# The impact of defense hormones on the interaction between plants and the soil microbial community

Zhang, J.

**Author**: Zhang, J.
**Title**: The impact of defense hormones on the interaction between plants and the soil microbial community
**Issue Date**: 2021-05-04

# Chapter 4

# Activation of the SA-associated plant defense pathway alters the functions of soil microbial communities in four sequential generations

Jing Zhang[1*]

T. Martijn Bezemer[1, 2]

Peter G.L. Klinkhamer[1]

Klaas Vrieling[1]

1) Plant Science and Natural Products, Institute of Biology Leiden (IBL), Leiden University, Sylviusweg 72, 2333BE Leiden, The Netherlands

2) The Netherlands Institute of Ecology (NIOO-KNAW), Department of Terrestrial Ecology, Droevendaalsesteeg 10, 6708 PB, Wageningen, The Netherlands

**Abstract**

Systemic acquired resistance (SAR) is an immune response of plants that regulates plant hormonal signaling pathways and strengthens the ability of the plant to withstand pathogenic microbes. Aboveground application of salicylic acid (SA) to the plant can induce SAR and we showed that it mitigates negative effects of the soil microbial community on the performance of the plant *Jacobaea vulgaris*. How SA-induced resistance affects the expression of functional genes and gene ontology in the rhizosphere and how this phenomenon extends over multiple generations is not well studied. In this study, a meta-transcriptomics approach was used to characterize gene expression profiles of microbial communities in 24 soil samples of SA-treated and control plants over 4 generations. 71.6 million reads were used for de-novo assembly of the microbial transcriptome, after which a total of 1.3 million unique contigs (genes) were identified. Multivariate analysis revealed that the SA treatment, generation and the interaction between these two affected the functional genes of the rhizosphere microbial communities of *J. vulgaris*. In general, the effect of the SA treatment on microbial gene expression was lowest in the first generation and strongest in the fourth generation. Microbes in soil samples of SA-treated and control plants showed 1663 differentially expressed genes. In the first generation only two genes differed significantly in gene expression between microbes from soils of SA treated and control plants while in the fourth generation 361 genes were differentially expressed between microbes from soils of SA treated and control plants. None of the significantly expressed SA-downregulated genes were present in all four generations, while only one SA-upregulated gene was observed in all four generations. Gene ontology (GO) analysis showed that soil microbial communities in rhizosphere soil of SA-treated plants increased the expression of thirteen GO terms in the second, third and fourth generation. These increased GO terms were mostly related to viral RNA genome replication, to interactions with host cells, to organelles of the host cells and to RNA polymerase activities. There were six GO terms of which the expression decreased in the second, third and fourth generation, and these were associated with processing nitrogen and macromolecules. Overall, our results show that aboveground activation of defenses in the plant affects the expression of functional genes in the soil microbial communities belowground. This suggests that plants may recruit functional

rhizosphere microbiomes that improve plant health and crop production in agriculture.

**Keywords**
Meta-transcriptomics, Soil microbial community, Functional genes, Plant-soil interactions, Induced resistance, Rhizosphere soil, Salicylic acid

## Introduction

Plants can alter the microbiome of the soil in which they grow, and in turn, microorganisms can influence plant performance. The rhizosphere microbiome, defined as the microbial community established near or on plant roots, can have negative, positive and neutral effects on the growth of a host plant (Van Wees et al., 2008; Raaijmakers, et al., 2009; Berendsen et al., 2012). Microbes such as plant growth-promoting bacteria (PGPB) and arbuscular mycorrhizal fungi (AMF) are typically characterized as plant beneficial, because of their contribution to plant health and nutrient uptake (Jeffries et al., 2003; Compant et al., 2010). In contrast, pathogenic microbes typically reduce plant growth and trigger defense mechanisms in the plant (Pieterse et al., 2001). However, the overall net effect of soil microbial communities on plant growth is often negative (Nijjer et al., 2007; Wardle et al., 2011). This might be due to e.g. competition between plants and microbes for available nutrients or soil pathogens (Berendse, 1994; Callaway et al., 2004; Mazzoleni et al., 2015; Cesarano et al., 2017). In response, plants have evolved hormone-driven defensive strategies to suppress these pathogenic impacts, such as systemic acquired resistance (SAR) and induced systemic resistance (ISR) (Bruce and Pickett, 2007; Berendsen et al., 2012; Huang et al., 2014; Ökmen and Doehlemann, 2014).



Systemic acquired resistance (SAR) is a distinct transduction pathway, which is involved in the biological processes that enhance the plant's immune system and defense against microbial pathogens (Reymond and Farmer, 1998; Walters and Heil, 2007; Pieterse et al., 2014; Haney and Ausubel, 2015). An infection caused by a pathogenic microbe can induce SAR, in which plants enhance their immune system by expressing genes coding for pathogenic-proteins (PR) in infested and uninfected tissues (Kachroo and Robin, 2013; Shah and Zeier, 2013; Gao et al., 2015). Apart from local induction by pathogenic microbes, SAR can also be induced by foliar sprays of the phytohormone salicylic acid (SA) (Reymond and Farmer, 1998). Applying a low concentration of SA directly to leaf tissues results in the activation of SA signaling pathways and this has been considered an effective way to activate defense signals in many plant species (Reymond and Farmer, 1998; Pozo and Azcón-Aguilar, 2007; Vlot et al., 2009).

In Chapter 2, we showed that the application of SA mitigates the negative effects of soil microbes on the growth of *J. vulgaris* although this effect did not increase further in subsequent generations of plant growth. A number of studies have examined the expression of functional genes in soil microbial communities. For example, Xue et al. (2016) showed that changing the temperature of soil significantly altered the gene expression in soil microbial communities and these genes were related to maintaining carbon and nitrogen stability in the soil, resulting in higher plant growth. Moreover, Castro et al. (2019) recently demonstrated that plants can change the expression of functional genes (i.e., carbon metabolic genes) in the soil microbial community in response to environmental changes such as drought. Here we hypothesize that application of SA to plants can also cause changes in the expression of functional genes in the soil microbial community and we hypothesize that the altered gene expression is related to the suppression of soil microbial pathogens of plants (Maurhofer et al., 1998; Verberne et al., 2000; Tanaka et al., 2015). Moreover, we expect, that the gene expression difference in the rhizosphere microbial community of control and SA treated plants will increased over generations of plant growth.

Previously, we analyzed the changes in the composition of the microbial community in the rhizosphere soil upon foliar application with SA and showed that the composition of rhizosphere bacterial communities differed among four plant generations of *J. vulgaris* and between soils from SA treated and control plants. However, the composition differed strongly among generations (Chapter 3). Functions of the soil microbial community can be performed by different microbial taxa (Burke et al., 2011; Liu et al., 2018; Liu et al., 2020) and hence we expect that there is functional redundancy in the soil microbial community and a consistent effect of SA application on gene expression in the microbial community.

In this study we ask the following questions: (1) Does the application of SA on leaves of *J. vulgaris* significantly alter the gene expression of the microbial community in the rhizosphere? (2) Does the effect differ between generations or is there an interaction between the SA treatment and generation on the gene expression in the microbial communities? (3) Which groups of genes or gene ontology pathways in the rhizosphere microbiome are influenced by SA-application over generations?

## Materials, methods and bioinformatics processing

The multi-generation growth experiment with *J. vulgaris* has been described in Chapter 3. In short, *J. vulgaris* plants were grown for four generations on soils inoculated with soil from the previous generation from the same treatment with a foliar SA application treatment and a control treatment. Each treatment had 10 replicates. For each treatment, the three successively labeled replicates (No. 1, 2, 3, No. 4, 5, 6 and No. 7, 8, 9) were mixed and used as one pooled replicate, Hence, the three pooled replicates were used for RNA extraction for each treatment in each generation and a total of 24 soil samples were used for RNA extraction (3 replicates x 2 treatments x 4 generations). RNAseq was carried out using the Illumina platform.

Processing of the data included quality control of raw reads (FastQC), data trimming (Trimmomatic 0.39), filtering out ribosomal RNAs (SortMeRNA), de novo assembly of reads (Trinity), remove duplicates (CD-HIT-EST algorithm), mapping back to the transcriptome (Bowtie2). For a detailed description see Chapter 3. Gene ontology enrichment was performed using Trinotate and Goseq against the SwissProt, NR (non-redundant) and Pfam databases (Bryant et al., 2017; Bateman, 2019; El-Gebali et al., 2019).



## Statistical analyses

Prior to analysis, the raw data were normalized. TMM (trimmed mean of M-values) normalization was used for read counts among all 24 samples (Robinson and Oshlack, 2010). A principal component analysis (PCA) was employed using the normalized number of genes to examine the composition of rhizosphere soil samples of SA-treated and control plants for the four generations. A PERMANOVA test was performed using the *adonis* function (number of permutations = 999) in R within the "vegan" package to verify the effects of the SA treatment and time on the composition of all expressed genes. To compare similarities among samples of treatment SA and control over four generations, a Pearson correlation for pairwise sample comparison based on the normalized raw read counts of all replicates in the control and SA treatments was performed in R and a heatmap was produced.

Differential gene expression (DE) analysis was performed for all possible combinations of replicates of sets of 8 samples (2 treatments x 4 generations) with EdgeR with raw read counts as input. EdgeR normalizes the data to TMM before further processing. After DE analysis in EdgeR, for all differentially expressed genes of the 8 samples Volcano plots were made for the contrast between SA-treated and control samples per generation. Log2 (FC) values were used as x-variable and -log10 (FDR) for the y-variable to produce a volcano plot of differentially expressed genes between control and SA-treated soil samples per generation. Genes that were significantly differentially expressed between SA-treated and control soil samples that could be annotated were listed. A clustered heatmap based on Euclidean distances (Danielsson, 1980) of gene expression derived from EdgeR per treatment after Z-scored transformation was generated in R using the package “pheatmap” (Kolde and Kolde, 2015).

To visualize the gene expression changes among different hormonal treatments and time categories, an NMDS (nonmetric multidimensional scaling) plot using the Bray-Curtis index as a measure of dissimilarity was generated using TMM normalized read counts. To verify changes in the composition of the 1663 expressed genes due to the SA treatment and time effect, a PERMANOVA test was performed using the adonis function (number of permutations = 999) in R within the “vegan” package.



Gene ontology (GO) enrichment was performed with “GoSeq" for each generation separately. Gene functional classification was determined for three categories: biological processes, cellular components and molecular functions. GO terms affiliated to Eukaryotes (e.g. mitochondria) were removed. The rich factor was calculated as the number of differentially expressed genes in the ontology divided by the number of all genes that were used as a background gene list.

## Results

### Comparing read counts between generations and treatments

A total of 898,4 million raw sequencing reads were obtained from the 24 metatranscriptomic libraries. The details of the library size and basic information about read quality were described in Chapter 3. A principal component analysis (PCA) using log2 transformed normalized CPM showed that the read counts of contigs in the microbial community of rhizosphere soil of the *J. vulgaris* samples among generations were well separated (Fig. 1), this was in line with the permutation test (PERMANOVA $R^2 = 0.22$, $F = 19.6$, $df_1 = 3$, $df_2 > 999$, $p < 0.01$). In addition, the effect of SA application was significant (PERMANOVA $R^2 = 0.07$, $F = 6.3$, $df_1 = 1$, $df_2 > 999$, $p < 0.05$). Gene expression patterns of SA-treated *J. vulgaris* and control samples were better separated in the third and fourth generation than in the first and second generation (Fig. 1). In the correlation matrix for all sample replicates generated with PtR (a tool for comparing sample replicates in Trinity) (Fig. 1), samples within treatments were positively correlated with each other and also there was a positive correlation between samples within generations especially for the first generation. The heatmap showed clear clustering of treatments within generations except for generation 1. The separation between the SA and the control treatment became more distinct over generations (Fig. 2).



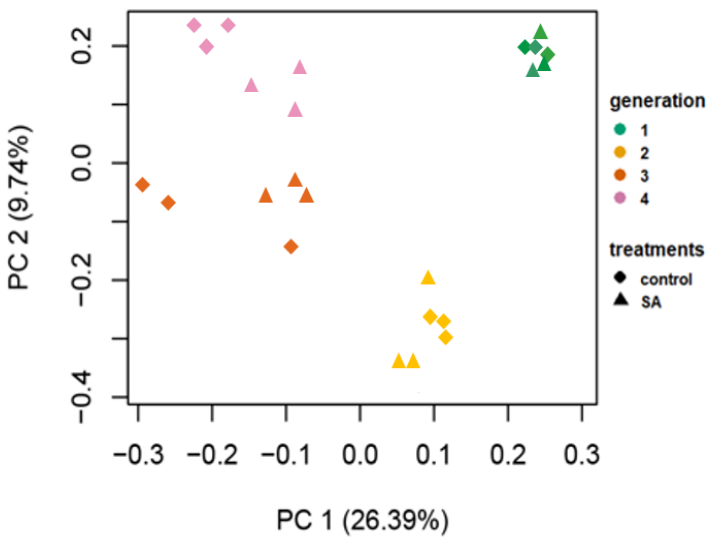



Fig. 1 Scatter plot from a principal component analysis (PCA) of TMM normalized CPM representing the overall gene expression patterns of different rhizosphere soil

samples of SA-treated and control *J. vulgaris* plants over generations. Shapes represent the treatments and colors represent generations.

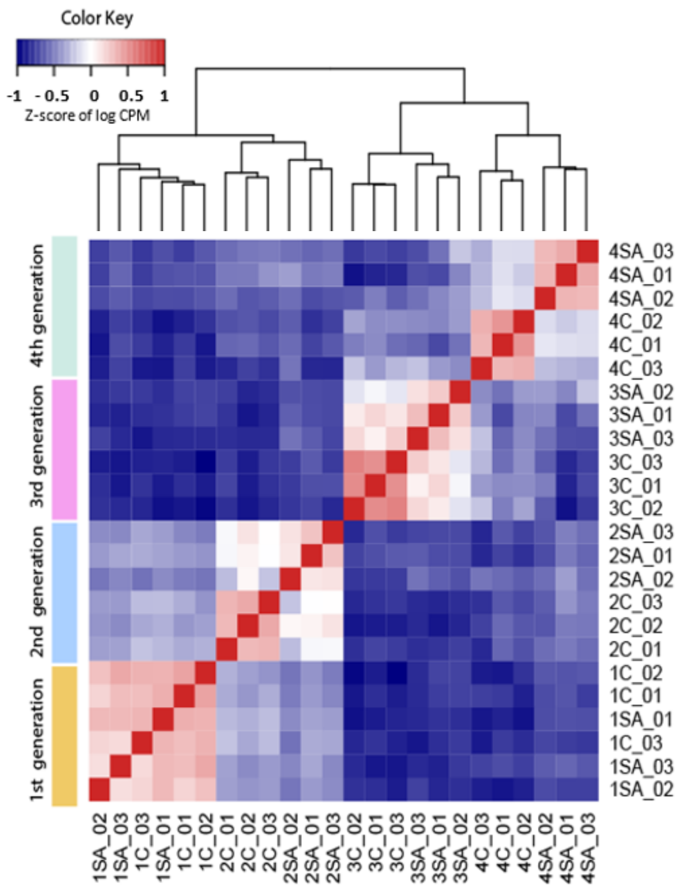



Fig. 2 Clustered heatmap visualizing the Pearson correlation matrix for pairwise sample comparisons based on TMM normalized read counts per million. The heat map shows the correlation in microbial gene expression in all paired replicates between rhizosphere soil samples of SA-treated and control *J. vulgaris* plants over four generations. The dendrogram illustrates the relationship-distance between samples and is calculated based on a Pearson correlation coefficient. The color key represents the z-score of log2 CPM. The legends on the sides represent: Generation (1-4), treatment (SA/Control) and replicate number (01-03).

4

**Differential gene expression**

In total, 0.36 million genes were detected. Of those genes, 1663 were differentially expressed between all possible combinations of replicates of sets of 8 samples (2 treatments x 4 generations). Hierarchical clustering on CPM for 1663 differentially expressed genes was performed to explore the patterns of gene expression of the microbial communities between all pairwise combinations of all the samples among SA and control treatments over four generations (Fig. 3). Except for the first generation, SA and control samples were separated from each other in different clusters (Fig. 3). However, among generations, different clusters of genes were differentially grouped. Differences were most pronounced between on the one hand, the first and second generation, and on the other hand, the third and fourth generation.

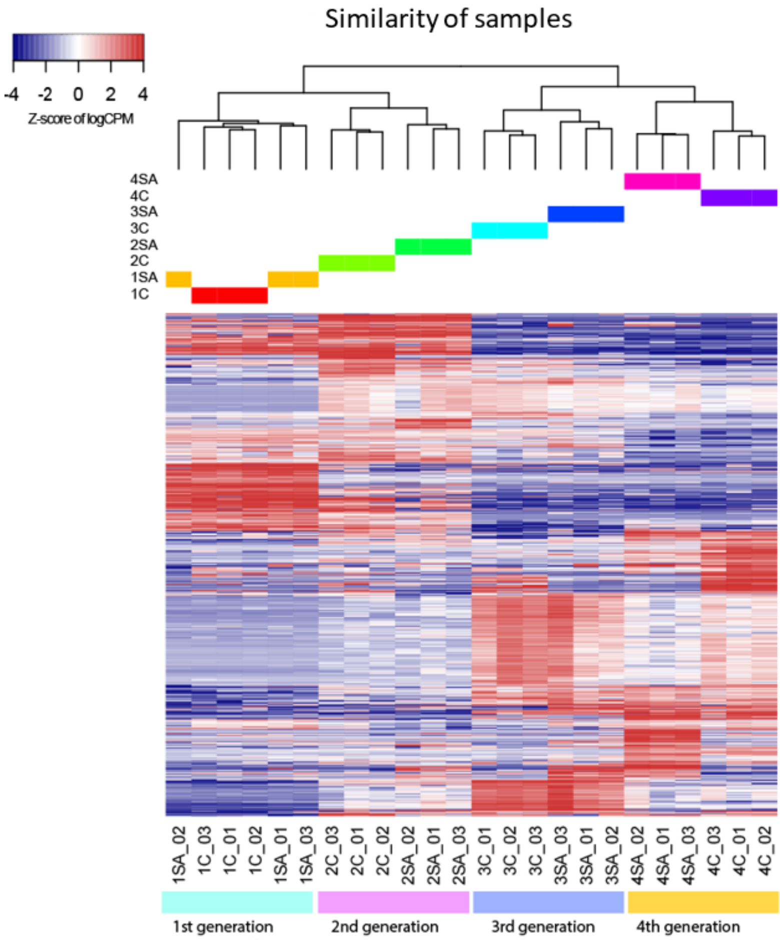



Fig. 3 Heatmap showing 1663 differentially expressed genes (FDR < 0.05 and fold change ≥ 2) between all possible combinations of replicates of rhizosphere soil samples of SA-treated and control *J. vulgaris* plants over four generations based on TMM normalized CPMs. The color key represents the z-score of log2 CPM. The dendrogram on the x-axis illustrates the hierarchal clustering of relationship-distance between replicates using TMM normalized log2-transformed CPM. The legend on the bottom represents: generation (1-4), treatment (SA/control) and replicate number (01-03).

The NMDS plot showed that the 1663 differentially expressed microbial genes detected with EdgeR were differentially expressed in the different generations

(Fig. 4, PERMANOVA $R^2 = 0.63$, $F = 21.8$, $df_1 = 3$, $df_2 = 1662$, $p < 0.01$) and also that genes were differently expressed between the SA treatment and the control (PERMANOVA $R^2 = 0.07$, $F = 7.0$, $df_1 = 1$, $df_2 = 1662$, $p < 0.01$). The effect of the SA treatment was not the same in each generation as indicated by the significant interaction (PERMANOVA $R^2 = 0.15$, $F = 5.2$, $df_1 = 3$, $df_2 = 1662$, $p < 0.001$).

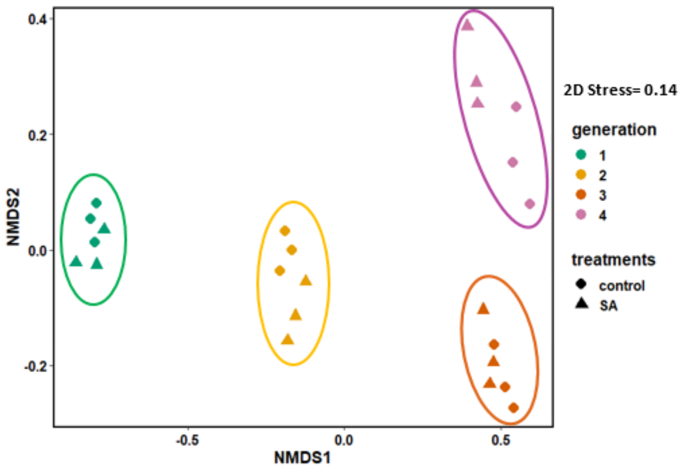





Fig. 4 Multivariate analysis of 1663 differentially expressed microbial genes between all replicates in rhizosphere samples from SA-treated and control *J. vulgaris* plants grown in four generations. Shown are sample scores from a nonmetric multidimensional scaling (NMDS) plot.

To identify the numbers of significantly down- or up-regulated genes in the SA treatment in each generation in the rhizosphere microbial community, volcano plots were made (Fig. 5). In the first generation, no downregulated genes were observed and only two upregulated genes were detected (Fig. 5a). This increased to 59 and 76 in the second, 89 and 26 in the third, and 187 and 174 in the fourth generation, respectively (Fig. 5b, c, d). Among all the significant differentially expressed genes, no genes were found that were downregulated after SA application in all four generations, while only one gene was observed that was upregulated in SA in all four generations (Fig. 6). Circa 90% of the genes that were significantly altered by the SA treatment could not be annotated. Among all the annotated genes, only two of the significant differentially expressed microbial genes were detected in three generations and eight genes were detected in two generations (Fig. 6). Not all the genes could be matched with a function in the database. Detailed information of successfully annotated genes was listed in Table S1.

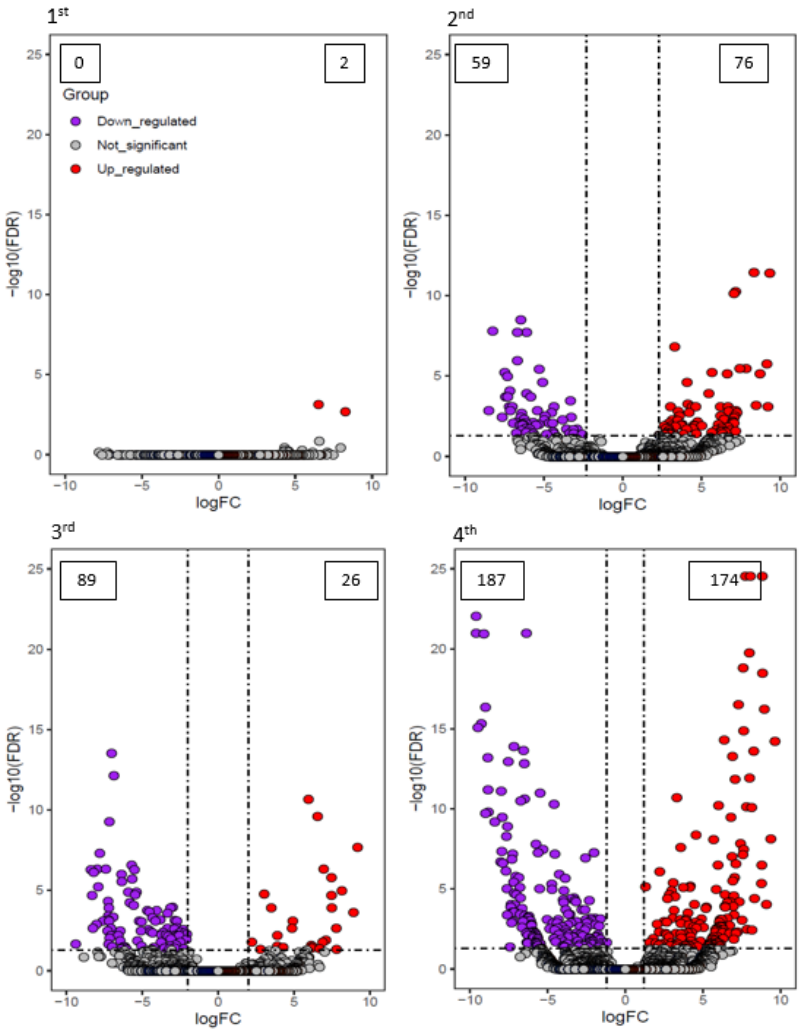



Fig. 5 Volcano plots of 1663 differentially expressed genes of the microbial community in rhizosphere samples of SA-treated and control *J. vulgaris* plants per generation. The x-axes show log2 fold changes of read counts of the genes of the SA treatment compared to thecontrol, and the y-axes show the -log10 adjusted for FDR

values. SA upregulated genes are presented in purple, and SA downregulated genes are displayed in red, while non-significant genes are shown as light grey dots. 1st, 2nd, 3rd and 4th represent the different generations. The numbers inside each box represent the number of significantly up/down-regulated genes. The two vertical dashed lines represent the positive or negative log2 fold changes in the number of readcounts in the SA treatment compared to the control in the generation when -log10(FDR) is 2 as presented by the horizontal dashed lines.

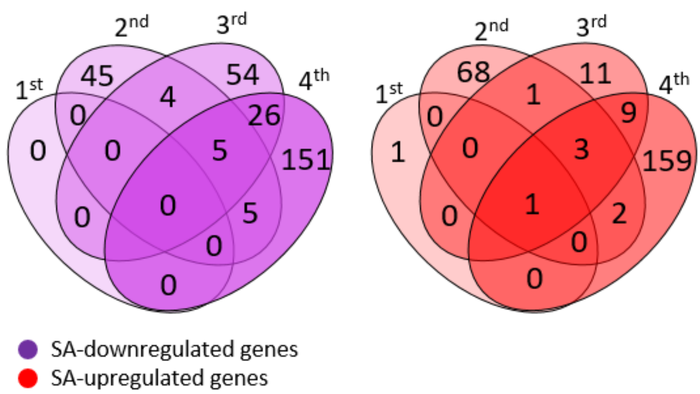


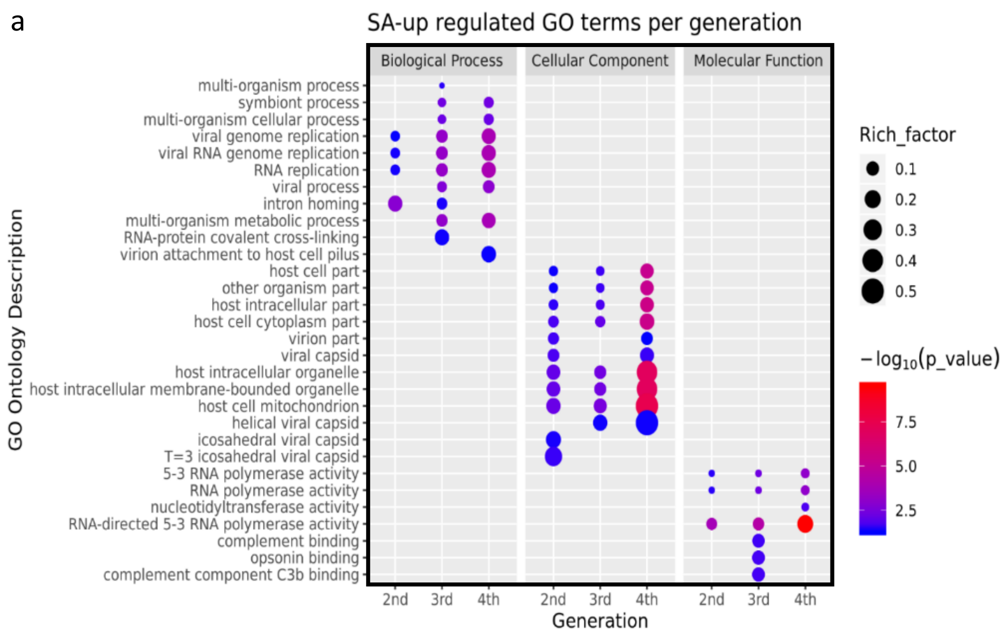



Fig. 6 Venn diagrams showing the number of shared and unique up and down-regulated microbial genes over generations in the rhizosphere of *J. vulgaris*. The numbers represent the significantly differently expressed genes from the volcano plot (Fig. 5). 1st, 2nd, 3rd and 4th represent the different generations.

**Gene ontology (GO) analysis**

To profile differentially expressed pathways, we performed a gene ontology (GO) analysis for the soil samples of SA-treated and control plants for each generation (Fig. 7, Table S2). No significantly upregulated or downregulated GO terms were observed in the first generation (Fig. 7a, 7b). In the second, third and fourth generations, genes from classes of the GO categories "biological processes", "cellular components" and "molecular functions" were differentially expressed (Fig. 7a). 13 GO terms were upregulated in the SA treatment in three generations, while 18 GO terms were upregulated in one or two generations (Fig. 7a). Of the 13 GO terms upregulated in three generations three belonged to the GO category "biological processes", and these GO terms were all related to viral RNA genome replication, seven belonged to the GO category "cellular components" and these GO terms were related to interactions with host cells and to organelles of the host cells and finally three belonged to the GO category "molecular function" and these GO terms were all related to RNA polymerase activity.

Only six GO terms were downregulated in the second, third and fourth generation in the rhizosphere of SA treated plants, while 58 GO terms were downregulated in one or two generations only (Fig. 7b). The six GO terms downregulated in three generations fell all in the GO category "biological processes" and the GO terms were related the localization of processes, to nitrogen processing and to processes involving macromolecules. None of GO terms involved in cellular components and molecular functions were present in these three generations.



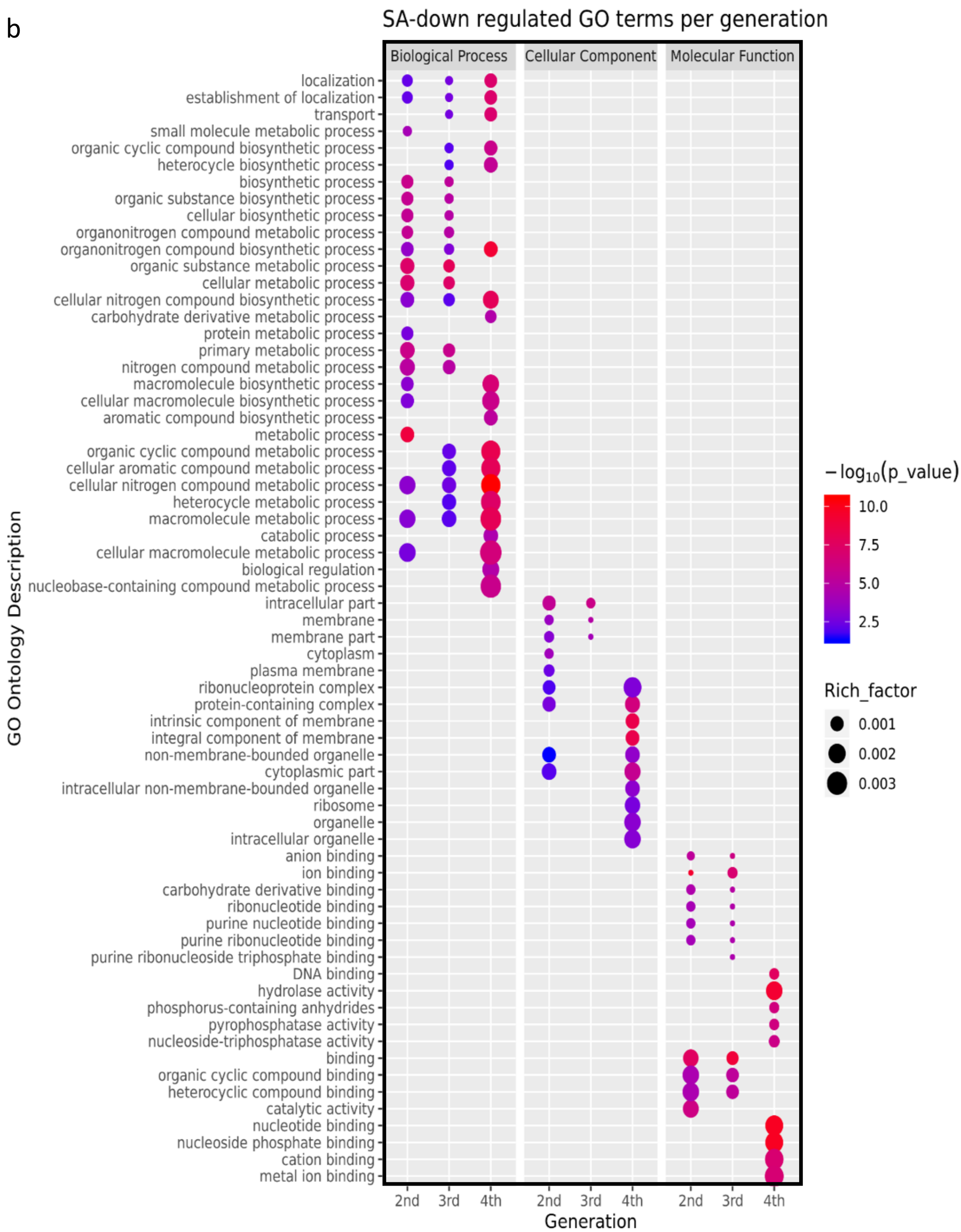



Fig. 7 Gene ontology (GO) enrichment analysis of significantly differentially expressed genes in the microbial community in the rhizosphere. A bubble chart shows

enrichment of differentially expressed GO terms. The Y-axis label lists the GO terms, the size of the bubbles represents the rich factor (= amount of differentially expressed genes enriched in the ontology/total amount of all genes in the background gene set) in different generations. Gene classification of the annotated GO terms was grouped in three categories. Colors of the bubbles represent the significance level of enrichment as calculated with Goseq. a and b represent up and down-regulated GO terms in the SA treatment, respectively. Note: in the first generation, there were no up- or down-regulated GO terms.

**Discussion**

In this study, a high-throughput metatranscriptomic sequencing approach was used to examine how the aboveground application of SA to the plant impacts the functional gene expression of the microbial communities in the rhizosphere over four subsequent generations of plant growth. Our study shows that the activation of the SA-associated plant defense pathways significantly affected the gene expression of the microbial communities in the rhizosphere, but this effect differed over four generations. Notably, the numbers of differentially expressed genes increased over generations, and there was almost no overlap of in the genes that were significantly expressed in the four generations. Moreover, foliar application of SA caused upregulation of genes of the microbial community related to GO terms associated to viral RNA genome replication, to interactions with host cells, to organelles of the host cells and to RNA polymerase activities, while downregulated GO terms of the microbial community were associated biosynthetic processes involving nitrogen and metabolic processes.



Our study shows that application of SA to plants changed the functional gene expression in the rhizosphere microbial community. This complements previous studies, which report that effects of different abiotic factors alter the expression of functional genes in the soil community (Xue et al., 2016; Castro et al., 2019). Interestingly, in our study, the highest number of significantly expressed genes was recorded in the fourth generation, which suggests that the effect of SA on gene expression becomes more pronounced over time. We did not find a selection-effect of SA on the rhizosphere bacterial community over multiple generations (results in Chapter 3). Hence, we cannot conclude that the increase in the number of significant expressed genes in our study was due to a specific rhizosphere bacterial community that became increasingly active.

Our finding that the expression of functional genes differed strongly among generations is in line with the previous findings that different taxonomic groups are present in the rhizosphere of SA treated *J. vulgaris* plants in each generation (Chapter 3). However, this clearly contrasts our prediction that there will be functional redundancy in the microbial community. In the same experiment also plant biomass was measured (Chapter 2) and SA treated plants in all generations did better than the control plants showing that from the plant's perspective different microbial taxa with different gene expressions in the rhizosphere provided similar functions. Our findings are in contrast to studies (e.g. Burke et al., 2011; Liu et al., 2018; Liu et al., 2020) that mention that particular functions of the soil microbial community are often distributed across multiple microbial taxa and more closely resemble other studies that show that environmental changes can cause selection of both different taxa and functions in the soil microbial communities (Haggerty and Dinsdale, 2017). It is important to note that in our study, in each generation we placed a subset of the microbial community in a sterile background. This may have led to selection for microbes and consequently different functions in each generation.



At the gene ontology level, we mapped 13 SA-upregulated and six SA-downregulated GO terms that were expressed in the second, third and fourth generation. The proportion of significantly expressed GO terms was high, compared to the proportion of significantly expressed genes. This is because most of the functional genes in this study could not be annotated, while at the ontology level more reads were matched with a function. As the taxa significantly selected by SA differed strongly from generation to generation, it is notable that there we detected many significant GO terms that were found in multiple generations.

Our results show that activating SA resistance in the plant drives gene expression in the rhizosphere microbiome. However, whether SA application to the plant suppressed soil pathogenic microbes remains unproven in our study. SA induced resistance is often reported to play an important role in resistance to a broad range of microbial pathogens, such as bacteria, fungi and viruses (Murphy et al., 1999;

Gilliland et al., 2003; Mayers et al., 2005; Kundu et al., 2011; Li et al., 2019; Yuan et al., 2019). Interestingly, at the ontology level, we found up-regulated GO terms that were involved in viral (RNA) genome replication and viral processes, and these GO terms increased in importance over generations. These results indicate that viruses in the soil may play a role in SA-induced resistance of host plants against soil microbes. It is well known that the soil contains bacteriophages as well as virus controlling microbial pathogens (Duckworth and Gulig, 2002; Svircev et al., 2018; Jamal et al., 2019; Kortright et al., 2019; Rehman et al., 2019). However, their exact role in the rhizosphere microbiome is still poorly understood and further studies should examine these virus-microbe-plant interactions in more detail.

In conclusion, our study shows that application of SA to the plant *J. vulgaris* causes differential gene expression in the rhizosphere microbial community. However, our data also shows that these effects vary among plant generations. Plant-defense-soil microbe interactions may be regulated by viruses or viral phages.

**Acknowledgements**

We would like to thank Karin van der Veen-van Wijk for assistance at the final biomass harvest, Connor Philippo for his practical assistance in extracting RNA and performing the quality control of raw sequencing reads, Bing Xie for his generous supporting in sharing his private server and help in troubleshooting, Fons Verbeek for sharing public resource of the server in the Leiden Institute of advanced computer science (LIACS). We also thank Yangan Chen for discussing general transcriptomic knowledge, and Martine Huberty for PCA statistical analysis. Jing Zhang would also like to thank the China Scholarship Council for financial support.

## References


Anantharaman, V. and Aravind, L., 2003. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome biology*, *4*(2), p.R11.

Andersson, C.E., Lagerbäck, P. and Carlson, K., 2010. Structure of bacteriophage T4 endonuclease II mutant E118A, a tetrameric GIY-YIG enzyme. *Journal of molecular biology*, *397*(4), pp.1003-1016.

Antoine, R., Jacob-Dubuisson, F., Drobecq, H., Willery, E., Lesjean, S. and Locht, C., 2003. Overrepresentation of a gene family encoding extracytoplasmic solute receptors in *Bordetella*. *Journal of bacteriology*, *185*(4), pp.1470-1474.

Aravind, L. and Koonin, E.V., 1998. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends in biochemical sciences*, *23*(12), pp.469-472.

Bateman, A., 2019. Uniprot: A Universal Hub of Protein Knowledge. In *Protein Science*, *28*, pp. 32-32.

Beatson, S.A., Minamino, T. and Pallen, M.J., 2006. Variation in bacterial flagellins: from sequence to structure. *Trends in microbiology*, *14*(4), pp.151-155.

Berendse, F., 1994. Competition between plant populations at low and high nutrient supplies. *Oikos*, pp.253-260.

Berendsen, R.L., Pieterse, C.M. and Bakker, P.A., 2012. The rhizosphere microbiome and plant health. *Trends in plant Science*, *17*(8), pp.478-486.

Bonocora, R.P. and Shub, D.A., 2001. A novel group I intron-encoded endonuclease specific for the anticodon region of tRNAfMet genes. *Molecular microbiology*, *39*(5), pp.1299-1306.

Boonrod, K., Galetzka, D., Nagy, P.D., Conrad, U. and Krczal, G., 2004. Single-chain antibodies against a plant viral RNA-dependent RNA polymerase confer virus resistance. *Nature biotechnology*, *22*(7), pp.856-862.

Braun, T.F., Khubbar, M.K., Saffarini, D.A. and McBride, M.J., 2005. *Flavobacterium johnsoniae* gliding motility genes identified by mariner mutagenesis. *Journal of bacteriology*, *187*(20), pp.6943-6952.

Briers, Y., Schmelcher, M., Loessner, M.J., Hendrix, J., Engelborghs, Y., Volckaert, G. and Lavigne, R., 2009. The high-affinity peptidoglycan binding domain of *Pseudomonas* phage endolysin KZ144. *Biochemical and biophysical research communications*, *383*(2), pp.187-191.

Bruce, T.J. and Pickett, J.A., 2007. Plant defence signalling induced by biotic attacks. *Current opinion in plant biology*, *10*(4), pp.387-392.

Bryant, D.M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M.B., Payzin-Dogru, D., Lee, T.J., Leigh, N.D., Kuo, T.H., Davis, F.G. and Bateman, J., 2017. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell reports*, *18*(3), pp.762-776.

Burd, C.G. and Dreyfuss, G., 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science*, *265*(5172), pp.615-621.

Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S. and Thomas, T., 2011. Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences*, U.S.A, *108*(34), pp.14288-14293.

Callaway, R.M. and Ridenour, W.M., 2004. Novel weapons: invasive success and the evolution of increased competitive ability. *Frontiers in Ecology and the Environment*, *2*(8), pp.436-443.

Castro, S.P., Cleland, E.E., Wagner, R., Al Sawad, R. and Lipson, D.A., 2019. Soil microbial responses to drought and exotic plants shift carbon metabolism. *The ISME journal*, *13*(7), pp.1776-1787.

Cesarano, G., De Filippis, F., La Storia, A., Scala, F. and Bonanomi, G., 2017. Organic amendment type and application frequency affect crop yields, soil fertility and microbiome composition. *Applied Soil Ecology*, *120*, pp.254-264.

Cho, K.H. and Salyers, A.A., 2001. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. *Journal of Bacteriology*, *183*(24), pp.7224-7230.

Compant, S., Clément, C. and Sessitsch, A., 2010. Plant growth-promoting bacteria in the rhizo-and endosphere of plants: their role, colonization, mechanisms involved and prospects for utilization. *Soil Biology and Biochemistry*, *42*(5), pp.669-678.

Convery, M.A., Rowsell, S., Storehouse, N.J., Ellington, A.D., Hirao, I., Murray, J.B., Peabody, D.S., Phillips, S.E. and Stockley, P.G., 1998. Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nature structural biology*, *5*(2), pp.133-139.

D'Amico, S., Collins, T., Marx, J.C., Feller, G., Gerday, C. and Gerday, C., 2006. Psychrophilic microorganisms: challenges for life. *EMBO reports*, *7*(4), pp.385-389.

De Mot, R., Schoofs, G., Roelandt, A., Declerck, P., Proost, P., Damme, J.V. and Vanderleyden, J., 1994. Molecular characterization of the major outermembrane protein OprF from plant root-colonizing *Pseudomonas fluorescens*. *Microbiology*, *140*(6), pp.1377-1387.

Duckworth, D.H. and Gulig, P.A., 2002. Bacteriophages. *BioDrugs*, *16*(1), pp.57-62.

Dunin-Horkawicz, S., Feder, M. and Bujnicki, J.M., 2006. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC genomics*, *7*(1), p.98.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A. and Sonnhammer, E.L.L., 2019. The Pfam protein families database in 2019. *Nucleic acids research*, *47*(D1) pp.D427-D432.

Felix, G., Duran, J.D., Volko, S. and Boller, T., 1999. Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *The Plant Journal*, *18*(3), pp.265-276.

Fernando Gil, J., Wibberg, D., Eini, O., Savenkov, E.I., Varrelmann, M. and Liebe, S., 2020. Comparative transcriptome analysis provides molecular insights into the interaction of beet necrotic yellow vein virus and beet soil-borne mosaic virus with their host sugar beet. *Viruses*, *12*(1), p.76.

Moscat, J., Diaz-Meco, M.T., Albert, A. and Campuzano, S., 2006. Cell signaling and function organized by PB1 domain interactions. *Molecular cell*, *23*(5), pp.631-640.

Gao, Q.M., Zhu, S., Kachroo, P. and Kachroo, A., 2015. Signal regulators of systemic acquired resistance. *Frontiers in plant Science*, *6*, p.228.

Gilliland, A., Singh, D.P., Hayward, J.M., Moore, C.A., Murphy, A.M., York, C.J., Slator, J. and Carr, J.P., 2003. Genetic modification of alternative respiration has differential effects on antimycin A-induced versus salicylic acid-induced resistance to *Tobacco mosaic* virus. *Plant Physiology*, *132*(3), pp.1518-1528.

Haggerty, J.M. and Dinsdale, E.A., 2017. Distinct biogeographical patterns of marine bacterial taxonomy and functional genes. *Global Ecology and Biogeography*, *26*(2), pp.177-190.

Haney, C.H. and Ausubel, F.M., 2015. Plant microbiome blueprints. *Science*, *349*(6250), pp.788-789.

Hardy, S.F., German, T.L., Loesch-Fries, L.S. and Hall, T.C., 1979. Highly active template-specific RNA-dependent RNA polymerase from barley leaves infected with brome mosaic virus. *Proceedings of the National Academy of Sciences*, *U.S.A,* 76(10), pp.4956-4960.

Henry, J.T. and Crosson, S., 2011. Ligand-binding PAS domains in a genomic, cellular, and structural context. *Annual review of microbiology*, *65*, pp.261-286.

Hong, Y., Cole, T.E., Brasier, C.M. and Buck, K.W., 1998. Evolutionary relationships among putative RNA-dependent RNA polymerases encoded by a mitochondrial virus-like RNA in the Dutch elm disease fungus, *Ophiostoma novo-ulmi*, by other viruses and virus-like RNAs and by the *Arabidopsis* mitochondrial genome. *Virology*, *246*(1), pp.158-169.

Hoppe, J., Ünal, C.M., Thiem, S., Grimpe, L., Goldmann, T., Gaßler, N., Richter, M., Shevchuk, O. and Steinert, M., 2017. PilY1 promotes *Legionella pneumophila* infection of human lung tissue explants and contributes to bacterial adhesion, host cell invasion, and twitching motility. *Frontiers in cellular and infection microbiology*, *7*, p.63.

Huang, X.F., Chaparro, J.M., Reardon, K.F., Zhang, R., Shen, Q. and Vivanco, J.M., 2014. Rhizosphere interactions: root exudates, microbes, and microbial communities. *Botany*, *92*(4), pp.267-275.

Ito, T., Matsui, Y., Ago, T., Ota, K. and Sumimoto, H., 2001. Novel modular domain PB1 recognizes PC motif to mediate functional protein-protein interactions. *The EMBO journal*, *20*(15), pp.3938-3946.

Jamal, M., Bukhari, S.M., Andleeb, S., Ali, M., Raza, S., Nawaz, M.A., Hussain, T., Rahman, S.U. and Shah, S.S., 2019. Bacteriophages: an overview of the control strategies against multiple bacterial infections in different fields. *Journal of basic microbiology*, *59*(2), pp.123-133.

Jeffries, P., Gianinazzi, S., Perotto, S., Turnau, K. and Barea, J.M., 2003. The contribution of arbuscular mycorrhizal fungi in sustainable maintenance of plant health and soil fertility. *Biology and fertility of soils*, *37*(1), pp.1-16.

Johnson, E.R. and McKay, D.B., 1999. Crystallographic structure of the amino terminal domain of yeast initiation factor 4A, a representative DEAD-box RNA helicase. *Rna*, *5*(12), pp.1526-1534.

Jung, A.L., Stoiber, C., Herkt, C.E., Schulz, C., Bertrams, W. and Schmeck, B., 2016. *Legionella pneumophila*-derived outer membrane vesicles promote bacterial replication in macrophages. *PLoS pathogens*, *12*(4), p.e1005592.

Kachroo, A. and Robin, G.P., 2013. Systemic signaling during plant defense. *Current Opinion in Plant Biology*, *16*(4), pp.527-533.

Kolde, R. and Kolde, M.R., 2015. Package ‘pheatmap’. *R Package*, *1*(7), p.790.

Kortright, K.E., Chan, B.K., Koff, J.L. and Turner, P.E., 2019. Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell host & microbe*, *25*(2), pp.219-232.



Krishna, S.S., Majumdar, I. and Grishin, N.V., 2003. Structural classification of zinc fingers: survey and summary. *Nucleic acids research*, *31*(2), pp.532-550.

Krogh, S., Jørgensen, S.T. and Devine, K.M., 1998. Lysis genes of the *Bacillus subtilis* defective prophage PBSX. *Journal of bacteriology*, *180*(8), pp.2110-2117.

Kundu, S., Chakraborty, D. and Pal, A., 2011. Proteomic analysis of salicylic acid induced resistance to Mungbean yellow mosaic India virus in Vigna mungo. *Journal of proteomics*, *74*(3), pp.337-349.

Kuo, W.T., Chin, K.H., Lo, W.T., Wang, A.H.J. and Chou, S.H., 2008. Crystal structure of the C-terminal domain of a flagellar hook-capping protein from *Xanthomonas campestris*. *Journal of molecular biology*, *381*(1), pp.189-199.

Li, T., Huang, Y., Xu, Z.S., Wang, F. and Xiong, A.S., 2019. Salicylic acid-induced differential resistance to the Tomato yellow leaf curl virus among resistant and susceptible tomato cultivars. *BMC plant biology*, *19*(1), pp.1-14.

Lindquist, J.A. and Mertens, P.R., 2018. Cold shock proteins: from cellular mechanisms to pathophysiology and disease. *Cell Communication and Signaling*, *16*(1), p.63.

Liu, Y.C., Machuca, M.A., Beckham, S.A., Gunzburg, M.J. and Roujeinikova, A., 2015. Structural basis for amino-acid recognition and transmembrane signalling by tandem Per-Arnt-Sim (tandem PAS) chemoreceptor sensory domains. *Acta Crystallographica Section D: Biological Crystallography*, *71*(10), pp.2127-2136.

Liu, Y.R., Delgado-Baquerizo, M., Bi, L., Zhu, J. and He, J.Z., 2018. Consistent responses of soil microbial taxonomic and functional attributes to mercury pollution across China. *Microbiome*, *6*(1), pp.1-12.

Liu, Y.R., Delgado-Baquerizo, M., Yang, Z., Feng, J., Zhu, J. and Huang, Q., 2020. Microbial taxonomic and functional attributes consistently predict soil CO2 emissions across contrasting croplands. *Science of The Total Environment*, *702*, p.134885.

Maurhofer, M., Reimmann, C., Schmidli-Sacherer, P., Heeb, S., Haas, D. and Défago, G., 1998. Salicylic acid biosynthetic genes expressed in *Pseudomonas fluorescens* strain P3 improve the induction of systemic resistance in tobacco against tobacco necrosis virus. *Phytopathology*, *88*(7), pp.678-684.

Mayers, C.N., Lee, K.C., Moore, C.A., Wong, S.M. and Carr, J.P., 2005. Salicylic acid-induced resistance to *Cucumber mosaic virus* in squash and *Arabidopsis thaliana*: contrasting mechanisms of induction and antiviral action. *Molecular plant-microbe interactions*, *18*(5), pp.428-434.

Mazzoleni, S., Bonanomi, G., Incerti, G., Chiusano, M.L., Termolino, P., Mingo, A., Senatore, M., Giannino, F., Cartenì, F., Rietkerk, M. and Lanzotti, V., 2015. Inhibitory and toxic effects of extracellular self-DNA in litter: a mechanism for negative plant-soil feedbacks? *New Phytologist*, *205*(3), pp.1195-1210.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. and Tollervey, D., 1997. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3′→ 5′ exoribonucleases. *Cell*, *91*(4), pp.457-466.

Murphy, A.M., Chivasa, S., Singh, D.P. and Carr, J.P., 1999. Salicylic acid-induced resistance to viruses and other pathogens: a parting of the ways? *Trends in plant science*, *4*(4), pp.155-160.

Nar, H., Huber, R., Meining, W., Schmid, C., Weinkauf, S. and Bacher, A., 1995. Atomic structure of GTP cyclohydrolase I. *Structure*, *3*(5), pp.459-466.

Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V., Van Etten, J.L. and Rossmann, M.G., 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proceedings of the National Academy of Sciences*, *U.S.A,* 99(23), pp.14758-14763.

Nijjer, S., Rogers, W.E. and Siemann, E., 2007. Negative plant-soil feedbacks may limit persistence of an invasive tree due to rapid accumulation of soil pathogens. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1625), pp.2621-2627.

Ökmen, B. and Doehlemann, G., 2014. Inside plant: biotrophic strategies to modulate host immunity and metabolism. *Current opinion in plant biology*, *20*, pp.19-25.

Pao, G.M. and Saier, M.H., 1995. Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. *Journal of Molecular Evolution*, *40*(2), pp.136-154.

Pieterse, C.M., Van Pelt, J.A., Van Wees, S.C., Ton, J., Léon-Kloosterziel, K.M., Keurentjes, J.J., Verhagen, B.W., Knoester, M., Van der Sluis, I., Bakker, P.A. and Van Loon, L.C., 2001. Rhizobacteria-mediated induced systemic resistance: triggering, signalling and expression. *European Journal of Plant Pathology*, *107*(1), pp.51-61.

Pieterse, C.M., Zamioudis, C., Berendsen, R.L., Weller, D.M., Van Wees, S.C. and Bakker, P.A., 2014. Induced systemic resistance by beneficial microbes. *Annual review of phytopathology*, *52*, pp.347-375.

Pozo, M.J. and Azcón-Aguilar, C., 2007. Unraveling mycorrhiza-induced resistance. *Current opinion in plant biology*, *10*(4), pp.393-398.

Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Liu, H., Kuehl, J.V., Melnyk, R.A., Lamson, J.S., Suh, Y. and Carlson, H.K., 2018. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*, *557*(7706), pp.503-509.

Raaijmakers, J.M., Paulitz, T.C., Steinberg, C., Alabouvette, C. and Moënne-Loccoz, Y., 2009. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant and soil*, *321*(1-2), pp.341-361.

Ramos, H.C., Rumbo, M. and Sirard, J.C., 2004. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends in microbiology*, *12*(11), pp.509-517.

Rehman, S., Ali, Z., Khan, M., Bostan, N. and Naseem, S., 2019. The dawn of phage therapy. *Reviews in medical virology*, *29*(4), p.e2041.

Reymond, P. and Farmer, E.E., 1998. Jasmonate and salicylate as global signals for defense gene expression. *Current opinion in plant biology*, *1*(5), pp.404-411.

Robinson, M.D. and Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, *11*(3), pp.1-9.

Saier, M.H., 2000. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiology and molecular biology reviews*, *64*(2), pp.354-411.

Schwach, F., Vaistij, F.E., Jones, L. and Baulcombe, D.C., 2005. An RNA-dependent RNA polymerase prevents meristem invasion by potato virus X and is required for the activity but not the production of a systemic silencing signal. *Plant physiology*, *138*(4), pp.1842-1852.

Sganga, M.W., Aksamit, R.R., Cantoni, G.L. and Bauer, C.E., 1992. Mutational and nucleotide sequence analysis of S-adenosyl-L-homocysteine hydrolase from *Rhodobacter capsulatus*. *Proceedings of the National Academy of Sciences*, *U.S.A* 89(14), pp.6328-6332.

Shah, J. and Zeier, J., 2013. Long-distance communication and signal amplification in systemic acquired resistance. *Frontiers in Plant Science*, *4*, p.30.

Shen, C., Shi, Y., Ni, Y., Deng, Y., Van Nostrand, J.D., He, Z., Zhou, J. and Chu, H., 2016. Dramatic increases of soil microbial functional gene diversity at the treeline ecotone of Changbai Mountain. *Frontiers in microbiology*, *7*, p.1184.



Sharma, M., Ellis, R.L. and Hinton, D.M., 1992. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proceedings of the National Academy of Sciences*, *U.S.A,* 89(14), pp.6658-6662.

Steen, A., Buist, G., Leenhouts, K.J., El Khattabi, M., Grijpstra, F., Zomer, A.L., Venema, G., Kuipers, O.P. and Kok, J., 2003. Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. *Journal of Biological Chemistry*, *278*(26), pp.23874-23881.

Stuwe, T., Hothorn, M., Lejeune, E., Rybin, V., Bortfeld, M., Scheffzek, K. and Ladurner, A.G., 2008. The FACT Spt16 "peptidase" domain is a histone H3-H4 binding module. *Proceedings of the National Academy of Sciences*, *U.S.A, 105*(26), pp.8884-8889.

Sumimoto, H., Kamakura, S. and Ito, T., 2007. Structure and function of the PB1 domain, a protein interaction module conserved in animals, fungi, amoebas, and plants. *Science's STKE*, *2007*(401), pp.re6-re6.

Svircev, A., Roach, D. and Castle, A., 2018. Framing the future with bacteriophages in agriculture. *Viruses*, *10*(5), p.218.

Tanaka, S., Han, X. and Kahmann, R., 2015. Microbial effectors target multiple steps in the salicylic acid production and signaling pathway. *Frontiers in Plant Science*, *6*, p.349.

Van Wees, S.C., Van der Ent, S. and Pieterse, C.M., 2008. Plant immune responses triggered by beneficial microbes. *Current opinion in plant biology*, *11*(4), pp.443-448.

Verberne, M.C., Verpoorte, R., Bol, J.F., Mercado-Blanco, J. and Linthorst, H.J., 2000. Overproduction of salicylic acid in plants by bacterial transgenes enhances pathogen resistance. *Nature biotechnology*, *18*(7), pp.779-783.

Vlot, A.C., Dempsey, D.M.A. and Klessig, D.F., 2009. Salicylic acid, a multifaceted hormone to combat disease. *Annual review of phytopathology*, *47*, pp.177-206.

Wardle, D.A., Bardgett, R.D., Callaway, R.M. and Van der Putten, W.H., 2011. Terrestrial ecosystem responses to species gains and losses. *Science*, *332*(6035), pp.1273-1277.

Walters, D. and Heil, M., 2007. Costs and trade-offs associated with induced resistance. *Physiological and Molecular Plant Pathology*, *71*(1-3), pp.3-17.

Westman, J., Hube, B. and Fairn, G.D., 2019. Integrity under stress: Host membrane remodelling and damage by fungal pathogens. *Cellular microbiology*, *21*(4), p.e13016.

Wilkens, S., Zhang, Z. and Zheng, Y., 2005. A structural model of the vacuolar ATPase from transmission electron microscopy. *Micron*, *36*(2), pp.109-126.

Williams, M.A., Taylor, E.B. and Mula, H.P., 2010. Metaproteomic characterization of a soil microbial community following carbon amendment. *Soil Biology and Biochemistry*, *42*(7), pp.1148-1156.

Witherell, G.W., Gott, J.M. and Uhlenbeck, O.C., 1991. Specific interaction between RNA phage coat proteins and RNA. *Progress in nucleic acid research and molecular biology*, *40*, pp.185-220.

Xue, K., Xie, J., Zhou, A., Liu, F., Li, D., Wu, L., Deng, Y., He, Z., Van Nostrand, J.D., Luo, Y. and Zhou, J., 2016. Warming alters expressions of microbial functional genes important to ecosystem functioning. *Frontiers in microbiology*, *7*, p.668.

Yao, Z., Zou, C., Peng, N., Zhu, Y., Bao, Y., Zhou, Q., Wu, Q., Chen, B. and Zhang, M., 2020. Virome identification and characterization of *Fusarium sacchari* and *F. andiyazi*: causative agents of *Pokkah boeng* disease in sugarcane. *Frontiers in microbiology*, *11*, p.240.

Yeh, L.C.C. and Lee, J.C., 1998. Yeast ribosomal proteins L4, L17, L20, and L25 exhibit different binding characteristics for the yeast 35S precursor rRNA. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, *1443*(1-2), pp.139-148.

Younas, F., Soltanmohammadi, N., Knapp, O. and Benz, R., 2018. The major outer membrane protein of *Legionella pneumophila* Lpg1974 shows pore-forming characteristics similar to the human mitochondrial outer membrane pore, hVDAC1. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1860*(8), pp.1544-1553.

Yuan, W., Jiang, T., Du, K., Chen, H., Cao, Y., Xie, J., Li, M., Carr, J.P., Wu, B., Fan, Z. and Zhou, T., 2019. Maize phenylalanine ammonia-lyases contribute to resistance to Sugarcane mosaic virus infection, most likely through positive regulation of salicylic acid accumulation. *Molecular plant pathology*, *20*(10), pp.1365-1378.

Zhang, G. and Darst, S.A., 1998. Structure of the *Escherichia coli* RNA polymerase α subunit amino-terminal domain. *Science*, *281*(5374), pp.262-266.

**Supplementary data**

Table S1 Log2 (FC) of 70 differentially expressed annotated genes, the expression of which are significantly altered by SA treatments in at least one generation in the rhizosphere of *J. vulgaris* plants in four generations. When Log2 (FC) is > 0, the gene is up-regulated in the SA treatment and when Log2 (FC) is < 0, the gene is down-regulated in the SA treatment. '-' indicates that the gene was not detected in the treatment; 'ns' represents the gene is not significantly altered by the SA treatment, but it is present. 1,2,3 and 4 represent the four generations.

| Isoform ID | Functional gene name | 1 | 2 | 3 | 4 | SA down-/up-regulation | Go category | GO term | Homologous organisms | Function | Literature |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN187408_c0_g1_i1 | PF05150.12^Legionella_OMP | - | -7.5 | - | - | down | Biological process | adhesion of symbiont to host cell | bacteria | Major outer membrane protein; The attachment of a symbiont to a host cell via adhesion molecules, general stickiness etc., either directly or indirectly. | Jung et al., 2016; Hoppe et al., 2017; Younas et al., 2018. |
| TRINITY_DN237954_c0_g1_i1 | PF05150.12^Legionella_OMP | - | -6.3 | - | - | down | Biological process | adhesion of symbiont to host cell | bacteria | Major outer membrane protein; The attachment of a symbiont to a host cell via adhesion molecules, general stickiness etc., either directly or indirectly. | Jung et al., 2016; Hoppe et al., 2017; Younas et al., 2018. |
| TRINITY_DN2161_c0_g4_i1 | PF00687.21^Ribosomal_L1 | ns | ns | -5.1 | -4.7 | down | Biological process | cellular metabolic process | bacteria | Ribosomal protein L1 is the largest protein from the large ribosomal subunit, also can involve moleclar function of RNA binding and cellular component of large ribosomal subunit. | |
| TRINITY_DN116491_c0_g1_i1 | PF00729.18^Viral_coat | - | -6.7 | - | - | down | Biological process | virual process | virus | Involing with multi-organism process in which a virus is a participant. The other participant is the host. Includes infection of a host cell, replication of the viral genome, and assembly of progeny virus particles. In some cases the viral genetic material may integrate into the host genome and only subsequently, under particular circumstances, 'complete' its life cycle. Viral genome integration into host DNA. | Nandhagopal et al., 2002. |
| TRINITY_DN221013_c0_g1_i5 | PF03863.13^Phage_mat-A | - | -7.7 | - | ns | down | Biological process or cellular component | virion | bacteriophage | The process by which a virion attaches to a host cell by binding to a pilus on the host cell surface. The complete fully infectious extracellular virus particle. | Witherell et al., 1991. |
| TRINITY_DN342313_c0_g1_i2 | PF03863.13^Phage_mat-A | ns | - | ns | -7.9 | down | Biological process or cellular component | virion | bacteriophage | The process by which a virion attaches to a host cell by binding to a pilus on the host cell surface. The complete fully infectious extracellular virus particle. | Witherell et al., 1991. |
| TRINITY_DN497_c0_g1_i6 | PF03863.13^Phage_mat-A | ns | ns | ns | -2.8 | down | Biological process or cellular component | virion | bacteriophage | The process by which a virion attaches to a host cell by binding to a pilus on the host cell surface. The complete fully infectious extracellular virus particle. | Witherell et al., 1991. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN2233_c0_g1_i1 | PF03863.13^Phage_mat-A | ns | ns | ns | -1.5 | down | Biological process or cellular component | virion | bacteriophage | The process by which a virion attaches to a host cell by binding to a pilus on the host cell surface. The complete fully infectious extracellular virus particle. | Witherell et al., 1991. |
| TRINITY_DN418_c0_g1_i14 | PF00460.20^Flg_bb_rod | ns | ns | ns | -4.6 | down | Cellular component | bacterial-type flagellum basal body, rod | bacteria | The central portion of the bacterial-type flagellar basal body, which spans the periplasm and threads through the rings. | |
| TRINITY_DN4152_c3_g1_i1 | PF05736.11^OprF | - | ns | -4.3 | ns | down | Cellular component | cell outer membrane | bacteria | This domain represents the presumed membrane-spanning region of the OprF proteins. This region is involved in channel formation and is thought to form an 8-stranded beta-barrel. | De Mot et al., 1994. |
| TRINITY_DN5762_c0_g1_i2 | PF11645.8^PDDEXK_5 | ns | ns | - | -3.7 | down | Cellular component | integral component of membrane | bacteria | This family of endonucleases includes a group I intron-encoded endonuclease. This family belongs to the PD (D/E)XK superfamily | Bonocora and Shub, 2001. |
| TRINITY_DN21112_c0_g1_i1 | PF03863.13^Phage_mat-A | - | - | ns | -5.6 | down | Cellular component | membrane | bacteria | Flagellins polymerise to form bacterial flagella. This family includes flagellins and hook associated protein 3. Structurally this family forms an extended helix that interacts with PF00700. | Felix et al., 1999 ; Ramos et al., 2004; Beatson et al., 2006. |
| TRINITY_DN1559755_c0_g1_i1 | PF00564.24^PB1 | - | - | -6.1 | ns | down | Cellular component | membrane | | Phox and Bem1 (PB1) domains contain approximately 80 amino acids and are found in a number of cytoplasmic signaling proteins. A PB1 domain may form heterodimers with a paired PB1 domain, although not all PB1 domains will associate with one another. A highly conserved internal sequence known as OPR, PC or AID motifs is necessary for PB1 domain function. Regions outside the OPR, PC and AID help confer specificity for binding. | Ito et al., 2001; Moscat et al., 2006; Sumimoto et al., 2007. |
| TRINITY_DN4077_c1_g2_i1 | PF00573.22^Ribosomal_L4 | - | -4.7 | ns | ns | down | Cellular component | ribosome | eubacteria | This family includes Ribosomal L4/L1 from eukaryotes and archaebacteria and L4 from eubacteria. L4 from yeast has been shown to bind rRNA. | Yeh and Lee, 1998. |
| TRINITY_DN4472_c0_g1_i5 | PF00281.19^Ribosomal_L5 | - | ns | ns | -6.8 | down | Cellular component | ribosome | | | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN164781_c0_g2_i2 | PF03725.15^RNase_PH_C | - | ns | ns | -5.8 | down | Molecular function | 5'-3' exoribonuclease activity | bacteria | This family includes 3'-5' exoribonucleases. Ribonuclease PH contains a single copy of this domain, and removes nucleotide residues following the -CCA terminus of tRNA. Polyribonucleotide nucleotidyltransferase (PNPase) contains two tandem copies of the domain. PNPase is involved in mRNA degradation in a 3'-5' direction. The exosome is a 3'-5' exoribonuclease complex that is required for 3' processing of the 5.8S rRNA. | Mitchell et al., 1997. |
| TRINITY_DN41896_c0_g1_i1 | PF05221.17^AdoHcyase | - | ns | - | -5.5 | down | Molecular function | adenosylhomocysteinase activity | | AdoHcyase is an enzyme of the activated methyl cycle, responsible for the reversible hydration of S-adenosyl-L-homocysteine into adenosine and homocysteine. | Sganga et al., 1992. |
| TRINITY_DN30386_c0_g2_i1 | PF12081.8^GldM_N | - | - | ns | -5.7 | down | Molecular function | ATP binding | bacteria | This domain is found in bacteria at the N-terminus of the GldM protein. This domain is typically between 169 to 182 amino acids in length. This domain has two completely conserved residues (Y and N) that may be functionally important. GldM is named for the member from *Cytophaga johnsonae* (*Flavobacterium johnsoniae*), which is required for a type of rapid gliding motility found in certain members of the *Bacteriodetes*. | Braun et al., 2005. |
| TRINITY_DN24155_c0_g2_i1 | PF01471.18^PG_binding_1 | - | ns | - | -6.0 | down | Molecular function | binding | bacetria | Putative peptidoglycan binding domain. It is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation | Krogh et al., 1998; Steen et al., 2003; Briers et al., 2009. |
| TRINITY_DN1498_c0_g1_i2 | PF01471.18^PG_binding_1 | - | - | - | -3.2 | down | Molecular function | binding | bacetria | Putative peptidoglycan binding domain. It is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation. | Krogh et al., 1998; Steen et al., 2003; Briers et al., 2009. |
| TRINITY_DN332299_c0_g1_i1 | PF13620.6^CarboxypepD_reg | - | ns | ns | -5.9 | down | Molecular function | carboxypeptidase activity | | | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN 664942_c0_g2_i1 | PF00270.29^DEAD | - | - | - | -5.4 | down | Molecular function | DEAD/H-box RNA helicase binding | | The DEAD/DEAH box helicases are a family of proteins whose purpose is to unwind nucleic acids. The DEAD box helicases are involved in various aspects of RNA metabolism, including nuclear transcription, pre mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay and organellar gene expression. | Johnson and McKay, 1999. |
| TRINITY_DN 16479_c0_g1_i8 | PF01541.24^GIY-YIG | ns | -7.0 | -7.1 | -7.1 | down | Molecular function | endonuclease activity | bacteriophage | It is involved in degradation of host DNA, permitting scavenging of host-derived nucleotides for phage DNA synthesis; in enzymes involved in DNA repair and recombination. | Sharma et al., 1992; Dunin-Horkawicz et al., 2006; Andersson et al., 2010. |
| TRINITY_DN 4160_c0_g2_i3 | PF01541.24^GIY-YIG | - | -6.4 | ns | ns | down | Molecular function | endonuclease activity | bacteriophage | It is involved in degradation of host DNA, permitting scavenging of host-derived nucleotides for phage DNA synthesis; in enzymes involved in DNA repair and recombination. | Sharma et al., 1992; Dunin-Horkawicz et al., 2006; Andersson et al., 2010. |
| TRINITY_DN 2301_c0_g4_i1 | PF00006.25^ATP-synt_ab | - | -5.1 | - | - | down | Molecular function | hydrolase activity or RNA binding | bacteria, archaea eukaryotes | Catalysis of the hydrolysis of various bonds, e.g. C-O, C-N, C-C, phosphoric anhydride bonds, etc. Hydrolase is the systematic name for any enzyme of EC class 3 or Interacting selectively and non-covalently with an RNA molecule or a portion thereof. | Wilkens et al., 2005. |
| TRINITY_DN 1070_c1_g1_i1 | PF13609.6^Porin_4 | - | -5.7 | -4.6 | -4.0 | down | Molecular function | porin activity | bacteria | Enables the transfer of substances, sized less than 1000 Da, from one side of a membrane to the other. The transmembrane portions of porins consist exclusively of beta-strands, which form a beta-barrel. They are found in the outer membranes of Gram-negative bacteria, mitochondria, plastids and possibly acid-fast Gram-positive bacteria. | Saier, 2000. |
| TRINITY_DN 6228_c0_g1_i3 | PF00989.25^PAS | - | - | -6.5 | -5.9 | down | Molecular function | protein binding | bacteria | L13are positioned at the amino terminus of signaling proteins such as sensor histidine kinases, cyclic-di-GMP synthases and hydrolases, and methyl-accepting chemotaxis proteins. | Henry and Crosson, 2011; Liu et al., 2015. |



4

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN135_c0_g1_i7 | PF13360.6^PQQ_2 | ns | ns | ns | -2.5 | down | Molecular function | pyrroloquinoline quinone binding | | Interacting selectively and non-covalently with pyrroloquinoline quinone, PQQ, the coenzyme or the prosthetic group of certain alcohol dehydrogenases and glucose dehydrogenases. | |
| TRINITY_DN20104_c0_g1_i6 | PF05919.11^Mitovir_RNA_pol | ns | -7.0 | ns | -6.8 | down | Molecular function | RNA-directed 5'-3' RNA polymerase activity | virus | Mitovirus RNA-dependent RNA polymerase proteins. The family also contains fragment matches in the mitochondria of *Arabidopsis thaliana*. | Hong et al., 1998; Yao et al., 2020. |
| TRINITY_DN264317_c0_g1_i1 | PF03431.13^RNA_replicase_B | - | ns | -8.0 | ns | down | Molecular function | RNA-directed 5'-3' RNA polymerase activity | virus | This family is of Leviviridae RNA replicases. The replicase is also known as RNA dependent RNA polymerase. | Fernando et al., 2020. |
| TRINITY_DN109372_c0_g1_i2 | PF05919.11^Mitovir_RNA_pol | ns | -5.5 | ns | -7.2 | down | Molecular function | RNA-directed 5'-3' RNA polymerase activity | virus | Mitovirus RNA-dependent RNA polymerase proteins. The family also contains fragment matches in the mitochondria of *Arabidopsis thaliana*. | Hong et al., 1998; Yao et al., 2020. |
| TRINITY_DN5435_c0_g1_i4 | PF05919.11^Mitovir_RNA_pol | - | -5.0 | - | -7.3 | down | Molecular function | RNA-directed 5'-3' RNA polymerase activity | virus | Mitovirus RNA-dependent RNA polymerase proteins. The family also contains fragment matches in the mitochondria of *Arabidopsis thaliana*. | Hong et al., 1998; Yao et al., 2020. |
| TRINITY_DN34032_c0_g1_i1 | PF05919.11^Mitovir_RNA_pol | - | ns | -7.2 | -6.6 | down | Molecular function | RNA-directed 5'-3' RNA polymerase activity | virus | Mitovirus RNA-dependent RNA polymerase proteins. The family also contains fragment matches in the mitochondria of *Arabidopsis thaliana*. | Hong et al., 1998; Yao et al., 2020. |
| TRINITY_DN19010_c0_g4_i3 | PF00978.21^RdRP_2 | ns | -6.3 | - | - | down | Molecular function | viral replication or RNA-directed RNA polymerase activity | virus | RNA-dependent RNA polymerase or RNA replicase is an enzyme that catalyzes the replication of RNA from an RNA template. Specifically, it catalyses synthesis of the RNA strand complementary to a given RNA template. | . Hardy et al., 1979; Boonrod et al., 2004; Schwan et al., 2005. |
| TRINITY_DN243032_c0_g1_i2 | PF05919.11^Mitovir_RNA_pol | - | - | - | -6.3 | down | Biological process | RNA-directed 5'-3' RNA polymerase activity | virus | Mitovirus RNA-dependent RNA polymerase proteins. The family also contains fragment matches in the mitochondria of *Arabidopsis thaliana*. | Hong et al., 1998; Yao et al., 2020. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN2369_c0_g1_i4 | PF00669.20^Flagellin_N | ns | ns | ns | -1.7 | down | Cellular component | membrane | bacteria | Flagellins polymerise to form bacterial flagella. This family includes flagellins and hook associated protein 3. Structurally this family forms an extended helix that interacts with PF00700. | Felix et al., 1999; Ramos et al., 2004; Beatson et al., 2006. |
| TRINITY_DN306_c0_g2_i5 | COG1278^Cold shock protein | ns | 3.8 | ns | ns | up | Biological process | Many biological processes like regulation of transcription, translation, | | Response to stress, any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a cold stimulus, a temperature stimulus below the optimal temperature for that organism. Cold shock proteins are multifunctional RNA/DNA binding proteins, characterized by the presence of one or more cold shock domains. I | D'Amico et al., 2006; Williams et al., 2010; Lindquist and Mertens, 2018. |
| TRINITY_DN1057_c0_g1_i4 | COG1278^Cold shock protein | ns | 4.4 | ns | - | up | Biological process | Many biological processes like regulation of transcription, translation, | | Response to stress, any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a cold stimulus, a temperature stimulus below the optimal temperature for that organism. Cold shock proteins are multifunctional RNA/DNA binding proteins, characterized by the presence of one or more cold shock domains. | D'Amico et al., 2006; Williams et al., 2010; Lindquist and Mertens, 2018. |
| TRINITY_DN473649_c0_g1_i3 | PF05551.11^zf-His_Me_endon | - | ns | - | 5.8 | up | Biological process | nucleic acid phosphodiester bond hydrolysis | | Zinc-binding loop region of homing endonuclease. This domain is the short zinc-binding loops region of a number of much longer chain homing endonucleases. | Krishna et al., 2003. |
| TRINITY_DN3358_c0_g1_i2 | PF05551.11^F25:F73n | - | - | - | 5.9 | up | Biological process | nucleic acid phosphodiester bond hydrolysis | | Zinc-binding loop region of homing endonuclease. This domain is the short zinc-binding loops region of a number of much longer chain homing endonucleases. This domain is the short zinc-binding loops region of a number of much longer chain homing endonucleases. | Krishna et al., 2003. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN422335_c0_g1_i1 | PF01618.16^MotA_ExbB | - | ns | ns | 3.3 | up | Biological process | protein transport | bacteria | This family groups together integral membrane proteins that appear to be involved translocation of proteins across a membrane. These proteins are probably proton channels. MotA is an essential component of the flageller motor that uses a proton gradient to generate rotational motion in the flageller. ExbB is part of the TonB-dependent transduction complex. The TonB complex uses the proton gradient across the inner bacterial membrane to transport large molecules across the outer bacterial membrane | Braun et al., 1999. |
| TRINITY_DN1442_c0_g1_i12 | PF13860.6^FlgD_ig | ns | ns | ns | 2.5 | up | Biological process | proteolysis | bacteria | This domain has an immunoglobulin like beta sandwich fold. It is found in the FlgD protein the flagellar hook capping protein. THe structure for this domain shows that it is inserted within a TUDOR like beta barrel domain | Kuo et al., 2008. |
| TRINITY_DN220624_c0_g1_i6 | PF00072.24^Response_reg | - | - | ns | 5.4 | up | Biological Process | signal transduction | | Response regulator receiver domain. This domain receives the signal from the sensor partner in bacterial two-component systems. It is usually found N-terminal to a DNA binding effector domain. | Pao and Saier, 1995. |
| TRINITY_DN6829_c0_g2_i1 | PF03863.13^Phage_mat-A | - | ns | ns | 2.2 | up | Biological process or cellular component | virion | bacteriophage | The process by which a virion attaches to a host cell by binding to a pilus on the host cell surface. Or The complete fully infectious extracellular virus particle. | Witherell et al., 1991. |
| TRINITY_DN924_c0_g3_i1 | PF00189.20^Ribosomal_S3_C | - | 4.6 | ns | - | up | Biological process or molecular function | cellular metabolic process or structural constituent of ribosome | | The cellular metabolic process in which a protein is formed, using the sequence of a mature mRNA or circRNA molecule to specify the sequence of amino acids in a polypeptide chain. Translation is mediated by the ribosome, and begins with the formation of a ternary complex between aminoacylated initiator methionine tRNA, GTP, and initiation factor 2, which subsequently associates with the small subunit of the ribosome and an mRNA or circRNA. Translation ends with the release of a polypeptide chain from the ribosome or The action of a molecule that contributes to the structural integrity of the ribosome. | Burd and Dreyfuss, 1994. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN5169_c0_g1_i4 | PF12771.7^SusD-like_2 | - | - | ns | 2.4 | up | Cellular component | cell outer membrane | | SusD is a secreted starch-binding protein with an N-terminal lipid tail that allows it to associate with the outer membrane. | |
| TRINITY_DN520_c0_g2_i1 | PF13715.6^CarbopepD_reg_2 | - | ns | ns | 2.7 | up | Cellular component | integral component of membrane | bacteria, archaea or eukaryotes | This domain family is found in bacteria, archaea and eukaryotes, and is approximately 90 amino acids in length. The family is found in association with PF07715 and PF00593. | Cho and Salyers, 2001. |
| TRINITY_DN297806_c0_g2_i3 | PF00669.20^Flagellin_N | ns | 3.9 | ns | ns | up | Cellular component | membrane | bacteria | Flagellins polymerise to form bacterial flagella. This family includes flagellins and hook associated protein 3. Structurally this family forms an extended helix that interacts with PF00700. | Felix et al., 1999; Ramos et al., 2004; Beatson et al., 2006. |
| TRINITY_DN429_c0_g2_i2 | PF03401.14^TctC | - | - | - | 2.2 | up | Cellular component | membrane | bacteria | These probable extra-cytoplasmic solute receptors are strongly overrepresented in several beta-proteobacteria. This family, formerly known as Bug - Bordetella uptake gene (bug) product - is a family of bacterial tripartite tricarboxylate receptors of the extracytoplasmic solute binding receptor-dependent transporter group of families, distinct from the ABC and TRAP-T families | Antoine et al., 2003. |
| TRINITY_DN3530_c0_g1_i1 | PF03458.13^UPF0126 | - | - | ns | 5.6 | up | Cellular component | membrane | bacteria | This domain contains three transmembrane helices, with conserved glycine, which are suggestive of an ion channel. Proteins containing this domain have since been shown to be important in glycine utilizaion, and thus are predicted to be glycine transporters. This domain is always found as pair in such sequences. | Price et al., 2018. |
| TRINITY_DN1598_c0_g1_i1 | PF07963.12^N_methyl | - | - | 3.9 | ns | up | Cellular component | membrane or integral component of membrane | prokaryotes | Prokaryotic N-terminal methylation motif. This short motif directs methylation of the conserved phenylalanine residue. It is most often found at the N-terminus of pilins and other proteins involved in secretion. | |
| TRINITY_DN653_c7_g1_i1 | PF16320.5^Ribosomal_L12_N | - | ns | ns | 2.9 | up | Cellular component | ribosome | | Catalysis of the hydrolysis of the terminal or penultimate peptide bond at the C-terminal end of a peptide or polypeptide. | |
| TRINITY_DN4687_c0_g1_i2 | PF00466.20^Ribosomal_L10 | - | ns | ns | 4.5 | up | Cellular component | ribosome | | Catalysis of the hydrolysis of the terminal or penultimate peptide bond at the C-terminal end of a peptide or polypeptide. | |

4

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN2568_c0_g1_i1 | PF00466.20^Ribosomal_L10 | - | - | - | 5.4 | up | Cellular component | ribosome | | Catalysis of the hydrolysis of the terminal or penultimate peptide bond at the C-terminal end of a peptide or polypeptide. | |
| TRINITY_DN3585_c0_g1_i4 | COG0402^deaminase | - | - | ns | 3.8 | up | Molecular function | deaminase activity | | Catalysis of the removal of an amino group from a substrate, producing ammonia | |
| TRINITY_DN3461_c0_g3_i1 | PF01471.18^PG_binding_1 | - | 3.5 | - | 5.7 | up | Molecular function | binding | bacetria | Putative peptidoglycan binding domain. It is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation | Krogh et al., 1998; Steen et al., 2003; Briers et al., 2009. |
| TRINITY_DN3461_c0_g1_i1 | PF01471.18^PG_binding_1 | - | 7.2 | 6.6 | ns | up | Molecular function | binding | bacetria | Putative peptidoglycan binding domain. It is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation | Krogh et al., 1998; Steen et al., 2003; Briers et al., 2009. |
| TRINITY_DN64210_c0_g2_i1 | PF01471.18^PG_binding_1 | - | - | 6.2 | - | up | Molecular function | binding | bacetria | Putative peptidoglycan binding domain. It is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation | Krogh et al., 1998; Steen et al., 2003; Briers et al., 2009. |
| TRINITY_DN35_c0_g1_i1 | PF01471.18^PG_binding_1 | - | ns | ns | 5.7 | up | Molecular function | binding | bacetria | Putative peptidoglycan binding domain. It is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation | Krogh et al., 1998; Steen et al., 2003; Briers et al., 2009. |
| TRINITY_DN12161_c0_g4_i3 | PF01193.24^RNA_pol_L | - | - | - | 5.4 | up | Molecular function | DNA-directed 5'-3' RNA polymerase activity or protein dimerization activity | | The prokaryotic equivalent of the Rpb3/Rpb11 platform is the alpha-alpha dimer. The dimerisation domain of the alpha subunit/Rpb3 is interrupted by an insert domain (PF01000). Some of the alpha subunits also contain iron-sulphur binding domains | Zhang and Darst, 1998. |
| TRINITY_DN6187_c0_g1_i1 | COG0302^GTP cyclohydrolase i | - | - | ns | 3.4 | up | Molecular function | GTP cyclohydrolase I activity | bacteria | This family includes GTP cyclohydrolase enzymes and a family of related bacterial proteins | Nar et al., 1995. |
| TRINITY_DN193314_c0_g1_i1 | PF00557.24^Peptidase_M24 | - | - | - | 5.7 | up | Molecular function | hydrolase activity | bacteria | This family contains metallopeptidases. It also contains non-peptidase homologues such as the N terminal domain of Spt16, which is a histone H3-H4 binding module. | Stuwe et al., 2008. |
| TRINITY_DN1632_c0_g1_i1 | PF13328.6^HD_4 | - | ns | ns | 2.2 | up | Molecular function | hydrolase activity | | HD domains are metal dependent phosphohydrolases. | Aravind and Koonin, 1998. |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| TRINITY_DN2812_c0_g1_i4 | COG0791^NLP P60 protein | - | - | ns | 3.6 | up | Molecular function | hydrolase activity | | The function of this domain is unknown. It is found in several lipoproteins. | Anantharaman and Aravind, 2003. |
| TRINITY_DN9091_c0_g1_i1 | PF02874.23^ATP-synt_ab | ns | ns | 3.9 | ns | up | Molecular function | hydrolase activity or RNA binding | bacteria, archaea eukaryotes | Catalysis of the hydrolysis of various bonds, e.g. C-O, C-N, C-C, phosphoric anhydride bonds, etc. Hydrolase is the systematic name for any enzyme of EC class 3 or Interacting selectively and non-covalently with an RNA molecule. | Wilkens et al., 2005. |
| TRINITY_DN1905_c1_g2_i1 | PF03431.13^RNA_replicase_B | - | ns | 8.2 | - | up | Molecular function | RNA-directed 5'-3' RNA polymerase activity | virus | This family is of Leviviridae RNA replicases. The replicase is also known as RNA dependent RNA polymerase. | Fernando et al., 2020. |
| TRINITY_DN1761_c0_g2_i1 | PF01819.17^Levi_coat | - | ns | ns | 3.1 | up | Molecular function | viral capsid | bacteriophage | The Levivirus coat protein forms the bacteriophage coat that encapsidates the viral RNA. 180 copies of this protein form the virion shell. The MS2 bacteriophage coat protein controls two distinct processes: sequence-specific RNA encapsidation and repression of replicase translation-by binding to an RNA stem-loop structure of 19 nucleotides containing the initiation codon of the replicase gene. The binding of a coat protein dimer to this hairpin shuts off synthesis of the viral replicase, switching the viral replication cycle to virion assembly rather than continued replication | Convery et al., 1998. |
| TRINITY_DN17868_c1_g1_i9 | PF09984.9^DUF2222 | - | - | ns | 3.3 | up | | | | Domain of unknown function | |
| TRINITY_DN8253_c0_g1_i9 | PF01595.20^DUF21 | - | ns | ns | 3.4 | up | | | | Domain of unknown function | |

4

Table S2 The definition of GO terms in Fig. 7.

| GO ID | GO term | SA regulation | Catergory | Definition |
|---|---|---|---|---|
| GO:0051704 | multi-organism cellular process | up | Biological process | Interaction between organisms \| physiological interaction between organisms \| physiological interaction with another organism |
| GO:0044403 | symbiotic process | up | Biological process | A process carried out by gene products in an organism that enable the organism to engage in a symbiotic relationship, a more or less intimate association, with another organism. Microscopic symbionts are often referred to as endosymbionts. |
| GO:0044764 | multi-organism cellular process | up | Biological process | Any process that is carried out at the cellular level, which involves another organism of the same or different species. |
| GO:0019079 | viral genome replication | up | Biological process | Any process involved directly in viral genome replication, including viral nucleotide metabolism. |
| GO:0039694 | viral RNA genome replication | up | Biological process | The replication of a viral RNA genome. |
| GO:0039703 | RNA replication | up | Biological process | The cellular metabolic process in which a cell duplicates one or more molecules of RNA. |
| GO:0016032 | viral process | up | Biological process | A multi-organism process in which a virus is a participant. The other participant is the host. Includes infection of a host cell, replication of the viral genome, and assembly of progeny virus particles. In some cases, the viral genetic material may integrate into the host genome and only subsequently, under particular circumstances, 'complete' its life cycle. |
| GO:0006314 | intron homing | up | Biological process | Lateral transfer of an intron to a homologous allele that lacks the intron, mediated by a site-specific endonuclease encoded within the mobile intron. It involves with cellular macromolecule metabolic and nucleic acid metabolic process |
| GO:0044033 | multi-organism metabolic process | up | Biological process | A metabolic process - chemical reactions and pathways, including anabolism and catabolism, by which living organisms transform chemical substances, which involves more than one organism. |
| GO:0018144 | RNA-protein covalent cross-linking | up | Biological process | The formation of a covalent cross-link between RNA and a protein. It involves in cellular protein modification. |
| GO:0039666 | virion attachment to host cell pilus | up | Biological process | The process by which a virion attaches to a host cell by binding to a pilus on the host cell surface. Pili are retractile filaments that protrude from gram-negative bacteria. Filamentous viruses can attach to the pilus tip, whereas icosahedral viruses can attach to the pilus side. |
| GO:0033643 | host cell part | up | Cellular Component | Any constituent part of a host cell. The host is defined as the larger of the organisms involved in a symbiotic interaction. |
| GO:0044217 | other organism part | up | Cellular Component | Any constituent part of a secondary organism with which the first organism is interacting. |
| GO:0033646 | host intracellular part | up | Cellular Component | Any constituent part of the living contents of a host cell; the matter contained within (but not including) the plasma membrane, usually taken to exclude large vacuoles and masses of secretory or ingested material. In eukaryotes it includes the nucleus and cytoplasm. The host is defined as the larger of the organisms involved in a symbiotic interaction. |
| GO:0033655 | host cell cytoplasm part | up | Cellular Component | Any constituent part of the host cell cytoplasm, all of the contents of a cell excluding the plasma membrane and nucleus, but including other subcellular structures. The host is defined as the larger of the organisms involved in a symbiotic interaction. |
| GO:0044423 | virion part | up | Cellular Component | Any constituent part of a virion, a complete fully infectious extracellular virus particle. |
| GO:0019028 | viral capsid | up | Cellular Component | The protein coat that surrounds the infective nucleic acid in some virus particles. It comprises numerous regularly arranged subunits, or capsomeres. |
| GO:0033647 | host intracellular organelle | up | Cellular Component | Organized structure of distinctive morphology and function, occurring within the host cell. Includes the nucleus, mitochondria, plastids, vacuoles, vesicles, ribosomes and the cytoskeleton. Excludes the plasma membrane. The host is |

4

| | | | | |
|---|---|---|---|---|
| | | | | defined as the larger of the organisms involved in a symbiotic interaction. |
| GO:0033648 | host intracellular membrane-bounded organelle | up | Cellular Component | Organized structure of distinctive morphology and function, as found in host cells, bounded by a single or double lipid bilayer membrane and occurring within the cell. Includes the nucleus, mitochondria, plastids, vacuoles, and vesicles. Excludes the plasma membrane. The host is defined as the larger of the organisms involved in a symbiotic interaction. |
| GO:0033650 | host cell mitochondrion | up | Cellular Component | A semiautonomous, self-replicating organelle as found in host cells that occur in varying numbers, shapes, and sizes in the cell cytoplasm. The host is defined as the larger of the organisms involved in a symbiotic interaction. |
| GO:0019029 | helical viral capsid | up | Cellular Component | The protein coat that surrounds the infective nucleic acid in some virus particles; the subunits are arranged to form a protein helix with the genetic material contained within. Tobacco mosaic virus has such a capsid structure. |
| GO:0019030 | icosahedral viral capsid | up | Molecular function | The protein coat that surrounds the infective nucleic acid in some virus particles; the subunits are arranged to form an icosahedron, a solid with 20 faces and 12 vertices. Icosahedral capsids have 12 pentamers plus 10(T-1) hexamers, where T is the triangulation number. Tobacco satellite necrosis virus has such a capsid structure. |
| GO:0039617 | T=3 icosahedral viral capsid | up | Molecular function | The protein coat that surrounds the infective nucleic acid in some virus particles where the subunits (capsomeres) are arranged to form an icosahedron with T=3 symmetry. The T=3 capsid is composed of 12 pentameric and 20 hexameric capsomeres. |
| GO:0034062 | 5'-3' RNA polymerase activity | up | Molecular function | Catalysis of the reaction: nucleoside triphosphate + RNA (n) = diphosphate + RNA (n+1); the synthesis of RNA from ribonucleotide triphosphates in the presence of a nucleic acid template, via extension of the 3'-end. |
| GO:0097747 | RNA polymerase activity | up | Molecular function | Catalysis of the reaction: nucleoside triphosphate + RNA (n) = diphosphate + RNA (n+1); the synthesis of RNA from ribonucleotide triphosphates in the presence of a nucleic acid template. |
| GO:0016779 | nucleotidyltransferase activity | up | Molecular function | Catalysis of the transfer of a nucleotidyl group to a reactant. The upper group belongs to trasnfeerase activety. |
| GO:0003968 | RNA-directed 5'-3' RNA polymerase activity | up | Molecular function | Catalysis of the reaction: nucleoside triphosphate + RNA (n) = diphosphate + RNA (n+1); uses an RNA template, i.e., the catalysis of RNA-template-directed extension of the 3'-end of an RNA strand by one nucleotide at a time. |
| GO:0001848 | complement binding | up | Molecular function | Interacting selectively and non-covalently with any component or product of the complement cascade. |
| GO:0001846 | opsonin binding | up | Molecular function | Interacting selectively and non-covalently with an opsonin, such as a complement component or antibody, deposited on the surface of a bacteria, virus, immune complex, or other particulate material. |
| GO:0001851 | complement component C3b binding | up | Molecular function | Interacting selectively and non-covalently with the C3b product of the complement cascade. |
| GO:0051179 | localization | down | Biological process | Any process, in which a cell, a substance, or a cellular entity, such as a protein complex or organelle, is transported, tethered to or otherwise maintained in a specific location. In the case of substances, localization may also be achieved via selective degradation. |
| GO:0051234 | establishment of localization | down | Biological process | Any process that localizes a substance or cellular component. This may occur via movement, tethering or selective degradation. |
| GO:0006810 | transport | down | Biological process | The directed movement of substances (such as macromolecules, small molecules, ions) or cellular components (such as complexes and organelles) into, out of or within a cell, or between cells, or within a multicellular organism by means of some agent such as a transporter, pore or motor protein. |
| GO:0044281 | small molecule metabolic process | down | Biological process | The chemical reactions and pathways involving small molecules, any low molecular weight, monomeric, non-encoded molecule. |

| | | | | |
|---|---|---|---|---|
| GO:1901362 | organic cyclic compound biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of organic cyclic compound. |
| GO:0018130 | heterocycle biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of heterocyclic compounds, those with a cyclic molecular structure and at least two different atoms in the ring (or rings). |
| GO:0009058 | biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of substances; typically, the energy-requiring part of metabolism in which simpler substances are transformed into more complex ones. |
| GO:1901576 | organic substance biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of an organic substance, any molecular entity containing carbon. |
| GO:0044249 | cellular biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of substances, carried out by individual cells. |
| GO:1901564 | organonitrogen compound metabolic process | down | Biological process | The chemical reactions and pathways involving organonitrogen compound. |
| GO:1901566 | organonitrogen compound biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of organonitrogen compound. |
| GO:0071704 | organic substance metabolic process | down | Biological process | The chemical reactions and pathways involving an organic substance, any molecular entity containing carbon. |
| GO:0044237 | cellular metabolic process | down | Biological process | The chemical reactions and pathways by which individual cells transform chemical substances. |
| GO:0044271 | cellular nitrogen compound biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of organic and inorganic nitrogenous compounds. |
| GO:1901135 | carbohydrate derivative metabolic process | down | Biological process | The chemical reactions and pathways involving carbohydrate derivative. |
| GO:0019538 | protein metabolic process | down | Biological process | The chemical reactions and pathways involving a protein. Includes protein modification. |
| GO:0044238 | primary metabolic process | down | Biological process | The chemical reactions and pathways involving those compounds, which are formed as a part of the normal anabolic and catabolic processes. These processes take place in most, if not all, cells of the organism. |
| GO:0006807 | nitrogen compound metabolic process | down | Biological process | The chemical reactions and pathways involving organic or inorganic compounds that contain nitrogen. |
| GO:0009059 | macromolecule biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of a macromolecule, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetitions of units derived, actually or conceptually, from molecules of low relative molecular mass. |
| GO:0034645 | cellular macromolecule biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of a macromolecule, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass, carried out by individual cells. |
| GO:0019438 | aromatic compound biosynthetic process | down | Biological process | The chemical reactions and pathways resulting in the formation of aromatic compounds, any substance containing an aromatic carbon ring. |
| GO:0008152 | metabolic process | down | Biological process | Metabolic process resulting in cell growth \| metabolism \| metabolism resulting in cell growth \| multicellular organism metabolic process \| single-organism metabolic process |
| GO:1901360 | organic cyclic compound metabolic process | down | Biological process | The chemical reactions and pathways involving organic cyclic compound. |
| GO:0006725 | cellular aromatic compound metabolic process | down | Biological process | The chemical reactions and pathways involving aromatic compounds, any organic compound characterized by one or more planar rings, each of which contains conjugated double bonds and delocalized pi electrons, as carried out by individual cells. |

4

| | | | | |
|---|---|---|---|---|
| GO:0034641 | cellular nitrogen compound metabolic process | down | Biological process | The chemical reactions and pathways involving various organic and inorganic nitrogenous compounds, as carried out by individual cells. |
| GO:0046483 | heterocycle metabolic process | down | Biological process | The chemical reactions and pathways involving heterocyclic compounds, those with a cyclic molecular structure and at least two different atoms in the ring (or rings). |
| GO:0043170 | macromolecule metabolic process | down | Biological process | The chemical reactions and pathways involving macromolecules, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetitions of units derived, actually or conceptually, from molecules of low relative molecular mass. |
| GO:0009056 | catabolic process | down | Biological process | The chemical reactions and pathways resulting in the breakdown of substances, including the breakdown of carbon compounds with the liberation of energy for use by the cell or organism. |
| GO:0044260 | cellular macromolecule metabolic process | down | Biological process | The chemical reactions and pathways involving macromolecules, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass, as carried out by individual cells. |
| GO:0065007 | biological regulation | down | Biological process | Any process that modulates a measurable attribute of any biological process, quality or function. |
| GO:0006139 | nucleobase-containing compound metabolic process | down | Biological process | Any cellular metabolic process involving nucleobases, nucleosides, nucleotides and nucleic acids. |
| GO:0033646 | intracellular part | down | Cellular Component | Any constituent part of the living contents of a host cell; the matter contained within (but not including) the plasma membrane, usually taken to exclude large vacuoles and masses of secretory or ingested material. In eukaryotes it includes the nucleus and cytoplasm. The host is defined as the larger of the organisms involved in a symbiotic interaction. |
| GO:0016020 | membrane | down | Cellular Component | A lipid bilayer along with all the proteins and protein complexes embedded in it an attached to it. |
| GO:0005737 | cytoplasm | down | Cellular Component | All of the contents of a cell excluding the plasma membrane and nucleus, but including other subcellular structures. |
| GO:0005886 | plasma membrane | down | Cellular Component | The membrane surrounding a cell that separates the cell from its external environment. It consists of a phospholipid bilayer and associated proteins. |
| GO:1990904 | ribonucleoprotein complex | down | Cellular Component | A macromolecular complex containing both protein and RNA molecules. |
| GO:0032991 | protein-containing complex | down | Cellular Component | A stable assembly of two or more macromolecules, i.e., proteins, nucleic acids, carbohydrates or lipids, in which at least one component is a protein and the constituent parts function together. |
| GO:0031224 | intrinsic component of membrane | down | Cellular Component | The component of a membrane consisting of the gene products having some covalently attached portion, for example part of a peptide sequence or some other covalently attached group such as a GPI anchor, which spans or is embedded in one or both leaflets of the membrane. |
| GO:0016021 | integral component of membrane | down | Cellular Component | The component of a membrane consisting of the gene products and protein complexes having at least some part of their peptide sequence embedded in the hydrophobic region of the membrane. |
| GO:0043228 | non-membrane-bounded organelle | down | Cellular Component | Organized structure of distinctive morphology and function, not bounded by a lipid bilayer membrane. Includes ribosomes, the cytoskeleton and chromosomes. |
| GO:0005737 | cytoplasmic part | down | Cellular Component | All of the contents of a cell excluding the plasma membrane and nucleus, but including other subcellular structures. |
| GO:0043228 | non-membrane-bounded organelle | down | Cellular Component | Organized structure of distinctive morphology and function, not bounded by a lipid bilayer membrane. Includes ribosomes, the cytoskeleton and chromosomes. |
| GO:0005840 | ribosome | down | Cellular Component | An intracellular organelle, about 200 A in diameter, consisting of RNA and protein. It is the site of protein biosynthesis resulting from translation of messenger RNA (mRNA). It |

| | | | | |
|---|---|---|---|---|
| | | | | consists of two subunits, one large and one small, each containing only protein and RNA. |
| GO:0043226 | organelle | down | Cellular Component | Organized structure of distinctive morphology and function. Includes the nucleus, mitochondria, plastids, vacuoles, vesicles, ribosomes and the cytoskeleton, and prokaryotic structures such as anammoxosomes and pirellulosomes. Excludes the plasma membrane. |
| GO:0043229 | intracellular organelle | down | Cellular Component | Organized structure of distinctive morphology and function, occurring within the cell. Includes the nucleus, mitochondria, plastids, vacuoles, vesicles, ribosomes and the cytoskeleton. Excludes the plasma membrane. |
| GO:0043168 | anion binding | down | Molecular function | Interacting selectively and non-covalently with anions, charged atoms or groups of atoms with a net negative charge. |
| GO:0043167 | ion binding | down | Molecular function | Interacting selectively and non-covalently with ions, charged atoms or groups of atoms. |
| GO:0097367 | carbohydrate derivative binding | down | Molecular function | Interacting selectively and non-covalently with a carbohydrate derivative. |
| GO:0032553 | ribonucleotide binding | down | Molecular function | Interacting selectively and non-covalently with a ribonucleotide, any compound consisting of a ribonucleoside that is esterified with (ortho) phosphate or an oligophosphate at any hydroxyl group on the ribose moiety. |
| GO:0032555 | purine ribonucleotide binding | down | Molecular function | Interacting selectively and non-covalently with a purine ribonucleotide, any compound consisting of a purine ribonucleoside that is esterified with (ortho) phosphate or an oligophosphate at any hydroxyl group on the ribose moiety. |
| GO:0032557 | pyrimidine ribonucleotide binding | down | Molecular function | Interacting selectively and non-covalently with a pyrimidine ribonucleotide, any compound consisting of a pyrimidine ribonucleoside that is esterified with (ortho) phosphate or an oligophosphate at any hydroxyl group on the ribose moiety. |
| GO:0035639 | purine ribonucleoside triphosphate binding | down | Molecular function | Interacting selectively and non-covalently with a purine ribonucleoside triphosphate, a compound consisting of a purine base linked to a ribose sugar esterified with triphosphate on the sugar. |
| GO:0003677 | DNA binding | down | Molecular function | Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid). |
| GO:0016787 | hydrolase activity | down | Molecular function | Catalysis of the hydrolysis of various bonds, e.g., C-O, C-N, C-C, phosphoric anhydride bonds, etc. Hydrolase is the systematic name for any enzyme of EC class 3. |
| GO:0016818 | phosphorus-containing anhydrides | down | Molecular function | Catalysis of the hydrolysis of any acid anhydride, which contains phosphorus. |
| GO:0016462 | pyrophosphatase activity | down | Molecular function | Catalysis of the hydrolysis of a pyrophosphate bond between two phosphate groups, leaving one phosphate on each of the two fragments. |
| GO:0017111 | nucleoside-triphosphatase activity | down | Molecular function | Catalysis of the reaction: a nucleoside triphosphate + H2O = nucleoside diphosphate + phosphate. |
| GO:0005488 | binding | down | Molecular function | The selective, non-covalent, often stoichiometric, interaction of a molecule with one or more specific sites on another molecule. |
| GO:0097159 | organic cyclic compound binding | down | Molecular function | Interacting selectively and non-covalently with an organic cyclic compound, any molecular entity that contains carbon arranged in a cyclic molecular structure. |
| GO:1901363 | heterocyclic compound binding | down | Molecular function | Interacting selectively and non-covalently with heterocyclic compound. |
| GO:0003824 | catalytic activity | down | Molecular function | Catalysis of a biochemical reaction at physiological temperatures. In biologically catalyzed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzymes possess specific binding sites for substrates, and are usually composed wholly or largely of protein, but RNA that has catalytic activity (ribozyme) is often also regarded as enzymatic. |
| GO:0000166 | nucleotide binding | down | Molecular function | Interacting selectively and non-covalently with a nucleotide, any compound consisting of a nucleoside that is esterified with (ortho) phosphate or an oligophosphate at any hydroxyl group on the ribose or deoxyribose. |

| | | | | |
|---|---|---|---|---|
| GO:1901265 | nucleoside phosphate binding | down | Molecular function | Interacting selectively and non-covalently with nucleoside phosphate. |
| GO:0043169 | cation binding | down | Molecular function | Interacting selectively and non-covalently with cations, charged atoms or groups of atoms with a net positive charge. |
| GO:0046872 | metal ion binding | down | Molecular function | Interacting selectively and non-covalently with any metal ion. |