcriptomes were sequenced on an Illumina HiSeq 2500 platform with 2x125 bp length reads and at least 6G clean data for each sample, *de novo* assembled by Trinity v2.2 without reference genome sequence (Grabherr et al. 2011) and annotated to GOs by the software Blast2GO using the UniProt database (The UniProt Consortium 2016). In this study, we focused on four GOs with terms “defense response to fungus” (GO: 0050832), “defense response to bacterium” (GO: 0042742), “defense response to insect” (GO: 0002213), and “defense response to virus” (GO: 0051607), which are involved in the defense response to four lineages of natural enemies. Based on the result of GO annotation, we picked out transcripts annotated by the four GOs for 99 well sequenced tree species and translated to protein sequences by TransDecoder and the Pfam database (Haas et al. 2013). For each GO, we did the all-by-all blast for the protein sequences set. Before clustering, the blast results were filtered with 0.4 hit fraction. And then homologous gene clusters were obtained by employing MCL software with 10<sup>-5</sup> for the e-value and 2.0 for inflation value. The steps from blasting to clustering were referring the pipeline of Yang & Smith (2014). At last, we counted the number of genes in each cluster for each species. This resulted in four matrices containing gene clusters in columns and 99 tree species in the rows (hereafter denoted as functional CNV matrices). To show the dissimilarity of functional CNV among species, a heat-map was drawn by two-way cluster with the heatmap package in R 4.0.2 (R core Team, 2020), by using six clusters with most gene clusters for each GO.

Before the calculations of functional CNV at species- and community-levels (by seedling station), the values in functional CNV matrices were standardized by dividing by the maximum of each cluster to limit the values in a range between 0 and 1. For each defense response GO, the gene copy number of each species was defined as the sum of the standardized values of all the clusters, and the gene copy number of each seedling station was defined as the averaged gene copy number of all individuals in that station.

## Neighborhood variables and neighborhood functional CNV dissimilarity

Based on the census data in 2010, we defined all the adult trees ([?] 1 cm d.b.h.) in a circle of 20 m radius around the focal seedling as its adult neighbors, and all seedlings in the same 1 m<sup>2</sup>seedling plot of the focal seedling as its seedling neighbors. Two measures of conspecific neighborhood density were calculated for each focal seedling: 1) the number of conspecific seedling neighbors (S\_con); and 2) the sum of basal area of conspecific adult neighbors (A\_con).

The species pairwise distance matrices were calculated using the functional CNV matrices by the method of Gower (1971) as the proxy of species dissimilarity using FD package in R4.0.2. We used Blomberg's  $k$  statistic to measure phylogenetic signal (Blomberg et al. 2003) for the first two principal coordinates of each species pairwise distance matrix using the picante package in R 4.0.2. The species pairwise distance matrices of each defense response GO were employed to calculate relative nearest taxon functional diversity (NTFd') as neighborhood functional CNV dissimilarity for each focal seedling by the methods of Webb et al. (2006) and Wang et al. (2020) for both seedling (S\_NTFd') and adult (A\_NTFd') neighbors. The Null distribution of minimum neighborhood dissimilarity was produced by shuffling the species names on the species pairwise distance matrix for 999 times. The Mean ( $M_{Null}$ ) and standard deviation ( $Sd_{Null}$ ) of the null distribution were calculated. The NTFd' index was generated from observed value of minimum neighborhood dissimilarity by subtracting  $M_{Null}$  and dividing by  $Sd_{Null}$ . A positive NTFd' value indicates that species are more functionally dissimilar (higher functional CNV diversity) and a negative NTFd' value indicates that species are more functionally similar (lower functional CNV diversity). All the four neighborhood variables of focus seedlings calculated for each year interval.

## Statistical analyses

In order to assess whether gene copy number could directly explain the performance differences of seedlings, partial regression analysis based on linear mixed-effects models (LMMs) were used to identify the effects of gene copy number of each defense response on seedling survival rate at species level and community level using the lme4 package in R 4.0.2 (Bates et al. 2015). For the model at species level, gene copy number of four defense responses was set as fixed effect and the year interval as a random effect to account for interannual change. For the model at community level, the habitat type of seedling stations was also added as a random effect to account for habitat heterogeneity. The Akaike's Information Criterion (AIC) values were used to determine the optimal model in stepwise backward multiple regression. The partial effect ( $r_{[?]}$ ) as described in Grace et al. (2016) was evaluated and shown in multivariate partial plots.

The generalized linear mixed-effects models (GLMMs) were used to evaluate the effects of functional CNV dissimilarities of seedling and adult neighbors on the survival of focal seedlings. GLMMs with binomial errors were established for each defense response separately using the lme4 package in R 4.0.2. The annual survival (alive or dead in May of second year) of seedlings in all the six year intervals (2009 to 2015) was the response variable. Eight candidate models were developed with different factors as fixed effects for each defense responses (Table 1). The log-transformation of initial seedling height (logHT) was added to all the models to exclude its effect on seedling survival. S\_con and A\_con were set as fixed effects to evaluate the effects of the density of the local conspecific neighbors, while relative nearest taxon functional diversity of seedling (S\_NTFd') and adult (A\_NTFd') neighbors which were calculated as neighborhood functional CNV dissimilarity were set as fixed effects to evaluate the effect of heterospecific neighborhood functional CNV dissimilarity on defense response. Before being added to models, these four variables (S\_con, A\_con, S\_NTFd' and A\_NTFd') were standardized by subtracting the mean of observed value and dividing by one standard deviation. Species and year interval were added as the random effects to account for the inherent difference of each species and interannual change in response to neighbors. Plot and station were also included in the models as nested random effects to account for spatial autocorrelation. The AIC values were used to identify the best-fit models for each defense responses.

## Results

### Defense response GOs to four linkages of natural enemies

We totally obtained 14471, 23949, 877, and 4797 genes from 99 tree species annotated as the four GOs (defense response to fungus, bacterium, insect and virus) respectively. For each GO, we obtained 129, 216, 6 and 45 homologous gene clusters contained in at least ten species. A chord diagram was drawn to show the mean copy number of the four GOs in top ten families with high seedling density (Fig 1). There were 51 clusters shared by GOs defense response to fungus and bacterium, while 78 (60.5%) and 161 (74.5%) clusters were specific to the two GOs respectively (Fig. 2). By comparison, the GO defense response to bacterium shared only four clusters with GO defense response to virus and only one cluster with GO defense response to insect. Moreover, The GOs defense response to fungus, bacterium and virus also share only one cluster (Fig.2).

Among these homologous gene clusters related to defense responses, we found kinases acting as sentinels in signaling pathway in pathogen resistance, transcription factors regulating plant immunity, enzymes in biosynthesis or RNA silencing, one transmembrane protein, and one ribonucleoprotein in defense response to virus (Table S2). Four out of the top six gene clusters were shared between defense response to fungus and bacterium, including three transcription factors (*MYB44*, *WRKY33*, and *WRKY40*) and one receptor kinase (*ERECTA*). *FER* and *RD21A* were also found involved in defense response to fungal pathogens. *PBS1* and *SNC1* with the function of protein kinase and disease resistance protein respectively, were in the top six clusters in defense responses to bacterium. As for defense response to insect, two gene clusters with the function of transcription factor (*CAF1-9* and *ERF18*) associated with rapid responses to insects and four enzymes (*PAD4*, *LIP2*, *CYP81D11*, and *CYP94B3*) involved in plant hormone jasmonate (JA) related defenses or lipase modulating leaf senescence were found. Among the gene clusters associated with defense response to virus, two gene clusters (*AGO2* and *DCL3*) have the function of antiviral RNA silencing, and *SAMHD1* homolog may act as a host restriction factor.

The heap map showed that the functional CNV dissimilarity of defense responses was lineage specific, and species in the same genus or family were always clustered together (Fig. S1). At least one of the first two principal coordinates of functional CNVs species pair-wised distance matrix for each defense response showed significant phylogenetic signal (Table S3).

### Effect of Gene copy numbers of defense responses on seedling survival

The optimal model at species level showed that only gene copy number of defense response to insect influenced seedling survival rate with a negative effect ( $r = -0.28$ ,  $P = 0.002$ , Fig. 3A). The seedling survival rate at community level was best described by the model including gene copy number of defense response to fungus, insect and virus. The gene copy numbers of the defense response to fungus, insect and virus have significantly negative effects (fungus:  $r_{[?]} = -0.07$ ,  $P = 0.021$ ; insect:  $r_{[?]} = -0.11$ ,  $P < 0.001$ ; virus:  $r_{[?]} = -0.08$ ,  $P = 0.010$ ) on seedling survival rate (Fig. 3B).

### Effect of neighborhood functional CNV dissimilarity on seedling survival

The AIC values were used to compare models for each defense response (Table 1 & S4) and the values of the estimates of the fixed neighborhood effects of the best described models for each GO were shown in Figure 3. Across the best models for all four enemy types, log-transformation of initial seedling height (logHT) has a positive effect and conspecific seedling density (S\_con) has a negative effect on seedling survival significantly, while conspecific adult neighbors have no significant effect (Table 1, Fig. 4, as logHT had a much larger effect size than all other variables, it was not shown in Fig3). When considering defense response to fungus, seedling survival was best explained by model I (Table 1), containing only logHT and S\_con. When considering defense response to bacterium and virus, seedling survival was best explained by models VII (Table 1), and the adult NTFd' had a positive effect on seedling survival ( $P = 0.022$ , and  $P = 0.030$  for defense response to bacterium and virus, respectively). When considering defense response to insect, the best described model was model V, and the seedling NTFd' strongly negatively affected seedling survival ( $P = 0.0097$ ).

## Discussion

To evaluate NDD mediated by multiple natural enemies, we evaluated the effects of functional CNVs of defense response extracted from transcriptome on seedling survival. The gene copy numbers of defense response to fungus, insect and virus are significantly and negatively associated with survival rates, which does not support the Hypothesis I. The Hypothesis II, that neighborhood functional CNV dissimilarity increase seedling survival, was verified in defense response to bacterium and virus, but we gain an opposite effect in defense response to insect. Below the specific results linked to each hypothesis are discussed.

### **Gene copy numbers of defense responses affect seedling survival**

The result of partial linear regression analysis showed that at species level the copy numbers in defense response to insect have a negative effect on survival rates, and at community level (seedling stations) the copy numbers of defense response to fungus, insect and virus showed a negative effect on seedling survival. These results are inconsistent with some previous studies that showed that resistance was enhanced as gene copy number increased (Slabaugh et al. 2003, Bradeen et al. 2009). However, low gene copy numbers of biotic resistance are also found in some plant species (Lin et al. 2013, Xue et al. 2020), which may help them to gain evolutionary advantages due to a trade-off effect between cost on biotic resistance and growth performance. Although high copy number of defense gene is expected to be advantageous for better resistance against pathogens, the copy number is also limited to balance biological cost including not only energy for transcription and translation but also their toxic effect (Lin et al. 2013). Another potential explanation is that the low gene copy number within the GO in a local community may be due to the low presence of attacks from natural enemies. This hypothesis was supported by a recent genomic research on balsam poplar, which showed that a lower pathogenic pressure resulted in a lower defense gene copy number (Prunier et al. 2019). In the evolutionary process, plant-pathogen and plant-herbivore interactions may influence gene gain-and-loss and gene copy number of plant species (Liu 2014) and the copy number of defense related genes can act as an indicator of the abundance of natural enemies (Zhai et al. 2011). In our study, all tree seedlings within one community are exposed to the same putative natural enemy pool, which means that a species with lower copy number of defense genes may be less attacked by natural enemies and thus has a higher survival rate.

Moreover, this result also indicated that tree seedlings seemed more susceptible to gene copy numbers of defense response to insect rather than to other natural enemies, suggesting that the pressure from insect pests could be a key driver in structuring seeding community.

### **Conspecific NDD and neighborhood functional CNV dissimilarity in defense affect seedling survival**

A significant NDD effect from con-specific seedlings was detected in our study, which agrees with the result of Wang et al. (2020) in the same forest. The significant NDD in the subtropical forest infers a high intra-specific competition because of the increased seedling density in these years after an extreme winter storm (Wang et al. 2020). Additionally, conspecific neighboring adults had no significant effect on seedling survival, which can be explained by storm-induced decrease in tree neighbor density (Man et al. 2011, Wang et al. 2020).

In theory, all the four lineages of natural enemies can mediate negative density dependence (NDD) caused by Janzen–Connell mechanisms (Song et al. 2021). Many studies so far have focused on fungal pathogens rather than other natural enemy types and it is suggested that fungal pathogen is the key driver of NDD (Bell et al. 2006, Liang et al. 2016, Song & Corlett 2021). However, we didn't detect significant effects of functional CNV dissimilarity on defense to fungus in the present study. There may be two potential reasons for this. Firstly, the transcriptome extracted from the leaves cannot reflect the defense responses to root-associated pathogens, which play a vital role in structuring plant community (Chen et al. 2019, Tedersoo et al. 2020). Secondly, a counteraction effect may occur since both fungal pathogens and mutualistic endophytes exist simultaneously in seedlings (Chen et al. 2019).

Our results also show, for the first time, that pressure from bacterial and viral pathogens may also be a key driver of NDD for seedlings in subtropical forest. Functional CNV dissimilarity to adult neighbors in

defense response to bacterium showed a significant and positive effect on seedling survival, which means that seedlings with distinct defense response to bacterial pathogens from adult neighbors have higher probability of survival as predicted by NDD. This pattern is similar to the results of Zambrano et al. (2017) who also found that neighborhood dissimilarity of several defense genes have positive effects on individual survival. Our result suggests that the functional CNV dissimilarity between plants species can well reflect the niche differentiation in response to specialized bacteria. In other words, species with similar defense responses to bacterium may have similar bacterial pathogen profiles and thus seedlings may be killed by bacterial pathogen accumulated by adult neighbors with similar defense responses to bacterium. Moreover, since the defense response to bacterium has significant phylogenetic signal, this result may also explain conspecific and phylogeny related NDD. At last, this result is reasonable because many bacterial pathogens can infect plant leaves and are fatal to seedlings (Xin et al. 2018) and our transcriptomic data from seedling leaves can well reflect the interaction between seedlings and phyllosphere bacteria.

The Functional CNV dissimilarity to adult neighbors in defense response to virus also had a significant and positive influence on seedling survival. In other words, seedlings growing near adult neighbor with distinct defense response to virus survive more, similar to the NDD pattern in defense response to bacterium. While viral pathogens infect almost all plants in terrestrial ecosystems (Lefeuvre et al. 2019), plant viruses also have high host specificity. For example, a study on host range of 29 plant virus species indicated that 69% viruses were restricted to a single plant family (Garcia-Arenal & McDonald 2003). Due to the high host-specificity of plant virus, the positive effect of functional CNV dissimilarity in defense response to virus on seedling survival is also reasonable.

Although the NDD mediated by insect herbivores have been well documented in natural forests (Forrister et al. 2019) and manipulative experiment using insecticide (Jia et al. 2020), the contributions of insect herbivores to NDD are still debated (Bagchi et al. 2010, Gripenberg et al. 2014, Song & Corlett 2021). In this study we detected an opposite pattern that functional CNV dissimilarity to seedling neighbors in defense response to insect had a negative effect on seedling survival (Fig. 3), suggesting that seedling survival was higher when surrounded by neighbor seedlings with similar defense response to insect. This result is similar to previous studies showing that the survival of focal species increases in the presence of functionally similar or phylogenetic related neighbors, which could be explained by habitat preference (Godoy et al. 2014, Lebrija-Trejos et al. 2014, Wang et al. 2020). However, it seems unlikely the case in this study, because we found the neighbor seedlings rather than adults with similar anti-insect responses have a significant positive effect on seedling survival, while the habitat preference states that the habitats occupied by closely related adult neighbors might provide suitable resources and enhance seedling survival (Lebrija-Trejos et al. 2014). This indicates that insect-mediated biotic filtering rather than habitat preference may be involved in the mechanism underlying seedling dynamics that only specific seedlings beyond the host range of local insects survive. Another possible reason for this pattern is competitive exclusion driven by competitive ability rather than niche differences (Cavender-Bares et al. 2004, Mayfield & Levine 2010). During the competitive process, species with high competitive ability can displace the others and result in a clustering pattern (Mayfield & Levine 2010). We also found that the gene copy number of defense response to insect is negatively associated with seedling survival. This result, together with negative effect of functional CNV dissimilarity of defense insect on neighbor seedlings, suggests that the copy number of defense response to insect can reflect competitive ability of plant species and result in functional clustering of the seedling community. Taken together, it can be concluded that the competitive exclusion mechanism caused by the fitness difference of defense response to insect should be a key driver of seeding communities.

## Conclusion

Using functional CNVs, the impact of plant defense responses to natural enemies to species coexistence and diversity maintenance was found in a subtropical forest. Survival rate of seedling communities was negatively correlated with gene copy number of defense responses to fungus, insect and virus. And dominant species with lower gene copy number of defense responses to insect tended to have a higher survival rate. Here we found that high dissimilarities in gene copy numbers of defense responses to virus and bacterium of seedlings

to adult neighbors increased seedling survival. This can result in a functionally over-dispersed community through the negative density dependence process. Low dissimilarities in gene copy numbers of defense responses to insect to seedling neighbors also increased seedling survival. This can lead to a functionally clustered community via insect-mediated biotic filtering or competitive exclusion process. Overall, our analyses with transcriptomic data related to defense response to four lineages of enemies illustrate that the defense responses of plants to natural enemies can affect seedling survival.

### Author's contributions

K.M., Y.L., H.S., Y.W. and B.H. formulated the study and designed the methodology; B.H., L.C., X.M., Y.W. and H.S. collected the data; H.S., Y.W. and B.H. performed statistical analysis; H.S. wrote the initial manuscript with help from F.J.B. and Y.L.; All the authors contributed to revisions and gave final approval for publication.

### Acknowledgments

We are grateful for the support from Xiaojuan Liu, Jiangshan Lai, Ke Cao, Zhiqing Xue, Jianhui Ma, Rui Zhang, Hui Shen and Yuehong Yan for their helpful advices and kind assistance for data analysis. This work was funded by the Strategic Priority Research Program of Chinese Academy of Sciences (grant number XDB31000000), the National Key Research and Development Project of China (2017YFA0605103), and National Natural Science Foundation of China (31170495).

### Data accessibility

The functional CNV, seedling census and neighborhood dataset, will be available from the Dryad Digital Repository and also can be requested from the corresponding author on reasonable request.

### References

- Adler, P. B., Hillerislambers, J. & Levine, J. M. (2007). A niche for neutrality. *Ecology Letters*, **10** , 95-104. <https://doi.org/10.1111/j.1461-0248.2006.00996.x>
- Adler, P. B., Smull, D., Beard, K. H., Choi, R. T., Furniss, T., Kulmatiski, A., Meiners, J. M., Tredennick, A. T. & Veblen, K. E. (2018). Competition and coexistence in plant communities: intraspecific competition is stronger than interspecific competition. *Ecology Letters*, **21** , 1319-1329. <https://doi.org/10.1111/ele.13098>
- Aljibory, Z. & Chen, M.-S. (2018). Indirect plant defense against insect herbivores: a review. *Insect Science*, **25** , 2-23. <https://doi.org/10.1111/1744-7917.12436>
- Bagchi, R., Gallery, R. E., Gripenberg, S., Gurr, S. J., Narayan, L., Addis, C. E., Freckleton, R. P. & Lewis, O. T. (2014). Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature*, **506** , 85-88. <https://doi.org/10.1038/nature12911>
- Bagchi, R., Press, M. C. & Scholes, J. D. (2010). Evolutionary history and distance dependence control survival of dipterocarp seedlings. *Ecology Letters*, **13** , 51-59. <https://doi.org/10.1111/j.1461-0248.2009.01397.x>
- Bai, Z., Chen, J., Liao, Y., Wang, M., Liu, R., Ge, S., Wing, R. A. & Chen, M. (2016). The impact and origin of copy number variations in the *Oryza* species. *BMC Genomics*, **17** , 261. <https://doi.org/10.1186/s12864-016-2589-2>
- Bates, D., Machler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistics Software*, **67** , 51. <https://doi.org/10.18637/jss.v067.i01>
- Bell, T., Freckleton, R. P. & Lewis, O. T. (2006). Plant pathogens drive density-dependent seedling mortality in a tropical tree. *Ecology Letters*, **9** , 569-574. <https://doi.org/10.1111/j.1461-0248.2006.00905.x>
- Blomberg, S. P., Garland, J. R. T. & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, **57** , 717-745. <https://doi.org/10.1111/j.0014-3820.2003.tb00285.x>

- Bradeen, J. M., Iorizzo, M., Molloy, D. S., Raasch, J., Kramer, L. C., Millett, B. P., Austin-Phillips, S., Jiang, J. & Carputo, D. (2009). Higher copy numbers of the potato RB transgene correspond to enhanced transcript and late blight resistance levels. *Molecular Plant-Microbe Interactions*, **22** , 437-446. <https://doi.org/10.1094/mpmi-22-4-0437>
- Bridges, C. B. (1936). The bar "gene" a duplication. *Science* ,**83** (2148), 210-211. <https://doi.org/10.1126/science.83.2148.210>
- Cavender-Bares, J., Ackerly, D. D., Baum, D. A. & Bazzaz, F. A. (2004). Phylogenetic Overdispersion in Floridian Oak Communities. *The American Naturalist*, **163** , 823-843. <https://doi.org/10.1086/386375>
- Chen, L., Mi, X., Comita, L. S., Zhang, L., Ren, H. & Ma, K. (2010). Community-level consequences of density dependence and habitat association in a subtropical broad-leaved forest. *Ecology Letters*, **13** , 695-704. <https://doi.org/10.1111/j.1461-0248.2010.01468.x>
- Chen, L., Swenson, N. G., Ji, N., Mi, X., Ren, H., Guo, L. & Ma, K. (2019). Differential soil fungus accumulation and density dependence of trees in a subtropical forest. *Science*, **366** ,124-128. <https://doi.org/10.1126/science.aau1361>
- Chesson & Peter (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology & Systematics*, **31** ,343-366. <https://doi.org/10.1146/annurev.ecolsys.31.1.343>
- Connell, J. H. (1971). On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In P. J. Den Boer & G. R. Gradwell (Eds.), *Dynamics of Populations* (Vol. 298, pp. 298– 312). Centre for Agricultural Publishing and Documentation.
- Dolatabadian, A., Patel, D. A., Edwards, D. & Batley, J. (2017). Copy number variation and disease resistance in plants. *Theoretical and Applied Genetics*, **130** , 2479-2490. <https://doi.org/10.1007/s00122-017-2993-2>
- Fichtner, A., Hardtle, W., Li, Y., Bruelheide, H., Kunz, M., Von Oheimb, G. & Marshall, D. (2017). From competition to facilitation: how tree species respond to neighbourhood diversity. *Ecology Letters*,**20** , 892-900. <https://doi.org/10.1111/ele.12786>
- Forrister, D. L., Endara, M.-J., Younkin, G. C., Coley, P. D. & Kursar, T. A. (2019). Herbivores as drivers of negative density dependence in tropical forest saplings. *Science*, **363** ,1213-1216. <https://doi.org/10.1126/science.aau9460>
- Garcia-Arenal, F. & McDonald, B. A. (2003). An analysis of the durability of resistance to plant viruses. *Phytopathology*,**93** , 941-952. <https://doi.org/10.1094/phyto.2003.93.8.941>
- Godoy, O., Kraft, N. J. B. & Levine, J. M. (2014). Phylogenetic relatedness and the determinants of competitive outcomes. *Ecology Letters*, **17** , 836-844. <https://doi.org/10.1111/ele.12289>
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, **27** , 857-871. <https://doi.org/10.2307/2528823>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. & Zeng, Q. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29** ,644-652. <https://doi.org/10.1038/nbt.1883>
- Grace, J. B., Anderson, T. M., Seabloom, E. W., Borer, E. T., Adler, P. B., Harpole, W. S., Hautier, Y., Hillebrand, H., Lind, E. M., Partel, M., Bakker, J. D., Buckley, Y. M., Crawley, M. J., Damschen, E. I., Davies, K. F., Fay, P. A., Firn, J., Gruner, D. S., Hector, A., Knops, J. M. H., MacDougall, A. S., Melbourne, B. A., Morgan, J. W., Orrock, J. L., Prober, S. M. & Smith, M. D. (2016). Integrative modelling reveals mechanisms linking productivity and plant species richness. *Nature*, **529** , 390-393. <https://doi.org/10.1038/nature16524>
- Gripenberg, S., Bagchi, R., Gallery, R. E., Freckleton, R. P., Narayan, L. & Lewis, O. T. (2014). Testing for enemy-mediated density-dependence in the mortality of seedlings: field experiments with five Neotropical

tree species. *Oikos*, **123** , 185-193. <https://doi.org/10.1111/j.1600-0706.2013.00835.x>

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N. & Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8** , 1494-1512. <https://doi.org/10.1038/nprot.2013.084>

Han, B., Umana, M. N., Mi, X., Liu, X., Chen, L., Wang, Y., Liang, Y., Wei, W. & Ma, K. (2017). The role of transcriptomes linked with responses to light environment on seedling mortality in a subtropical forest, China. *Journal of Ecology*, **105** , 592-601. <https://doi.org/10.1111/1365-2745.12760>

Hazelwood, K., Beck, H. & Paine, C. E. T. (2021). Negative density dependence in the mortality and growth of tropical tree seedlings is strong, and primarily caused by fungal pathogens. *Journal of Ecology*, **109** , 1909-1918. <https://doi.org/10.1111/1365-2745.13615>

HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M. & Mayfield, M. M. (2012). Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review of Ecology, Evolution, and Systematics*, **43** , 227-248. <https://doi.org/10.1146/annurev-ecolsys-110411-160411>

Janzen, D. (1970). Herbivores and the number of tree species in tropical forests. *The American Naturalist*, **104** , 501-528. <https://doi.org/10.1086/282687>

Jia, S., Wang, X., Yuan, Z., Lin, F., Ye, J., Lin, G., Hao, Z. & Bagchi, R. (2020). Tree species traits affect which natural enemies drive the Janzen-Connell effect in a temperate forest. *Nature Communications*, **11** , 286. <https://doi.org/10.1038/s41467-019-14140-y>

Kater, M. M., Dreni, L., & Colombo, L. (2006). Functional conservation of MADS-box factors controlling floral organ identity in rice and *Arabidopsis* . *Journal of experimental botany* ,**57** (13), 3433-3444. <https://doi.org/10.1093/jxb/erl097>

Katju, V. & Bergthorsson, U. (2013). Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Frontiers in Genetics*, **4** , 273-273. <https://doi.org/10.3389/fgene.2013.00273>

Lai, J., Mi, X., Ren, H. & Ma, K. (2009). Species-habitat associations change in a subtropical forest of China. *Journal of Vegetation Science*, **20** , 415-423. <https://doi.org/10.1111/j.1654-1103.2009.01065.x>

Lebrija-Trejos, E., Wright, S. J., Hernandez, A. & Reich, P. B. (2014). Does relatedness matter? Phylogenetic density-dependent survival of seedlings in a tropical forest. *Ecology*, **95** ,940-951. <https://doi.org/10.1890/13-0623.1>

Lefeuvre, P., Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P. & Varsani, A. (2019). Evolution and ecology of plant viruses. *Nature Reviews Microbiology*, **17** , 632-644. <https://doi.org/10.1038/s41579-019-0232-3>

Liang, M., Liu, X., Gilbert, G. S., Zheng, Y., Luo, S., Huang, F. & Yu, S. (2016). Adult trees cause density-dependent mortality in conspecific seedlings by regulating the frequency of pathogenic soil fungi. *Ecology Letters*, **19** , 1448-1456. <https://doi.org/10.1111/ele.12694>

Liu, D., Leib, K., Zhao, P., Kogel, K. H., & Langen, G. (2014). Phylogenetic analysis of barley WRKY proteins and characterization of HvWRKY1 and -2 as repressors of the pathogen-inducible gene HvGER4c. *Molecular genetics and genomics* , **289** (6), 1331-1345. <https://doi.org/10.1007/s00438-014-0893-6>

Lin, X., Zhang, Y., Kuang, H. & Chen, J. (2013). Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. *BMC Genomics*,**14** , 335. <https://doi.org/10.1186/1471-2164-14-335>

- Liu, X. (2014). Gene copy number variation among and within plant species. Eidgenossische Technische Hochschule Zurich. <https://doi.org/10.3929/ethz-a-010432389>
- Magadam, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of genetics* , **92** (1), 155-161. <https://doi.org/10.1007/s12041-013-0212-8>
- Man, X., Mi, X. & Ma, K. (2011). Effects of an ice strom on community structure of an evergreen broad-leaved forest in Gutianshan National Natural Reserve, Zhejiang Province. *Biodiversity Science*,**19** , 197-205. <https://doi.org/10.3724/SP.J.1003.2011.09220>
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A. & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*,**461** , 747-753. <https://doi.org/10.1038/nature08494>
- Mayfield, M. M. & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, **13** , 1085-1093. <https://doi.org/10.1111/j.1461-0248.2010.01509.x>
- Nunan, N., Schmidt, H., & Raynaud, X. (2020). The ecology of heterogeneity: soil bacterial communities and C dynamics. *Philosophical Transactions of the Royal Society B* ,**375** (1798), 20190249. <https://doi.org/10.1098/rstb.2019.0249>
- Ohno S. (1970) Evolution by gene duplication. Springer-Verlag, New York, USA.
- Prunier, J., Giguere, I., Ryan, N., Guy, R., Soolanayakanahally, R., Isabel, N., MacKay, J. & Porth, I. (2019). Gene copy number variations involved in balsam poplar (*Populus balsamifera* L.) adaptive variations. *Molecular Ecology*, **28** , 1476-1490. <https://doi.org/10.1111/mec.14836>
- Pu, X. & Jin, G. (2018). Conspecific and phylogenetic density-dependent survival differs across life stages in two temperate old-growth forests in Northeast China. *Forest Ecology & Management*, 424,95-104. <https://doi.org/10.1016/j.foreco.2018.04.055>
- Ramirez-Prado, J. S., Abulfaraj, A. A., Rayapuram, N., Benhamed, M. & Hirt, H. (2018). Plant immunity: from signaling to epigenetic control of defense. *Trends in Plant Science*, **23** , 833-844. <https://doi.org/10.1016/j.tplants.2018.06.004>
- R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://doi.org/10.1038/sj.hdy.6800737>
- Rinker, D. C., Specian, N. K., Zhao, S., & Gibbons, J. G. (2019). Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. *Proceedings of the National Academy of Sciences* , **116** (27), 13446-13451. <https://doi.org/10.1073/pnas.1901093116>
- Roller, B. R. K., Stoddard, S. F, Schmidt, T. M. (2016). Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nature Microbiology*, **1** (11), 1-7. <https://doi.org/10.1038/nmicrobiol.2016.160>
- Schirrmeister, B. E., Dalquen, D. A., Anisimova, M., & Bagheri, H. C. (2012). Gene copy number variation and its significance in cyanobacterial phylogeny. *BMC microbiology* , **12** (1), 1-15. <https://doi.org/10.1186/1471-2180-12-177>
- Seixas, F. A., Edelman, N. B., & Mallet, J. (2021). Synteny-based genome assembly for 16 species of Heliconius butterflies, and an assessment of structural variation across the genus. *Genome biology and evolution* , **13** (7), evab069. <https://doi.org/10.1093/gbe/evab069>
- Slabaugh, M. B., Yu, J.-K., Tang, S., Heesacker, A., Hu, X., Lu, G., Bidney, D., Han, F. & Knapp, S.J. (2003). Haplotyping and mapping a large cluster of downy mildew resistance gene candidates in sun-

- flower using multilocus intron fragment length polymorphisms. *Plant Biotechnology Journal*, **1** , 167-185. <https://doi.org/10.1046/j.1467-7652.2003.00016.x>
- Song, X. & Corlett, R. T. (2021). Enemies mediate distance- and density-dependent mortality of tree seeds and seedlings: a meta-analysis of fungicide, insecticide and exclosure studies. *Proceedings of the Royal Society B: Biological Sciences*, **288** ,20202352. <https://doi.org/10.1098/rspb.2020.2352>
- Song, X., Lim, J. Y., Yang, J. & Luskin, M. S. (2021). When do Janzen-Connell effects matter? A phylogenetic meta-analysis of conspecific negative distance and density dependence experiments. *Ecology Letters*, **24** , 608-620. <https://doi.org/10.1111/ele.13665>
- Sutherland, W. J., Freckleton, R. P., Godfray, H. C. J., Beissinger, S. R., Benton, T., Cameron, D. D., Carmel, Y., Coomes, D. A., Coulson, T., Emmerson, M. C., Hails, R. S., Hays, G. C., Hodgson, D. J., Hutchings, M. J., Johnson, D., Jones, J. P. G., Keeling, M. J., Kokko, H., Kunin, W. E., Lambin, X., Lewis, O. T., Malhi, Y., Mieszkowska, N., Milner-Gulland, E. J., Norris, K., Phillimore, A. B., Purves, D. W., Reid, J. M., Reuman, D. C., Thompson, K., Travis, J. M. J., Turnbull, L. A., Wardle, D. A. & Wiegand, T. (2013). Identification of 100 fundamental ecological questions. *Journal of Ecology*, **101** , 58-67. <https://doi.org/10.1111/1365-2745.12025>
- Swenson, N. G. (2012). The functional ecology and diversity of tropical tree assemblages through space and time: from local to regional and from traits to transcriptomes. *International Scholarly Research Notices*, **2012** , 16. <https://doi.org/10.5402/2012/743617>
- Swenson, N. G., Iida, Y., Howe, R., Wolf, A., Umana, M. N., Petprakob, K., Turner, B. L. & Ma, K. (2017). Tree co-occurrence and transcriptomic response to drought. *Nature Communications*, **8** , 1996. <https://doi.org/10.1038/s41467-017-02034-w>
- Tedersoo, L., Bahram, M. & Zobel, M. (2020). How mycorrhizal associations drive plant population and community biology. *Science*, **367** , eaba1223. <https://doi.org/10.1126/science.aba1223>
- The UniProt Consortium (2016). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, **45** ,D158-D169. <https://doi.org/10.1093/nar/gkw1099>
- Wang, Y., Cadotte, M. W., Chen, J., Mi, X., Ren, H., Liu, X., Yu, M., Zhang, J. & Ma, K. (2020). Neighborhood interactions on seedling survival were greatly altered following an extreme winter storm. *Forest Ecology and Management*, **461** , 117940. <https://doi.org/10.1016/j.foreco.2020.117940>
- Webb, C. O., Gilbert, G. S. & Donoghue, M. J. (2006). Phylodiversity-dependent seedling mortality, size structure, and disease in a Bornean rain forest. *Ecology*, **87** , S123-S131. [https://doi.org/10.1890/0012-9658\(2006\)87\[123:PSMSSA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[123:PSMSSA]2.0.CO;2)
- Wiesner-Hanks, T. & Nelson, R. (2016). Multiple Disease Resistance in Plants. *Annual Review of Phytopathology*, **54** ,229-252. <https://doi.org/10.1146/annurev-phyto-080615-100037>
- Xin, X.-F., Kvitko, B. & He, S. Y. (2018). *Pseudomonas syringae* : what it takes to be a pathogen. *Nature Reviews Microbiology*, **16** , 316-328. <https://doi.org/10.1038/nrmicro.2018.17>
- Xue, J. Y., Zhao, T., Liu, Y., Zhang, Y. X. & Shao, Z. Q. (2020). Genome- Wide Analysis of the Nucleotide Binding Site Leucine-Rich Repeat Genes of Four Orchids Revealed Extremely Low Numbers of Disease Resistance Genes. *Frontiers in Genetics*, **10** ,1286. <https://doi.org/10.3389/fgene.2019.01286>
- Yang, Y. & Smith, S. A. (2014). Orthology Inference in Nonmodel Organisms Using Transcriptomes and Low-Coverage Genomes: Improving Accuracy and Matrix Occupancy for Phylogenomics. *Molecular Biology and Evolution*, **31** , 3081-3092. <https://doi.org/10.1093/molbev/msu245>
- Zambrano, J., Iida, Y., Howe, R., Lin, L., Umana, M. N., Wolf, A., Worthy, S. J. & Swenson, N. G. (2017). Neighbourhood defence gene similarity effects on tree performance: a community transcriptomic approach. *Journal of Ecology*, **105** , 616-626. <https://doi.org/10.1111/1365-2745.12765>

Zhai, J., Jeong, D. H., Paoli, E. D., Park, S., Rosen, B. D., Li, Y., Gonzalez, A. J., Yan, Z., Kitto, S. L. & Grusak, M. A. (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes & Development*, **25** , 2540-2553. <https://doi.org/10.1101/gad.177527.111>

Zhang, J. (2003) Evolution by gene duplication: an update. *Trends in ecology & evolution* ,**18** , 192–198. [https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8)

Zhang, J., Liu, L., Shu, J.-P., Jin, D.-M., Shen, H., Chen, H.-F., Zhang, R. & Yan, Y.-H. (2019). Transcriptomic Evidence of Adaptive Evolution of the Epiphytic Fern *Asplenium nidus* .*International Journal of Genomics*, **2019** , 1-9. <https://doi.org/10.1155/2019/1429316>

Żmieńko, A., Samelak, A., Kozłowski, P. & Figlerowicz, M. (2014). Copy number polymorphism in plant genomes. *Theoretical and Applied Genetics*, **127** , 1-18. <https://doi.org/10.1007/s00122-013-2177-7>

Table 1 Akaike information criterion (AIC) values of the generalized linear mixed models of seedling survival for each GO.

	AIC of models considering each defense GO	AIC of models considering each def
Model	Fungus	Bacteria
Null: SUR~logHT+	17472.34	17472.34
I: SUR~logHT +S_con++	<b>17465.30¶</b>	17465.30
II: SUR~logHT+ A_con++	17474.31	17474.31
III: SUR~logHT+ S_NTFd'§	17474.23	17474.27
IV: SUR~logHT+ A_NTFd'§	17474.34	17468.94
V: SUR~logHT+ S_con+S_NTFd'	<b>17467.11¶</b>	17467.24
VI: SUR~logHT+ A_con+ A_NTFd'	17476.31	17470.94
VII: SUR~logHT+ S_con+ A_NTFd'	<b>17467.30¶</b>	<b>17461.78¶</b>
Full	17470.88	17465.49

Note: +SUR is seedling survival. logHT is the log-transformed initial seedling height. ++S\_con is the number of con-specific seedlings and A\_con is basal area of conspecific adults. §A\_NTFd' is the nearest taxon functional distance for neighbor adults and S\_NTFd' is the nearest taxon functional distance for neighbor seedlings. A\_NTFd' and S\_NTFd' was calculated based on functional CNVs per defense responses separately. ¶Bold values indicate the models selected based on the lowest AIC value per defense response and bold grey values indicate the models with  $\Delta AIC \geq 2$  compared to the best-fit model.

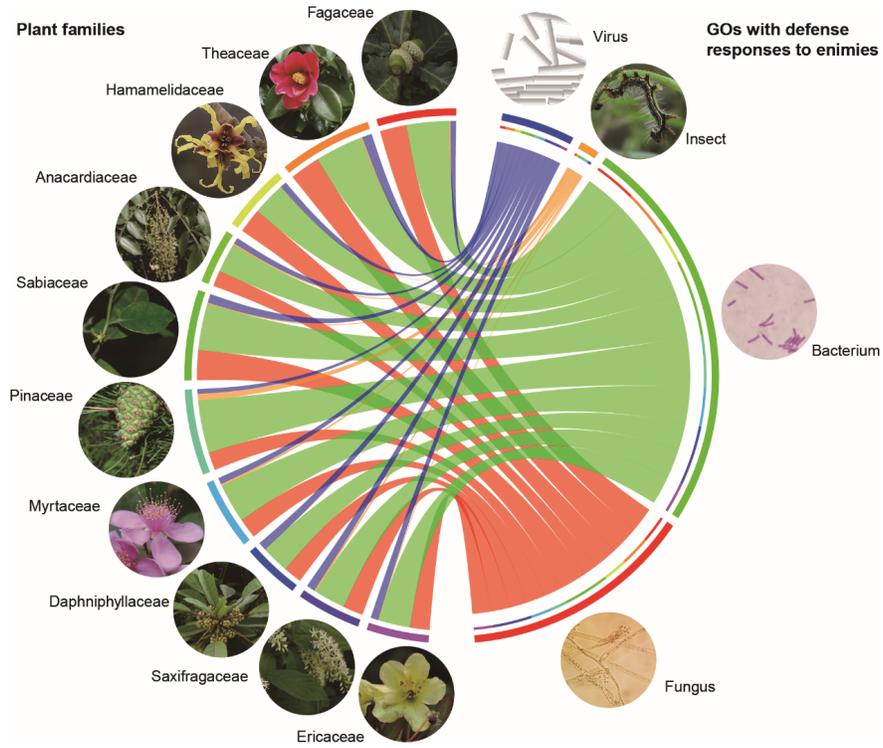


Figure 1 A chord diagram show the mean copy number of the four GOs (defense response to fungus, bacterium, insect and virus) in top ten families with most seedling density.

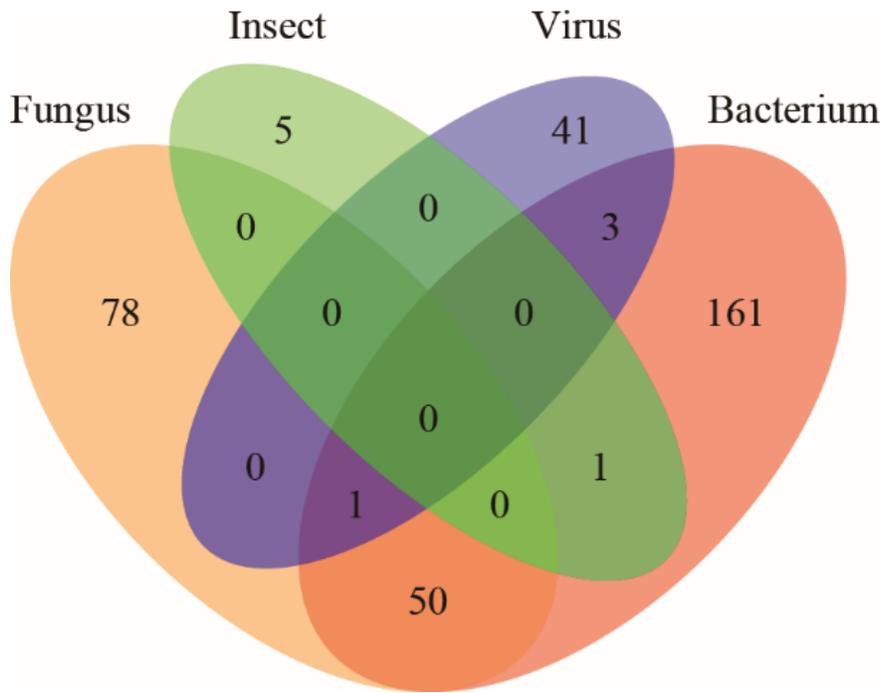


Figure 2 Venn diagram showing shared and unique clusters among each defense response GO.

## Hosted file

image3.emf available at <https://authorea.com/users/447966/articles/546867-molecular-defense-responses-to-natural-enemies-determine-seedling-survival-in-a-subtropical-forest>

Figure 3 The multivariate partial plots show predicted effects of gene copy number on seedling survival rate in species level (A) and community level (B) for the GOs defense response to fungus, bacterium, insect and virus, fitted by partial linear regression. Solid lines mean significant correlations and dash lines mean insignificant relationships. Shaded areas represent 95% confidence intervals. The *P*-values shown for each GO and ns means not significant.

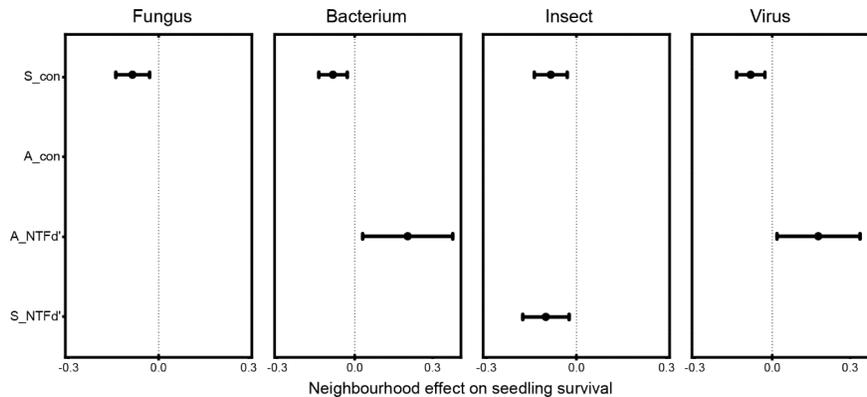


Figure 4 Estimates of fixed effects ( $\pm$  95% CI) for the best seedling survival model (See table 1) when considering each defense GO. Filled dots represent significant values ( $P < 0.05$ ). S\_con is the number of con-specific seedlings. A\_con is basal area of conspecific adults. A\_NTFd' is the nearest taxon functional distance for neighbor adults and S\_NTFd' is the nearest taxon functional distance for neighbor seedlings. A\_NTFd' and S\_NTFd' were calculated based on functional CNVs per defense responses separately.

## Supporting Information

Figure S1 two-way cluster heat-map show the functional CNV dissimilarity of 99 sequenced species in GTS FDP using gene copy numbers of top six clusters for each GO, which is defense response to fungus, bacterium, insect and virus. A more red color means higher copy numbers.

Table S1 tree species list in this study

Table S2 the homologous gene cluster information of top six clusters for each GOs.

Table S3 Measures of phylogenetic signal for gene copy number dissimilarity of four defense response GOs.

Table S4 results of the best-fit model and models with  $\Delta AIC[?]\geq 2$  as compared to the best-fit model for each defense response GO.