Balancing selection shapes the Intracellular Pathogen Response in natural *Caenorhabditis elegans* populations

Lisa van Sluijs[1,2], Kobus Bosman[1], Frederik Pankok[1], Tatiana Blokhina[1], Joost A. G. Riksen[1], Basten L. Snoek[1], Gorben P. Pijlman[2], Jan E. Kammenga[1,3], Mark G. Sterken[1,2,3]

[1]Laboratory of Nematology, Wageningen University and Research, 6708 PB Wageningen, the Netherlands

[2]Laboratory of Virology, Wageningen University and Research, 6708 PB Wageningen, the Netherlands

[3]Corresponding authors

Keywords:

Caenorhabditis elegans, Orsay virus, transcriptomics, Intracellular Pathogen Response (IPR), *pals*-genes, genetic variation

Corresponding authors:

Jan E. Kammenga

Wageningen University and Research, Laboratory of Nematology

Droevendaalsesteeg 1, 6708PB, Wageningen, the Netherlands

0031317482998

Jan.Kammenga@wur.nl

Mark G. Sterken

Wageningen University and Research, Laboratory of Nematology

Droevendaalsesteeg 1, 6708PB, Wageningen, the Netherlands

0031317482998

Mark.Sterken@wur.nl

Abstract

Viral infection requires a swift reaction from the host, which includes a transcriptional response. This transcriptional response is shaped by genetic variation in the host population, resulting in phenotypic differences in viral susceptibilities within the infected population. However, regulatory genetic variation in antiviral responses is not well investigated in the context of the natural host populations. Here, we infect genetically diverse isolates of the model organism *Caenorhabditis elegans* with Orsay virus to determine their transcriptional response to infection. The Hawaiian *C. elegans* isolate CB4856 shows low viral susceptibility, despite lacking the upregulation of Intracellular Pathogen Response (IPR) genes which are known to counteract viral infection. We subsequently investigated whether temporal differences in IPR timing could explain viral susceptibility, yet find that the low viral susceptibility of CB4856 unlikely results from an accelerated IPR gene activation. Instead, our data suggests that regulatory genetic variation, in particular within two key IPR regulators, determines the host transcriptional defence. Genetic analysis of 330 wild isolates reveals that CB4856 belongs to a minority of strains that show high genetic diversity within the *pals*-genes that are part of the IPR. The two IPR regulators, *pals-22* and *pals-25*, are located in a genomic region shaped by balancing selection, therefore different evolutionary strategies in IPR regulation exist. Nevertheless, the vast majority of wild isolates harbour little genetic variation in the *pals*-genes. Together, this suggests that the worldwide conservation of the IPR host transcriptional defence in *C. elegans* results from a high evolutionary pressure that pathogens could impose.

Introduction

The continuous battle between host and virus drives host genetic variation to arise in antiviral mechanisms such as transcriptional responses. Regulatory genetic variation affects the viral susceptibility after infection, making some individuals within the population more resistant than others (Wang et al. 2018; Franco et al. 2013; Piasecka et al. 2018; van Sluijs et al. 2017). Yet, the universality and mode-of-action of genetic diversity in shaping antiviral transcriptional responses within natural populations remains largely unknown.

Caenorhabditis elegans and its natural pathogen Orsay virus (OrV) are used as a powerful genetic model system to study host-virus interactions (Félix et al. 2011). OrV is a positive-sense single-stranded RNA virus infecting *C. elegans* intestinal cells where it causes local disruptions of the cellular structures (Félix et al. 2011; Franz et al. 2013). Two major groups of antiviral genes respond to viral infection in *C. elegans*: genes related to the RNA interference (RNAi) pathway (Félix et al. 2011; Tanguy et al. 2017; Sterken et al. 2014;

Sarkies et al. 2013; Ashe et al. 2013, 2015; Guo et al. 2013) and genes related to the Intracellular Pathogen Response (IPR) (Bakowski et al. 2014; Reddy et al. 2017, 2019; Chen et al. 2017). The RNAi pathway activity is controlled by the gene *sta-1* which in turn is activated by the viral sensor *sid-3* that is hypothesized to directly interact with the Orsay virus (Tanguy et al. 2017). Subsequently, the antiviral RNAi components *dcr-1*, *drh-1*, and *rde-1* degrade the viral RNA (Sarkies et al. 2013), but the RNAi genes themselves remain equally expressed during infection (Chen et al. 2017; Sarkies et al. 2013). The IPR counteracts infection by intracellular pathogens (including OrV) and increases the ability to handle proteotoxic stress (Bakowski et al. 2014; Reddy et al. 2017, 2019). The gene *pals-22* co-operates together with *pals-25* to control the IPR pathway by functioning as a molecular switch between growth and antiviral defence. *Pals-22* promotes development and lifespan, whereas *pals-25* stimulates pathogen resistance. Together *pals-22* and *pals-25* regulate a set of 80 genes that are upregulated upon intracellular infection including 25 genes in the *pals*-family and several members of the ubiquitination response (Reddy et al. 2017, 2019).. Both *pals*-22 and *pals-25* do not change gene expression following OrV infection. In total, the *pals*-gene family contains 39 members mostly found in five genetic clusters on chromosome I, III, and V.. Recently, a third antiviral defence was identified which degrades the viral genome after uridylation by the gene *cde-1* (Le Pen et al. 2018). Together, these antiviral pathways are key in controlling OrV infection in *C. elegans*.

The transcriptional responses following infection have so far been studied in the *C. elegans* laboratory strain N2, in RNAi deficient mutants in the N2 background such as *rde-1* and *dcr-1* and in the RNAi-deficient wild isolate JU1580 (Tanguy et al. 2017; Sarkies et al. 2013; Reddy et al. 2017; Chen et al. 2017; Bakowski et al. 2014; Ashe et al. 2013). These studies indicate the presence of intraspecies variation in the transcriptional response. For example, *C. elegans rde-1* and *dcr-1* mutants and JU1580 are deficient in the RNAi pathway and all show high viral susceptibility, yet their transcriptional responses differ substantially (Sarkies et al. 2013; Félix et al. 2011). Further, several *pals*-genes differ in expression levels between N2 and the RNAi mutants on one hand and JU1580 on the other hand (Sarkies et al. 2013). However, the link between the genetic background and the transcriptional responses is not fully understood. Understanding the transcriptional responses in diverse genetic backgrounds will provide insight into understanding the natural variation within the antiviral defence mechanisms.

Here we studied the transcriptional response of the *C. elegans* Hawaiian isolate CB4856 in comparison to the response in N2. CB4856 has high genetic diversity compared to N2 (Thompson, et al. 2015) and the transcriptional profile of this strain has been well-studied under multiple conditions e.g. (Snoek et al. 2017;

Capra et al. 2008; Li et al. 2006; Viñuela et al. 2012). We found that CB4856 is less susceptible to viral infection than the N2 strain, but lacks upregulation of the antiviral IPR genes. The temporal dynamics of the transcriptional response in three genetic backgrounds, N2, JU1580, and CB4856 demonstrates that gene expression patterns for multiple genes responding to OrV infection are highly dynamic, but does not show evidence for an accelerated activation of established antiviral IPR members in CB4856. The strains N2 and CB4856 differ in (basal) gene expression of several *pals*-genes, including *pals-22* and *pals-25* that control the IPR transcriptional response. Our data suggests that regulatory genetic variation within the *pals*-family underlies this difference. Contrary, most strains among a set of 330 wild isolates show little genetic variation for genes in the *pals*-family, suggesting that high selective pressure exists to conserve the IPR transcriptional response.

Materials and Methods

Nematode strains and culturing

C. elegans wild-types strains N2 (Bristol) and CB4856 (Hawaii) were kept on 6-cm Nematode Growth Medium (NGM) dishes containing *Escherichia coli* strain OP50 as food source (Brenner 1974). Strains were kept in maintenance culture at 12°C and the standard growing temperature for experiments was 20°C. Fungal and bacterial infections were cleared by bleaching (Brenner 1974). The strains were cleared of males prior to the experiments by selecting L2 larvae and placing them individually in a well in a 12-wells plate at 20°C. Thereafter, the populations were screened for male offspring after 3 days and only the 100% hermaphrodite populations were transferred to fresh 9-cm NGM dishes containing *E. coli* OP50 and grown until starved.

Orsay virus infection assay

Orsay virus stocks were prepared according to the protocol described before (Félix et al. 2011). After bleaching, nematodes were infected using 20, 50, or 100µL Orsay virus/500µL infection solution as previously described (Sterken et al. 2014). Mock infections were performed by adding M9 buffer instead of Orsay virus stock (Brenner 1974). The samples for the viral load and transcriptional analysis were infected in Eppendorf tubes with 50µL Orsay virus/500µL infection solution 26 hours post bleaching (L2-stage) (8 biological replicates per treatment per genotype). The nematodes were collected 30 hours after infection. The samples for the transcriptional analysis of the time-series were infected with 50µL Orsay virus/500µL infection solution at 40 hours post bleaching (L3-stage). The nematodes were collected at time points: 1.5, 2, 3, 8, 10, 12, 20.5, 22, 24, 28, 30.5, or 32 hours post-infection (1 biological replicate per treatment per genotype per time point). Viral loads of the samples were determined by RT-qPCR as described by (Sterken et al. 2014).

RNA isolation

The RNA of the samples in the transcriptional analysis (infected 26hpb and collected 56hpb) was isolated using Maxwell® 16 Tissue LEV Total RNA Purification Kit, Promega according to the manufacturer's instructions including two modifications. First, 10 µL proteinase K was added to the samples (instead of 25 µL). Second, after the addition of proteinase K samples were incubated at 65°C for 10 minutes while shaking at 350 rpm. Quality and quantity of the RNA were measured using the NanoDrop-1000 spectrophotometer (Thermo Scientific, Wilmington DE, USA).

The RNA of the samples in the time series was isolated using the RNeasy Micro Kit from Qiagen (Hilden, Germany). The 'Purification of Total RNA from Animal and Human Tissues' protocol was followed, with a modified lysing procedure; frozen pellets were lysed in 150 µl RLT buffer, 295 µl RNAse-free water, 800 µg/ml proteinase K and 1% ß-mercaptoethanol. The suspension was incubated at 55°C at 1000 rpm in a Thermomixer (Eppendorf, Hamburg, Germany) for 30 minutes or until the sample was clear. After this step the manufacturer's protocol was followed. Quality and quantity of the RNA were measured using the NanoDrop-1000 spectrophotometer (Thermo Scientific, Wilmington DE, USA) and RNA integrity was determined by agarose gel electrophoresis (3 µL of sample RNA on 1% agarose gel).

cDNA synthesis, labelling and hybridization

The 'Two-Color Microarray-Based Gene Expression Analysis; Low Input Quick Amp Labeling' -protocol, version 6.0 from Agilent (Agilent Technologies, Santa Clara, CA, USA) was followed, starting from step five. The *C. elegans* (V2) Gene Expression Microarray 4X44K slides, manufactured by Agilent were used.

Data extraction and normalization

The microarrays were scanned by an Agilent High Resolution C Scanner with the recommended settings. The data was extracted with Agilent Feature Extraction Software (version 10.7.1.1), following manufacturers' guidelines. Normalization of the data was executed separately for the transcriptional response data (infected at 26 and collected at 56 hours post bleaching) and the transcriptional response of the time series. For normalisation, "R" (version 3.3.1. x64) with the Limma pacakage was used. The data was not background corrected before normalization (as recommended by (Zahurak et al. 2007)). Within-array normalization was done with the Loess method and between-array normalization was done with the Quantile method (Smyth and Speed 2003). The obtained single channel normalized intensities were $\log_2$ transformed and the transcriptional response data (infected 26hpb) was batch corrected for the two different virus stocks that were used for infection. The obtained (batch corrected) $\log_2$ intensities were used for further analysis using the package 'tidyverse' (1.2.1) in "R" (3.3.1, x64).

The transcriptome datasets generated are deposited at ArrayExpress (E-MTAB-7573 and E-MTAB-7574). The data of the 12 N2 mock samples of the time series has previously been described (Snoek et al. 2015).

Principal component analysis

A principal component analysis was conducted on the gene-expression data of the both the transcriptional response and the transcriptional response of the time series. For this purpose, the data was transformed to a $\log_2$ ratio with the mean, using

$$R_{i,j} = log_2(\frac{y_{i,j}}{\bar{y}_i})$$

where R is the $\log_2$ relative expression of spot i (i = 1, 2, ..., 45220) in strain j (N2, CB4856, or JU1580), and *y* is the intensity (not the $\log_2$-transformed intensity) of spot i in strain j. The principal component analyses were performed independently per experiment. The transformed data was used in a principal component analysis, where the first six axes that explain above 4.9% of the variance were further examined.

Linear model

The $\log_2$ intensity data of the nematodes that were infected 26hpb and collected 56hpb was analysed using the linear model

$$Y_i = G + T + GxT + \varepsilon$$

with Y being the $\log_2$ normalised intensity of spot i (1, 2, ..., 45220). Y was explained over genotype (G; either N2 or CB4856), treatment (T, either infected or mock), the interaction between genotype and treatment and the error term ε. The significance threshold was determined by the *p.adjust* function, using the Benjamini & Hochberg correction (FDR < 0.1 for T and GxT, FDR < 0.05 for G) (Benjamini and Hochberg 2009).

The $\log_2$ intensity data for samples of the time series was analysed using the linear model

$$Y_i = D + G + T + \varepsilon$$

with Y being the $\log_2$ normalised intensity of spot i (1, 2, ..., 45220). Y was explained over development (D, time of isolation: 1.5, 2, 3, 8, 10, 12, 20.5, 22, 24, 28, 30.5, or 32 hours post-infection), genotype (G; either N2 or CB4856), treatment (T; either infected or mock) and the error term ε. The significance threshold was determined by the *p.adjust* function, using the Benjamini & Hochberg correction (FDR < 0.05) (Benjamini and Hochberg 2009).

Functional enrichment analysis

Gene group enrichment analyses were performed using a hypergeometric test and several databases with annotations. The databases used were: the WS258 gene class annotations, the WS258 GO-annotation, anatomy terms, phenotypes, RNAi phenotypes, developmental stage expression, and disease related genes (www.wormbase.org) (Lee et al. 2018; Stein et al. 2002); the MODENCODE release 32 transcription factor binding sites (www.modencode.org) (Gerstein et al. 2000), which were mapped to transcription start sites (as described by (Tepper et al. 2013)). Furthermore, a comparison with previously identified genes involved in OrV infection was made using a custom made database (Table S1).

Enrichments were selected based on the following criteria: size of the category n>3, size of the overlap n>2. The overlap was tested using a hypergeometric test, of which the p-values were corrected for multiple testing using Bonferroni correction (as provided by p.adjust in R, 3.3.1, x64). Enrichments were calculated based on unique gene names, not on spots.

Probe alignment

Probe sequences of the *pals*-genes (*C. elegans* (V2) Gene Expression Microarray 4X44K slides, Agilent) were aligned to the genome sequence of CB4856 (PRJNA275000) using the BLAST function of Wormbase (WS267).

Genetic variation analysis

Data on *C. elegans* wild isolates were obtained from the CeDNR website (release 20180527) (Cook et al. 2017). The data was further processed using custom made scripts (https://git.wur.nl/mark_sterken/Orsay_transcriptomics). In short, the number of polymorphisms in the *pals*-family within a strain was compared to the total number of natural polymorphisms found in that that strain. The N2 strain was used as the reference strain. A chi-square test (FDR < 0.0001) was used to determine whether strains showed less or more variation than expected within the *pals*-gene family.

The number of polymorphisms within the *pals*-gene family was further specified per gene. Tajima's D values were calculated per gene within the *C. elegans* genome using the PoPGenome package (Pfeifer et al. 2014). The number of polymorphisms within the *pals*-gene family were compared to the geographical origin of the strain obtained from the CeDNR database (Cook et al. 2017). The data were visualised using the packages 'maps' (3.3.0) and 'rworldmap' (1.3-6) using custom written scripts (https://git.wur.nl/mark_sterken/Orsay_transcriptomics).

Results

The C. elegans strains N2, JU1580, and CB4856 differ in viral susceptibility

Three strains with a different genetic background were exposed to Orsay virus (OrV) to determine their viral susceptibility. Different levels of OrV (20, 50, or 100µL OrV/500µL infection solution) were used to investigate which amount was needed to reach the maximum viral load at 30 hours after infection. The strains were infected 26 hours post bleaching and were collected as young adults before the start of egg-laying (Fig 1A). In line with previous research the strain JU1580 was more susceptible than the strain N2 (Sterken et al. 2014; Félix et al. 2011). CB4856 showed a lower viral load than N2 and JU1580, therefore being the least susceptible of the three strains (Fig 1B). The maximum viral load could be reached using 50µL OrV/500µL infection solution for all three strains which was therefore used in subsequent experiments (Fig 1B). The genotypes N2 and CB4856 were used for further transcriptional analysis.

The transcriptional response upon Orsay virus infection depends on the genetic background

We infected 26 hour old (L2) nematodes of the strains N2 and CB4856 and allowed the infection to develop for 30 hours. Young adult nematodes (before start of egg-laying) were isolated for transcriptomic analysis. In parallel, the same experiment was conducted with a mock-infection at 26 hours (Figure 1A). First, the transcriptional response of the strains N2 and CB4856 after OrV infection was analysed by means of a principal component analysis (PCA). Genotype explained the main difference in gene expression patterns (36.1%), which is in line with previous results, for example (Snoek et al. 2017; Capra et al. 2008; Li et al. 2006). The PCA showed that the effect of OrV infection is relatively small, explaining less than 5% of the total variance (Fig S1). Therefore, only a mild transcriptional response to infection is to be expected in further analyses.

Gene expression analysis by a linear model showed that 18 genes were differentially expressed upon infection by OrV (FDR < 0.1) (Fig 2A) and 15 genes were differentially expressed by a combination of both treatment and genotype (FDR < 0.1) (Fig 2B). These two groups of genes were largely overlapping (Fig 2C), because most genes only respond to infection in the genotype N2 (Fig 2D). Moreover, among the 7541 genes that were differentially expressed between N2 and CB4856 - under both mock and infected conditions – were 180 genes known to be involved in OrV infection (hypergeometric test, FDR < 0.05). As CB4856 populations showed lower viral loads, the lack of a transcriptional response may result from fewer infected nematodes or infected cells, but it may also result from a difference in (the timing of) the transcriptional response. The genes that were found to respond to OrV infection in the N2 strain were enriched for the *pals*-family gene class and the

genes were associated with dopaminergic neurons, amphid- and sheath cells (FDR < 0.05), which was in agreement with previous studies (Chen et al. 2017; Sarkies et al. 2013).

Time-dependent transcription of Orsay virus response genes

The transcriptional response of three genotypes, N2, CB4856, and JU1580, was measured over a time-course to investigate the dynamics of the transcriptional response. Nematodes were infected in the L3 stage (40 hours after bleaching) and collected 1.5-32h after infection (Fig 3A). Thereby, a high resolution timeseries of transcriptional data for three genotypes and two treatments was obtained. We hypothesised that the low viral susceptibility of CB4856 could result from early activation of IPR genes, thereby counteracting the viral infection more promptly than N2.

A PCA indicated that time of isolation explains most of the variance (38.4%) in the gene expression for all three strains, both with and without infection. This fits previous results, where the transcriptional response over the course of the L4 stage was shown to be highly dynamic (Snoek et al. 2015). The similarity between mock and infected samples suggested that the infection with OrV does not affect the transcriptional patterns of development. Genotype (either N2, CB4856, or JU1580) explained another large part of the variance (21.8%) (Fig S3). The effect of virus on global gene expression was relatively small and explained less than 4.9% of the variance.

Subsequently, the time series data was analysed using a linear model to identify variation in gene expression over time as a result of development, genotype, or infection. Over 10,000 genes were differentially expressed over time (FDR < 0.05) confirming that development was a major determinant of gene expression (Snoek et al. 2015). None of the genes were affected by infection when analysed with the linear model. Yet, the use of a linear model will only detect linear expression dynamics. Therefore, we investigated the temporal expression of several infection-responsive genes found in the N2 and CB4856 experiment in more detail (Figure 3B, Figure S3). The average increase in gene expression over time was estimated by calculating the correlation coefficient (r) for all Genes responding to OrV infection (excluding *best-24* and F58F6.3 that showed downregulation upon infection before) (Figure S3B). Gene expression for infection-responsive genes in N2 increased roughly 40% (r_{mock} = 0.00 and $r_{infected}$ = 0.40). In JU1580 there is smaller increase of about 20% (r_{mock} = -0.04 and $r_{infected}$ = 0.14). Contrary, in CB4856 genes that respond upon OrV infection in N2 became lower expressed in both mock and infected conditions over time (r_{mock} = -0.25 and $r_{infected}$ = -0.21). Subsequently, the dynamics per gene were investigated and showed distinct expression patterns depending on the genetic

background (Figure 3B, S3A). Genes expression of F26F2.1 and F26F2.4 rises gradually in the genotypes N2 and JU1580 after infection (F26F2.1: $\Delta r = 0.64$ and $\Delta r = 0.76$ respectively, F26F2.4: $\Delta r = 0.88$ and $\Delta r = 0.74$ respectively) but not in the genotype CB4856 (F26F2.1: $\Delta r = -0.06$, F26F2.4: $\Delta r = -0.19$ respectively) indicating that these IPR genes were not activated in CB4856. Also within the *pals*-gene family expression patterns were different between CB4856 and N2 and JU1580. For example, over the timeframe of 30 hours expression of the gene *pals-11* increased most in N2 ($\Delta r = 1.23$), however for CB4856 expression peaked at ~20 hours after infection reaching the highest maximum throughout the investigated period. In general, for CB4856 expression of the genes responding to OrV infection, was more dynamic than in N2 and JU1580 showing more fluctuation within the measured timeframe (Figure S3A). Together, this indicates that genetic variation can determine the developmental or infection response patterns for OrV-response genes. Of note, the IPR genes in CB4856 were similarly expressed under mock and infected conditions, even though the temporal dynamics may vary when compared to N2. Therefore, we conclude that the viral resistance of CB4856 unlikely results from an accelerated IPR, but instead results from genetic variation acting on basal gene expression.

Transcription of genes in the pals*-family is not linked to OrV infection in CB4856*

Members of the *pals*-gene family were previously shown to be involved in the IPR transcriptional response defending against OrV infection and other intracellular pathogens in the strain N2 (Sarkies et al. 2013; Chen et al. 2017; Reddy et al. 2017, 2019). However, in the strain JU1580 several of the *pals*-genes were not differentially expressed after infection (Sarkies et al. 2013) just as they were not differentially expressed in the genotype CB4856 (Fig 2C). Interestingly, the RNAi deficient strain JU1580 is highly susceptible upon viral infection (Sterken et al. 2014; Félix et al. 2011; Ashe et al. 2013), whereas CB4856 is more resistant than both JU1580 and N2 (Fig 1B). To gain a better understanding of this unexpected result, the transcriptional response in the *pals*-gene family was explored in more detail from the genotypes N2 and CB4856 30h after infection.

Many of the *pals*-gene family members became higher expressed after infection in N2, but not in the strain CB4856 (Fig 4A). In agreement with literature, the *pals*-genes in the cluster on chromosome III (*pals-17* until *pals-25*) were not differentially expressed after OrV infection (Leyva-Díaz et al. 2017; Reddy et al. 2017). Two of these genes, the paralogs *pals-22* and *pals-25*, regulate the IPR transcriptional response (Reddy et al. 2017, 2019) and for both genes the expression is higher in N2 than the CB4856 genotype under mock as well as infected conditions (Fig 4A), which may affect gene expression of other IPR members. Multiple *pals*-genes with unknown function were highly expressed under mock conditions in CB4856, but not in N2. The most extreme

example was *pals-14*. Upon infection in N2 the expression of *pals-14* increased roughly 4-fold, where after the *pals-14* expression in N2 equalled that in CB4856 under both uninfected and infected conditions (Fig 4B).

Genetic variation could underlie differences in gene expression. Therefore, the number of polymorphisms between N2 and CB4856 were calculated per *pals*-gene and compared to total genetic variation present for that gene in natural populations (Cook et al. 2017). The strain CB4856 harbours much of the total genetic variation known to occur in nature for the genomic *pals*-clusters on chromosome III and V (Cook et al. 2017) (Fig 4C). Given that our experiment used microarrays, differential expression could be due to hybridisation errors (Snoek et al. 2017; Alberts et al. 2007). Therefore, the probe sequences were compared to the genome of CB4856, showing that most probes align at the correct location (Table S2). The alignment shows that expression patterns for at least 20 out of 31 investigated *pals*-genes - including *pals-14, pals-22,* and *pals-25* - were highly likely to hybridize to the probe sequence (>95% overlap on the expected chromosomal location) (Thompson et al. 2015). The remaining 11 *pals*- genes that had low alignment scores were not taken along in further analyses (Table S2).

Most of the genetically diverse *pals*-genes on chromosome III and V have previously been shown to display local regulation of gene expression (*cis*-QTL) (Table S3). Moreover, at least 10 genes across different *pals*-clusters were regulated by genes elsewhere in the genome (*trans*-eQTL) (Table S3). Most of the eQTL were consistently found across multiple studies, environmental conditions and labs (Li et al. 2006; Rockman et al. 2010; Sterken et al. 2014; Li et al. 2010; Viñuela et al. 2010, 2012; Snoek et al. 2017). More specifically, *pals*-22 and *pals-25* that regulate expression of other IPR genes were genetically distinct and differentially expressed between N2 and CB4856 in these studies (Li et al. 2006; Rockman et al. 2010; Sterken et al. 2014; Li et al. 2010; Viñuela et al. 2010, 2012; Snoek et al. 2017). Therefore we conclude that regulatory genetic variation within *pals*-gene family members plays a role in the difference in the IPR between the strains N2 and CB4856.

The pals-*family regulating the IPR transcriptional response upon intracellular pathogen infection experiences selective pressure*

Regulatory genetic variation within the *pals*-family linked to differences in the IPR transcriptional response between N2 and CB4856. Therefore, the total genetic variation within the *pals*-family was investigated for the 330 wild isolates in the CeNDR database (Cook et al. 2017). We examined whether the *pals*-family contained genetic variation as compared to an average gene family in *C. elegans* using a chi-square test. For 48 wild isolates the genetic variation in the *pals*-family is higher than expected (chi-square test, FDR < 0.0001 and ratio

pals-gene variants/total variants > 1), but 204 isolates of the 330 analysed strains contains less than expected variation (chi-square test, FDR < 0.0001 and ratio *pals*-gene variants/total variants < 1) (Figure 5A).

To investigate the genetic diversity of the *pals*-genes within *C. elegans*, we tested if DNA sequence divergence appeared random, or that selective forces were acting on the *pals*-family by computing the Tajima's D values per gene (Figure 5B). Overall, Tajima's D values in *C. elegans* populations were low as a result of overall low genetic diversity (TD_{mean} = -1.08, TD_{median} = -1.12) (Andersen et al. 2012). Yet, four *pals*-genes (*pals-17*, *pals-18*, *pals-19*, and *pals-30*) show positive Tajima's D values that indicate either balancing selection acts on these genes or that there is a low frequency of rare alleles. The most extreme example is *pals-30* that has an Tajima's D value of 4.8, the highest value of all *C. elegans* genes. In total, 11 out of 39 *pals*-genes have values that fall within the 10% highest Tajima's D values for *C. elegans* (TD > -0.42), including IPR regulators *pals-22* and *pals-25*.

Subsequently, the genetic diversity within each *pals*-gene was analysed to explore if balancing selection might be acting on genes within this family (Figure S4, Table S4). Several *pals*-genes contain hardly any genetic variation and are therefore conserved on a worldwide scale. This conserved group contains the *pals-5* gene which acts downstream in the IPR (Reddy et al. 2017). Other *pals*-genes are highly variable and can contain hundreds of polymorphisms in a single gene. Interestingly, for most genes in this group few haplotypes exist worldwide. For example, for the gene *pals-25* strains harbour either an N2-like allele, an allele containing ~30 polymorphisms (CB4856 belongs to this group) or an allele containing ~95 polymorphisms. In total, 19 out of 24 highly variable *pals*-genes show a clear grouping within 2 or 3 haplotypes supporting that balancing selection could be acting on these genes. In conclusion, individual *pals*-genes are either conserved or few variants occur for genetically distinct *pals*-genes. In particular the *pals*-genes with a division into a few haplotypes show atypically high Tajima's D values when compared to other *C. elegans* genes. Thus, the *pals*-gene family is undergoing evolutionary selection distinct from other genes in the genome possibly resulting from balancing selection.

Some geographical locations may encounter higher selective pressures than others. Therefore, a geographical map was constructed to compare the *C. elegans* isolation location and the genetic variation found in the *pals*-genes. However, after mapping the amount of natural variation to the geographical location no clear pattern could be found (Figure S5). Interestingly, some strains that were isolated from the same location, show highly different rates of genetic diversity within the *pals*-family even though these strains may encounter similar (amounts of) pathogens. For example strain WN2002 was isolated in Wageningen and contains 3 times more

polymorphisms in the *pals*-family than in other genes. Strain WN2066 was isolated from the same compost heap as WN2002, yet contains only 0.27% variation in the *pals*-family compared to an overall genetic variation of 2.67%. Therefore, it is unclear what drives global conservation versus local divergence of the IPR transcriptional response.

Discussion

In this study we used genetically distinct strains of the nematode *C. elegans* to measure the transcriptional response upon Orsay virus infection in a genotype- and time-dependent matter. We found that genetic variation in *C. elegans* determines the Intracellular Pathogen Response (IPR): a transcriptional response that counteracts pathogens by increased proteostasis (Reddy et al. 2017, 2019). In the reference strain N2 this response leads to activation of several members in the *pals*-gene family that in turn activate ubiquitination related genes such as *skr-3*, *skr-4*, *skr-5*, and *cul-6* (Reddy et al. 2019). In contrast, genes in the *pals*-family are not upregulated in the strains JU1580 (Sarkies et al. 2013) and CB4856. The *pals*-cluster on chromosome III contains the gene pair *pals-22* and *pals-25* that together control the IPR (Reddy et al. 2019). Both *pals-22* and *pals-25* are lower expressed in CB4856 than in N2, likely underlying the difference in IPR between both strains. Multiple *cis*-QTL indicate that the expression of the *pals*-genes on chromosome III and V is controlled by local regulatory genetic variation (Li et al. 2006; Rockman et al. 2010; Sterken et al. 2014; Li et al. 2010; Viñuela et al. 2010, 2012; Snoek et al. 2017). Taken together, the genetic variation found in the *pals*-genes of CB4856 regulates the expression of IPR genes. An investigation of the worldwide variation within 330 *C. elegans* isolates reveals that the *pals*-genes show uncharacteristic patterns of genetic variation for this species. Most of the *pals*-genes are either conserved or few haplotypes exist causing their genetic conservation to be high on a worldwide scale. Population genetic analyses reveal that these genes are experiencing selective pressure which could be a result of balancing selection. Therefore, we argue that genetic variation in the *pals*-gene family regulates an evolutionary important transcriptional response to environmental stress, such as infection.

The exact cellular function of most IPR members is unclear

The function of most *pals*-genes is still unclear, even though many of them do become differentially expressed after intracellular infections. The biochemical function of the one common factor in this family, the so-called 'protein containing ALS2cr12 (ALS2CR12) signature' is still unknown. Besides functioning as an intracellular defence pathway, the *pals*-genes are involved in handling proteotoxic stress. More specifically, the gene pair *pals-22* and *pals-25* form a molecular switch between anti-stress and developmental pathways (Reddy et al. 2019). Gene expression of 22 *pals*-genes is enriched in dauer and/or L4 male expressed genes which suggests these *pals*-genes may have additional functions besides their functioning in the IPR (Gerstein et al. 2014; Leyva-Díaz et al. 2017). Indeed, one of the *pals*-genes, *pals-22*, is known to have dual biological functions. *pals-22* controls the IPR and silences repetitive RNA (Leyva-Díaz et al. 2017), although whether both functions are

executed via a similar mechanism remains unknown. Moreover, there may be functional redundancy within the *pals*-genes themselves, even though many *pals*-genes have highly dissimilar DNA and protein sequences (Leyva-Díaz et al. 2017). The gene pair *pals-22* and *pals-25* has been shown to regulate gene expression of other members in the IPR, however in *pals-22;pals-25* double mutants the other IPR genes can still respond to infection (Reddy et al. 2019).

IPR genes of the pals-*family are under selection*

Multiple *pals*-genes show high Tajima's D values, in particular the genes on the first and second cluster on chromosome III (0.1 and 1.4Mb) that all but one fall within the top 10% of genes with the highest Tajima's D values. Positive Tajima's D values indicate either balancing selection or a lack of rare alleles whereas negative Tajima's D values show that rare alleles are present at high frequencies or there has been a recent selective sweep or population bottleneck (Tajima 1989). For *C. elegans* most of the genes show negative Tajima's D values due to a recent selective sweep affecting chromosome I, IV, V, and X greatly reducing the genetic variation within the species (Andersen et al. 2012). After a selective sweep, novel genetic variation arises depending on the mutation rate and spreads depending on the outcrossing and migration rate of the species. The mutation rate in *C. elegans* is comparable to other multicellular species such as *Drosophila* (Denver et al. 2009; Haag-Liautard et al. 2007; Denver et al. 2004), however the outcrossing rate in *C. elegans* is exceptionally low. Estimated outcrossing rates in the wild range from 1-0.1% per generation (Andersen et al. 2012; Barrière and Félix 2005, 2007). In agreement, males are hardly found in wild populations and outbred populations show lower fitness (Barrière and Félix 2005; Richaud et al. 2018; Yon Rhee et al. 2008; Dolgin et al. 2007), although higher male frequencies can be beneficial under stress conditions (Teotonio et al. 2012; Morran et al. 2011, 2009b, 2009a).

As *C. elegans* mostly reproduces by selfing, genetic drift is proposed to be the main source of natural variation (Cutter 2010). In agreement, there has been little evidence for selection in *C. elegans*, but balancing selection can occur when environmental conditions determine the fitness of the strain. In that case, the benefit of the allele depends on the present environment and several alleles may be maintained actively in the population (Greene et al. 2016; Teotonio et al. 2017). The *pals*-genes with relatively high Tajima's D values on chromosome III are located in a region estimated to have diverged 10^6 generations ago. Despite this early divergence, few haplotypes occur for this region which could result from long-term balancing selection (Thompson, et al. 2015). The transcriptional regulators of the IPR, *pals-22* and *pals-25*, fall within this region

and are expected to have robust transcription within the IPR pathway for most *C. elegans* isolates in general. This is in line with our results that show only a minority of strains, including CB4856, carry distinct regulatory genetic variants. The transcriptome in *C. elegans* was shown to experience high stabilizing selection, thus favouring robust transcription of genes on a population level despite genetic variation (Denver et al. 2005). Yet, some strains potentially harbour regulatory genetic variation which may be beneficial in certain environments. Possibly, when environmental stress is present, constant growth may be preferred over the *pals-22*/*pals-25* molecular switch controlling growth and anti-stress programs. Therefore, the *pals*-gene evolution could result from an environmental factor such as pathogen presence.

Finding out which environmental factor could explain the population genetic patterns within the *pals*-genes of the IPR will be challenging. The IPR pathway has been shown to respond to multiple environmental stressors including intestinal and epidermal pathogens, but also heat stress (Reddy et al. 2019, 2017). Despite the increasing amount of ecological data for both *C. elegans* (Cook et al. 2017) and its pathogens (Richaud et al. 2018; Zhang et al. 2016), it is not yet sufficient to draw any firm conclusions whether co-occurrence of host and pathogen drives evolution within the *pals*-family. However, some evidence exists that host-pathogen interactions can affect the genotypic diversity at a population level. In Orsay (France), the location where Orsay virus is found, diversity in pathogen susceptibility potentially explains the maintenance of several minority genotypes. These genotypes are outcompeted in the absence of the intracellular pathogen *Nematocida parisii*, but perform better in the presence of the pathogen(Richaud et al. 2018). Additional transcriptional and genotypic investigation of the strains isolated in Orsay could reveal if IPR activity explains these observations. Moreover, *C. elegans* populations may experience distinct heat stress on the microscale they live in than expected based on the overall weather conditions at the sampling location (Petersen et al. 2014). Experimental evolutionary experiments hold the potential to bridge to gap between the lab and the field by investigating if the presence of intracellular pathogens or application of heat stress invokes any genetic and transcriptional changes within the *pals*-family (Gray and Cutter 2014; Teotonio et al. 2017).

The IPR response is genotype dependent

The viral susceptibility of *C. elegans* does not link directly to the IPR transcriptional response as CB4856 and JU1580 both lack upregulation of most *pals*-genes, but differ strongly in viral susceptibility. Compared to CB4856 the strain JU1580 harbours less genetic variation within the *pals*-family (11% and 6% respectively) when compared to the reference strain N2. Moreover, the expression pattern over time for *pals-22* and *pals-25*

was similar between N2 and JU1580 (data not shown). Thus, the absence of a (strong) IPR in JU1580 could have another source than in CB4856. We hypothesise that crosstalk with the (early) RNAi response is necessary to activate the IPR. Conjunction between the RNAi pathway and *pals-22* has been suggested before by results from Leyva-Díaz *et al*. showing that *pals-22* silences transgenes in wild-type animals. Yet when the RNAi gene *rde-4* is mutated, *pals-22* cannot execute its silencing function anymore (Leyva-Díaz et al. 2017). Antiviral RNAi in *C. elegans* functions via both *rde-4* dependent and *rde-4* independent mechanisms. In the *rde-4* dependent mechanism *rde-4* was proposed to process viral dsRNAs together with *drh-1* (Guo et al. 2013). A dependency of the IPR on the RNAi pathway may explain why IPR activation would be lacking in the *drh-1* deficient strain JU1580. Furthermore, most OrV transcriptional studies so far have focused on genes with a function in the IPR or the RNAi pathway. Yet transcriptional studies have yielded several genes with functions that do not (yet) link to any of the two pathways (Tanguy et al. 2017; Sarkies et al. 2013; Chen et al. 2017). The function of these genes in OrV infection could further enlighten the complexity of host-virus interactions.

CB4856 resistance to OrV does not result in a clear transcriptional response

The CB4856 strain does not show a transcriptional response to OrV once the maximum viral load is reached. There are several non-exclusive explanations as to why CB4856 does not show transcriptional changes. First, the viral loads may not be high enough to induce transcriptional changes in CB4856. Yet some of the CB4856 samples analysed contain viral loads equalling the viral loads of N2 samples that show transcriptional changes. Second, some *pals*-genes in CB4856 vary in temporal gene expression. Higher expression of IPR genes upon infection may speed up the process of counteracting viral infection to minimize its impact, thereby allowing for quick transcriptional recovery. Last, the IPR transcriptional antiviral response in CB4856 may be less efficient than that of N2, but this is compensated by other genes. N2 and CB4856 are among the most genetically diverse *C. elegans* strains resulting in a large set of different protein variants (Thompson et al. 2015). Possibly some CB4856 variants, for example one encoding an antiviral protein, could have a higher binding affinity to its target than the N2 variant. In that case, upregulation of the CB4856 variant may not be necessary to in the end still effectively counteract the viral infection.

Future perspectives

This study built a large dataset for the transcriptional response upon OrV infection in the genotypes N2 and CB4856 and temporally transcriptional data for three genotypes. For the first time the OrV transcriptional

response was measured in the genotype CB4856 and in a temporally detailed manner for this genotype and the genotypes N2 and JU1580. However, the study design also contains some limitations. First, the use of microarrays that are designed for the N2 genotype is estimated to affect about 1600 spots (3.5%) due to hybridisation differences (Snoek et al. 2017). Subsequent studies could make use of RNA-seq techniques, although this data also needs to be carefully handled to ensure that genetic variants are mapped to the correct reference genes in the N2 or CB4856 genome (The *C. elegans* Sequencing Consortium 1998; Thompson et al. 2015; Piskol et al. 2013; Deelen et al. 2015). Second, although we did not detect any transcriptional changes in the strain CB4856, there may be local changes in the intestinal cells and for certain individuals. Orsay virus infection does not infect all individuals within the population and its cellular tropism is limited to the intestinal cells (Félix et al. 2011; Franz et al. 2013). Therefore, local changes will be largely diluted in a population readout. Transcriptional techniques, such as single-cell RNA-seq or TOMO-seq (Ebbing et al. 2018; Trapnell et al. 2017), that identify gene expression within infected cells will provide more details about the transcriptional response in the cells at stake. Third, the samples for the timeseries were taken in a temporally highly detailed manner by investigating 12 timepoints within 30 hours. Yet, the effect of development is large, confounding the identification of the relatively small effect of viral infection over time. Further research investigating the effect of development on gene expression and infection over time may reveal additional information on the temporal expression of Genes responding to OrV infection.

This study provides insights into the natural context of the evolutionary conserved genetic and the plastic, transcriptional response after infection. Our results show that genetic variation within *C. elegans* natural isolates correlates to diversity in the transcriptional response upon OrV infection, thus indicating how genetic variation can shape the transcriptional response after infection. We show that few genetic variations occur worldwide within clusters of *pals*-genes that regulate the IPR transcriptional response. Therefore, we suggest that genes that function in the IPR transcriptional responses experience an evolutionary pressure such as presence of intracellular pathogens. However, the importance of the transcriptional responses appears to vary as well, as the activity of the IPR response does not directly link to the viral susceptibility. This study provides new insights into the diversity of ways that host can develop both genetic and transcriptional responses to protect themselves from harmful infections.

References

Alberts R, Terpstra P, Li Y, Breitling R, Nap JP, Jansen RC. 2007. Sequence polymorphisms cause many false cis eQTLs. *PLoS One* **2**: e622. 10.1371/journal.pone.0000622.

Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, Félix M-A, Kruglyak L. 2012. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet* **44**: 285–290. 10.1038/ng.1050.

Ashe A, Bélicard T, Le Pen J, Sarkies P, Frézal L, Lehrbach NJ, Félix MA, Miska EA. 2013. A deletion polymorphism in the *Caenorhabditis elegans* RIG-I homolog disables viral RNA dicing and antiviral immunity. *Elife* **2**: e00994. 10.7554/eLife.00994.

Ashe A, Sarkies P, Le Pen J, Tanguy M, Miska EA. 2015. Antiviral RNA Interference against Orsay Virus Is neither Systemic nor Transgenerational in *Caenorhabditis elegans*. *J Virol* **89**: 12035–12046. 10.1128/JVI.03664-14.

Bakowski MA, Desjardins CA, Smelkinson MG, Dunbar TA, Lopez-Moyado IF, Rifkin SA, Cuomo CA, Troemel ER. 2014. Ubiquitin-Mediated Response to Microsporidia and Virus Infection in *C. elegans*. *PLoS Pathog* **10**: e1004200. 10.1371/journal.ppat.1004200.

Barrière A, Félix MA. 2005. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Curr Biol* **15**: 1176–1184. 10.1016/j.cub.2005.06.022.

Barrière A, Félix MA. 2007. Temporal dynamics and linkage disequilibrium in natural *Caenorhabditis elegans* populations. *Genetics* **176**: 999–1011. 10.1534/genetics.106.067223.

Benjamini Y, Hochberg Y. 2009. Controlling the False Discovery Rate□: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc* **57**: 289–300.

Brenner S. 1974. The Genetics of *Caenorhabditis elegans*. *Genetics* **77**: 71–94. 10.1111/j.1749-6632.1999.tb07894.x.

Capra EJ, Skrovanek SM, Kruglyak L. 2008. Comparative developmental expression profiling of two *C. elegans* isolates. *PLoS One* **3**: e4055. 10.1371/journal.pone.0004055.

Chen K, Franz CJ, Jiang H, Jiang Y, Wang D. 2017. An evolutionarily conserved transcriptional response to viral infection in *Caenorhabditis* nematodes. *BMC Genomics* **18**: 303. 10.1186/s12864-017-3689-3.

Cook DE, Zdraljevic S, Roberts JP, Andersen EC. 2017. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res* **45**: D650–D657. 10.1093/nar/gkw893.

Cutter AD. 2010. Molecular evolution inferences from the *C. elegans* genome. *WormBook*. 10.1895/wormbook.1.149.1.

Deelen P, Zhernakova D V., de Haan M, van der Sijde M, Bonder MJ, Karjalainen J, van der Velde KJ, Abbott KM, Fu J, Wijmenga C, et al. 2015. Calling genotypes from public RNA-sequencing data enables identification of genetic variants that affect gene-expression levels. *Genome Med* **7**: 1–13. 10.1186/s13073-015-0152-4.

Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc Natl Acad Sci* **106**: 16310–16314. 10.1073/pnas.0904895106.

Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**: 679–682. 10.1109/T-ED.1982.21043.

Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. 2005. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat Genet* **37**: 544–548. 10.1038/ng1554.

Dolgin ES, Charlesworth B, Baird SE, Cutter AD. 2007. Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. *Evolution* **61**: 1339–1352. 10.1111/j.1558-5646.2007.00118.x.

Ebbing A, Vértesy Á, Betist MC, Spanjaard B, Junker JP, Berezikov E, van Oudenaarden A, Korswagen HC. 2018. Spatial Transcriptomics of *C. elegans* Males and Hermaphrodites Identifies Sex-Specific Differences in Gene Expression Patterns. *Dev Cell* **47**: 801–813.e6. 10.1016/j.devcel.2018.10.016.

Félix MA, Ashe A, Piffaretti J, Wu G, Nuez I, Bélicard T, Jiang Y, Zhao G, Franz CJ, Goldstein LD, et al. 2011. Natural and experimental infection of *Caenorhabditis* nematodes by novel viruses related to nodaviruses. *PLoS Biol* **9**: e1000586. 10.1371/journal.pbio.1000586.

Franco LM, Bucasas KL, Wells JM, Niño D, Wang X, Zapata GE, Arden N, Renwick A, Yu P, Quarles JM, et al. 2013. Integrative genomic analysis of the human immune response to influenza vaccination. *Elife* **2013**: 1–18. 10.7554/eLife.00299.

Franz CJ, Renshaw H, Frezal L, Jiang Y, Félix M-A, Wang D. 2013. Orsay, Santeuil and Le Blanc viruses primarily infect intestinal cells in *Caenorhabditis* nematodes. *Virology* **448**: 255–264. 10.1016/j.virol.2013.09.024.

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. 2000. Integrative Analysis of the *Caenorhabditis elegans* Genome by the modENCODE Project. *Science* **330**: 1775 LP-1787.

Gerstein MB, Rozowsky J, Yan K-K, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, et al. 2014. Comparative analysis of the transcriptome across distant species. *Nature* **512**: 445–448. 10.1038/nature13424.

Gray JC, Cutter AD. 2014. Mainstreaming *Caenorhabditis elegans* in experimental evolution. *Proc Biol Sci* **281**: 20133055. 10.1098/rspb.2013.3055.

Greene JS, Dobosiewicz M, Butcher RA, McGrath PT, Bargmann CI. 2016. Regulatory changes in two chemoreceptor genes contribute to a *Caenorhabditis elegans* QTL for foraging behavior. *Elife* **5**: 1–19. 10.7554/eLife.21454.

Guo X, Zhang R, Wang J, Lu R. 2013. Antiviral RNA Silencing Initiated in the Absence of RDE-4, a Double-Stranded RNA Binding Protein, in *Caenorhabditis elegans*. *J Virol* **87**: 10721–10729. 10.1128/JVI.01305-13.

Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Charlesworth B, Keightley PD. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* **445**: 82–85. 10.1038/nature05388.

Le Pen J, Jiang H, Di Domenico T, Kneuss E, Kosałka J, Leung C, Morgan M, Much C, Rudolph KLM, Enright AJ, et al. 2018. Terminal uridylyltransferases target RNA viruses as part of the innate immune system. *Nat Struct Mol Biol* **25**: 778–786. 10.1038/s41594-018-0106-9.

Lee RYN, Howe KL, Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Davis P, Gao S, Grove C, et al. 2018. WormBase 2017: Molting into a new stage. *Nucleic Acids Res* **46**: D869–D874. 10.1093/nar/gkx998.

Leyva-Díaz E, Stefanakis N, Carrera I, Glenwinkel L, Wang G, Driscoll M, Hobert O. 2017. Silencing of repetitive DNA is controlled by a member of an unusual *Caenorhabditis elegans* gene family. *Genetics* **207**: 529–545. 10.1534/genetics.117.300134.

Li Y, Álvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JAG, Hazendonk E, Prins P, Plasterk RHA, Jansen RC, et al. 2006. Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet* **2**: 2155–2161. 10.1371/journal.pgen.0020222.

Li Y, Breitling R, Snoek LB, Van Der Velde KJ, Swertz MA, Riksen J, Jansen RC, Kammenga JE. 2010. Global genetic robustness of the alternative splicing machinery in *Caenorhabditis elegans*. *Genetics* **186**: 405–410. 10.1534/genetics.110.119677.

Morran LT, Cappy BJ, Anderson JL, Phillips PC. 2009a. Sexual partners for the stressed: Facultative outcrossing in the self-fertilizing nematode *Caenorhabditis elegans*. *Evolution* **63**: 1473–1482. 10.1111/j.1558-5646.2009.00652.x.

Morran LT, Parmenter MD, Phillips PC. 2009b. Mutation load and rapid adaptation favour outcrossing over self-fertilization. *Nature* **462**: 350–352. 10.1038/nature08496.

Morran LT, Schmidt OG, Gelarden IA, Parrish RC, Lively CM. 2011. Running with the Red Queen: Host-parasite coevolution selects for biparental sex. *Science* **333**: 216–218. 10.1126/science.1206360.

Petersen C, Dirksen P, Prahl S, Strathmann E, Schulenburg H. 2014. The prevalence of *Caenorhabditis elegans* across 1.5 years in selected North German locations: the importance of substrate type, abiotic parameters, and Caenorhabditis competitors. *BMC Ecol* **14**: 4. 10.1186/1472-6785-14-4.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An efficient swiss army knife for population genomic analyses in R. *Mol Biol Evol* **31**: 1929–1936. 10.1093/molbev/msu136.

Piasecka B, Duffy D, Urrutia A, Quach H, Patin E, Posseme C, Bergstedt J, Charbit B, Rouilly V, MacPherson CR, et al. 2018. Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. *Proc Natl Acad Sci* **115**: E488–E497. 10.1073/pnas.1714765115.

Piskol R, Ramaswami G, Li JB. 2013. Reliable identification of genomic variants from RNA-seq data. *Am J Hum Genet* **93**: 641–651. 10.1016/j.ajhg.2013.08.008.

Reddy KC, Dror T, Sowa JN, Panek J, Chen K, Lim ES, Wang D, Troemel ER. 2017. An Intracellular Pathogen Response Pathway Promotes Proteostasis in *C. elegans*. *Curr Biol* **27**: 3544–3553.e5. 10.1016/j.cub.2017.10.009.

Reddy KC, Dror T, Underwood RS, Osman GA, Desjardins CA, Cuomo CA, Barkoulas M, Troemel ER. 2019. Antagonistic paralogs control a switch between growth and pathogen resistance in C. elegans. *Plos Pathog* **15**: e1007528. https://doi.org/10.1371/journal.ppat.1007528.

Richaud A, Zhang G, Lee D, Lee J, Félix M-AA. 2018. The Local Coexistence Pattern of Selfing Genotypes in Natural Metapopulations. *Genetics* **208**: 807–821. 10.1534/genetics.117.300564/-/DC1.1.

Rockman M V., Skrovanek SS, Kruglyak L. 2010. Selection at Linked Sites Shapes Heritable Phenotypic Variation in *C. elegans*. *Science* **330**: 372–376. 10.1126/science.1194208.

Sarkies P, Ashe A, Le Pen J, McKie MA, Miska EA. 2013. Competition between virus-derived and endogenous small RNAs regulates gene expression in *Caenorhabditis elegans*. *Genome Res* **23**: 1258–1270. 10.1101/gr.153296.112.

Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. *Methods* **31**: 265–273. 10.1016/S1046-2023(03)00155-5.

Snoek BL, Sterken MG, Bevers RPJ, Volkers RJM, van't Hof A, Brenchley R, Riksen JAG, Cossins A, Kammenga JE. 2017. Contribution of trans regulatory eQTL to cryptic genetic variation in *C. elegans*. *BMC Genomics* **18**: 1–15. 10.1186/s12864-017-3899-8.

Snoek LB, Sterken MG, Volkers RJM, Klatter M, Bosman KJ, Bevers RPJ, Riksen JAG, Smant G, Cossins AR, Kammenga JE. 2015. A rapid and massive gene expression shift marking adolescent transition in *C. elegans*. *Sci Rep* **4**: 3912. 10.1038/srep03912.

Stein L, Sternberg PW, Durbin R, Thierry-Mieg J, Spieth J. 2002. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res* **29**: 82–86. 10.1093/nar/29.1.82.

Sterken MG, Snoek LB, Bosman KJ, Daamen J, Riksen JAG, Bakker J, Pijlman GP, Kammenga JE. 2014. A heritable antiviral RNAi response limits orsay virus infection in *Caenorhabditis elegans* N2. *PLoS One* **9**: e89760. 10.1371/journal.pone.0089760.

Tajima F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* **123**: 585–595.

Tanguy M, Véron L, Stempor P, Ahringer J, Sarkies P, Miska EA. 2017. An alternative STAT signaling pathway acts in viral immunity in *Caenorhabditis elegans*. *MBio* **8**: 1–16. 10.1128/mBio.00924-17.

Teotonio H, Carvalho S, Manoel D, Roque M, Chelo IM. 2012. Evolution of outcrossing in experimental populations of *Caenorhabditis elegans*. *PLoS One* **7**: e35811. 10.1371/journal.pone.0035811.

Teotonio H, Estes S, Phillips PC, Baer CF. 2017. Experimental Evolution with *Caenorhabditis* Nematodes. *Genetics* **206**: 691–716. 10.1534/genetics.115.186288/-/DC1.1.

Tepper RG, Ashraf J, Kaletsky R, Kleemann G, Murphy CT, Bussemaker HJ. 2013. PQM-1 complements DAF-16 as a key transcriptional regulator of DAF-2-mediated development and longevity. *Cell* **154**: 676–690. 10.1016/j.cell.2013.07.006.

The *C. elegans* Sequencing Consortium. 1998. Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science* **282**: 2012–2018. 10.1126/science.282.5396.2012.

Thompson OA, Snoek LB, Nijveen H, Sterken MG, Volkers RJM, Brenchley R, van't Hof A, Bevers RPJ, Cossins AR, Yanai I, et al. 2015. Remarkably divergent regions punctuate the genome assembly of the Caenorhabditis elegans hawaiian strain CB4856. *Genetics* **200**: 975–989. 10.1534/genetics.115.175950.

Trapnell C, Furlan SN, Waterston RH, Huynh C, Cao J, Adey A, Shendure J, Steemers FJ, Daza R, Cusanovich DA, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**: 661–667. 10.1126/science.aam8940.

van Sluijs L, Pijlman GP, Kammenga JE. 2017. Why do individuals differ in viral susceptibility? A story told by model organisms. *Viruses* **9**: 1–13. 10.3390/v9100284.

Viñuela A, Snoek LB, Riksen JAG, Kammenga JE. 2012. Aging Uncouples Heritability and Expression-QTL in *Caenorhabditis elegans*. *G3 Genes|Genomes|Genetics* **2**: 597–605. 10.1534/g3.112.002212.

Viñuela A, Snoek LB, Riksen JAG, Kammenga JE. 2010. Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res* **20**: 929–937. 10.1101/gr.102160.109.

Wang L, Pittman KJ, Barker JR, Salinas RE, Stanaway IB, Williams GD, Carroll RJ, Balmat T, Ingham A, Gopalakrishnan AM, et al. 2018. An Atlas of Genetic Variation Linking Pathogen-Induced Cellular Traits to Human Disease. *Cell Host Microbe* **24**: 308–323.e6. 10.1016/j.chom.2018.07.007.

Yon Rhee S, Wood V, Dolinski K, Draghici S. 2008. Use and misuse of the gene ontology annotations. *Nat Rev Genet* **9**: 509–515. 10.1038/nrg2363.

Zahurak M, Parmigiani G, Yu W, Scharpf RB, Berman D, Schaeffer E, Shabbeer S, Cope L. 2007. Pre-processing Agilent microarray data. *BMC Bioinformatics* **8**: 1–13. 10.1186/1471-2105-8-142.

Zhang G, Sachse M, Prevost M-C, Luallen RJ, Troemel ER, Felix M-A. 2016. A Large Collection of Novel Nematode- Infecting Microsporidia and Their Diverse Interactions with Caenorhabditis elegans and Other Related Nematodes. *PLoS Pathog* **12**. 10.1371/journal.ppat.1006093.

Declarations

Availability of data and materials

All strains used can be requested from the authors. The transcriptome datasets generated are deposited at ArrayExpress (E-MTAB-7573 and E-MTAB-7574).

Competing interests

The authors declare that they have no competing interests.

Funding

LvS was funded by the NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) (824.15.006), MGS was funded by the Graduate School Production Ecology & Resource Conservation (PE&RC).

Authors' contributions

BLS, GPP, JEK and MGS conceived and designed the experiments. LvS, KB, FP, TB, JAGR, and MGS conducted the experiments. LvS and MGS conducted transcriptome and main analyses. LvS, GPP, JEK, and MGS wrote the manuscript. All authors read and provided comments on the manuscript.

Acknowledgements

The authors want to thank Erik Andersen for hosting and sharing natural variation data on CeNDR and his advice on population genetic analyses.

Figure legends

Figure 1 Viral susceptibility of *C. elegans* **N2, CB4856 and JU1580** – A) Strains were infected 26 hours post bleaching by liquid exposure to OrV for 1 hour. 30 hours after the exposure nematodes were collected and flash frozen before further processing. B) Viral loads (log_2) for the genotypes N2, CB4856 and JU1580 after exposure to 20, 50 or 100µL OrV/500µL infection solution (student *t*-test; * $p < 0.05$).

Figure 2 Differentially expressed genes upon OrV infection in *C. elegans* **N2 and CB4856** – A) Vulcano plot showing the effect of treatment on global gene expression patterns. Spots (blue) above the FDR threshold (dotted line) were considered differentially expressed. B) Vulcano plot showing the effect of the interaction between treatment and genotype on global gene expression patterns. Genes (blue) above the FDR threshold (dotted line) were considered differentially expressed. C) Venn diagram indicating the number of genes that are differentially expressed and associated with an effect treatment and/or genotype x treatment. D) Log_2 ratios for N2 mock, N2 infected, CB4856 mock and CB4856 infected samples of the 30 genes that are differentially expressed upon treatment or genotype x treatment.

Figure 3 Temporally gene expression patterns in *C. elegans* **N2, CB4856 and JU1580** – A) Strains were infected 40 hours post bleaching by liquid exposure to OrV for 1 hour. Nematodes were collected at 1.5, 2, 3, 8, 10, 12, 20.5, 22, 24, 28, 30.5, or 32 hours after exposure and flash frozen before further processing. B) Temporal gene expression patterns of (un)infected N2, CB4856 and JU1580 for a few example genes.

Figure 4 – Gene expression and genetic variation in the *pals*-gene family in *C. elegans* **N2 and CB4856** – A) Log_2 ratios of the gene expression of *pals*-genes for N2 mock, N2 infected, CB4856 mock and CB4856 infected samples. B) Log_2 intensities of four *pals*-genes in N2 mock, N2 infected, CB4856 mock and CB4856 infected samples. C) The number of natural variants that occur in CB4856 when compared to N2 as the reference strain (Cook et al. 2017). The *pals*-genes are clustered on chromosome I, III, VI, and V.

Figure 5 – Natural variation in the *C. elegans* ***pals*-gene family worldwide** – A) The percentage of genetic variants in the *pals*-gene family compared to the overall natural variation for each of the 330 wild isotypes (Cook et al. 2017). Blue dots indicate the amount of variation in the *pals*-genes is different than expected from the overall natural variation (Chi-square test, FDR < 0.0001). B) Tajima's D values per gene in the *C. elegans*

genome calculated based on the 330 wild isolates in the CeNDR database (Cook et al. 2017). Blue dots indicate Tajima's D values for *pals*-genes.

Supplementary figures

Figure S1 Principal component analysis for gene expression in (un)infected *C. elegans* **N2 and CB4856** – Principal component analysis for the gene expression data obtained for the nematodes that were infected 26 hours post bleaching and collected 30 hours post infection. PC axes that explain at least 5% of the total variation are shown. The genotype (N2 or CB4856) is indicated by colour and the treatment (mock or OrV infection) is indicated by shape.

Figure S2 Principal component analysis for gene expression in (un)infected *C. elegans* **N2, CB4856 and JU1580** – Principal component analysis for the gene expression data obtained for the nematode that were infected 40 hours post bleaching and collected at 1.5, 2, 3, 8, 10, 12, 20.5, 22, 24, 28, 30.5, or 32 hours post infection. PC axes that explain at least 4.9% of the total variation are shown. The genotype (N2, CB4856, or JU1580) is indicated by colour, treatment (mock or OrV infection) is indicated by shape and the timepoint is indicated by size.

Figure S3 Analysis of *C. elegans* **expression patterns for OrV-response genes in the time series data** – A) Gene expression patterns for N2 mock, N2 infected, CB4856 mock, CB4856 infected, JU1580 mock, and JU1580 infected nematodes of the 30 Genes responding to OrV infection that were found in dataset of infected N2 and CB4856 nematodes (infected at 26 and collected at 56 hours post bleaching). B) Correlation coefficients of the gene expression of N2 mock, N2 infected, CB4856 mock, CB4856 infected, JU1580 mock, and JU1580 infected nematodes over time per OrV response gene.

Figure S4 Genetic variation per *pals*-gene per *C. elegans* **strain** – The total number of polymorphisms within the *pals*-gene family is plotted against the number of known polymorphisms per *pals*-gene. Each dot represents a strain of the CeNDR database and the colour of the strain indicates if the *pals*-family within this strain is depleted or enriched in polymorphisms as determined by the chi-square test (FDR < 0.0001).

Figure S5 Geographical distribution of natural variation within the *C. elegans* ***pals*-gene family –** A) Location of CeNDR isolates worldwide. The amount of natural variation (%) within the *pals*-pathway is indicated by the colour of the dot. All natural isolates have been grouped in a quantile (the first quantile exhibits least natural variation in the *pals*-pathway, the fourth exhibits most natural variation). B) Zoomed in

representation of Figure 4A of the strains collected in Europe. C) Zoomed in representation of Figure 4A of the strains collected on Hawaii.

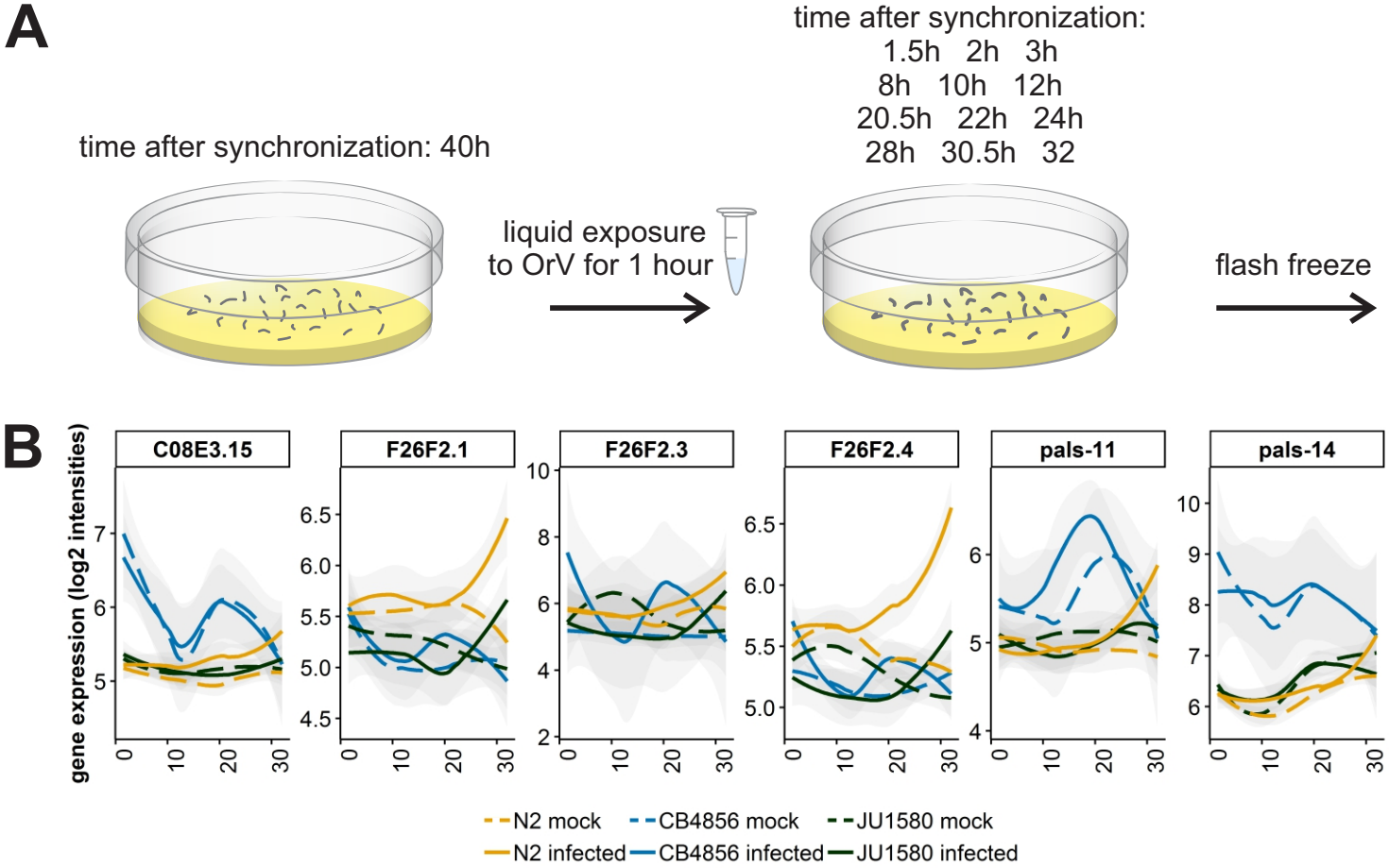
Supplementary tables

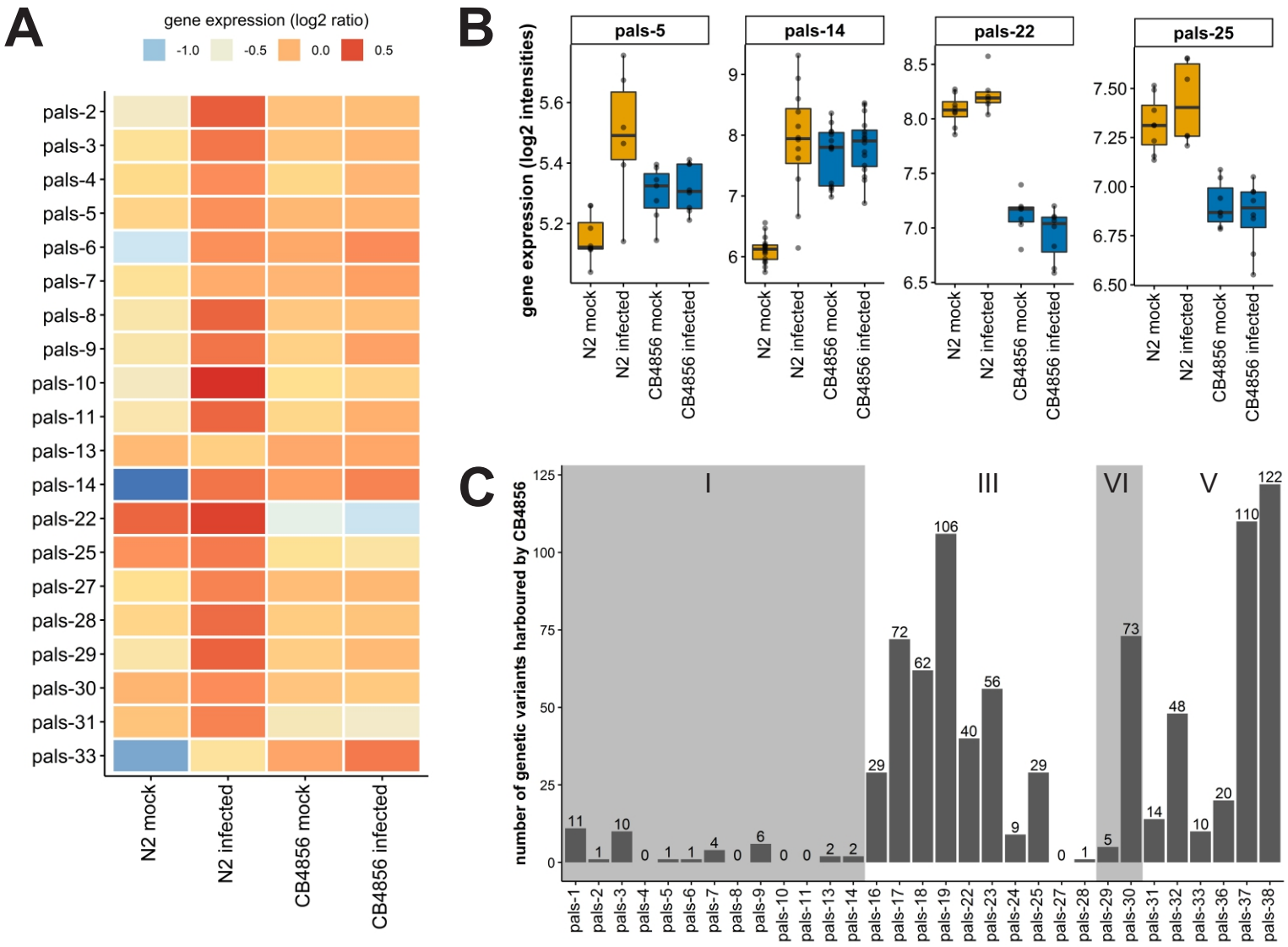
Table S1 Overview of *C. elegans* genes involved in OrV infection described in literature

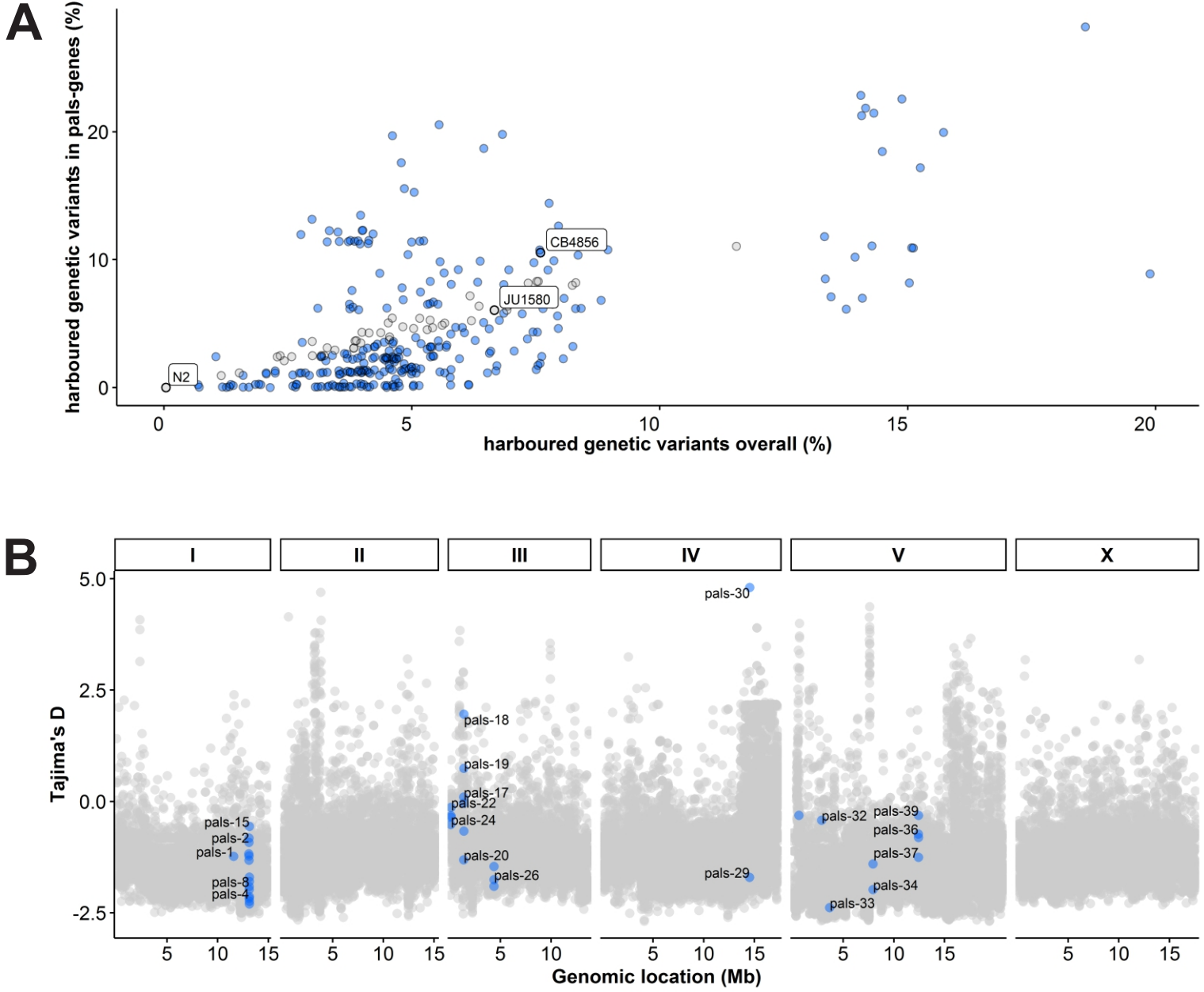
Table S2 Alignment of *C. elegans* CB4856 sequence to probes that amplify the *pals*-genes

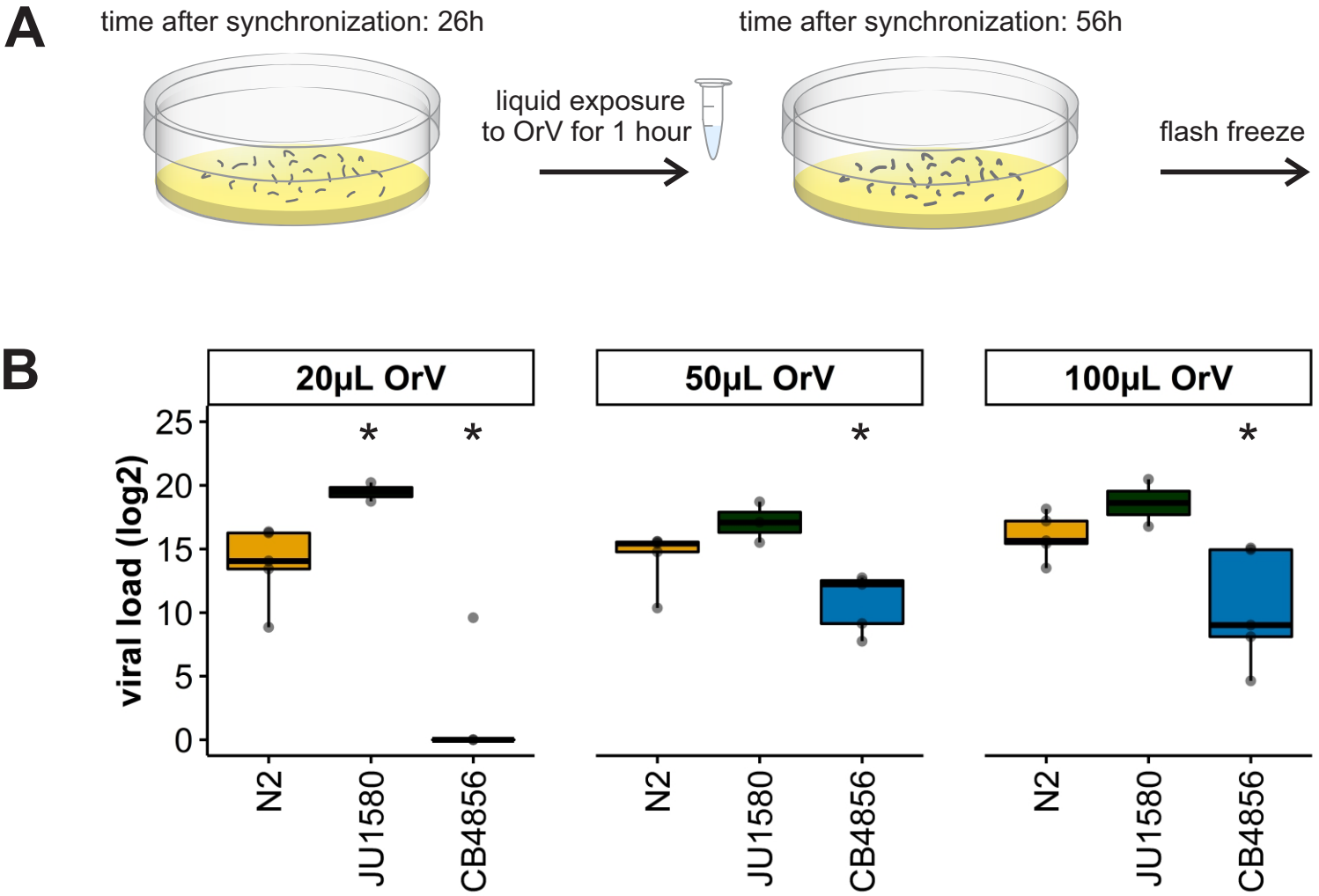
Table S3 *cis*- and *trans*-eQTL found for the *pals*-genes

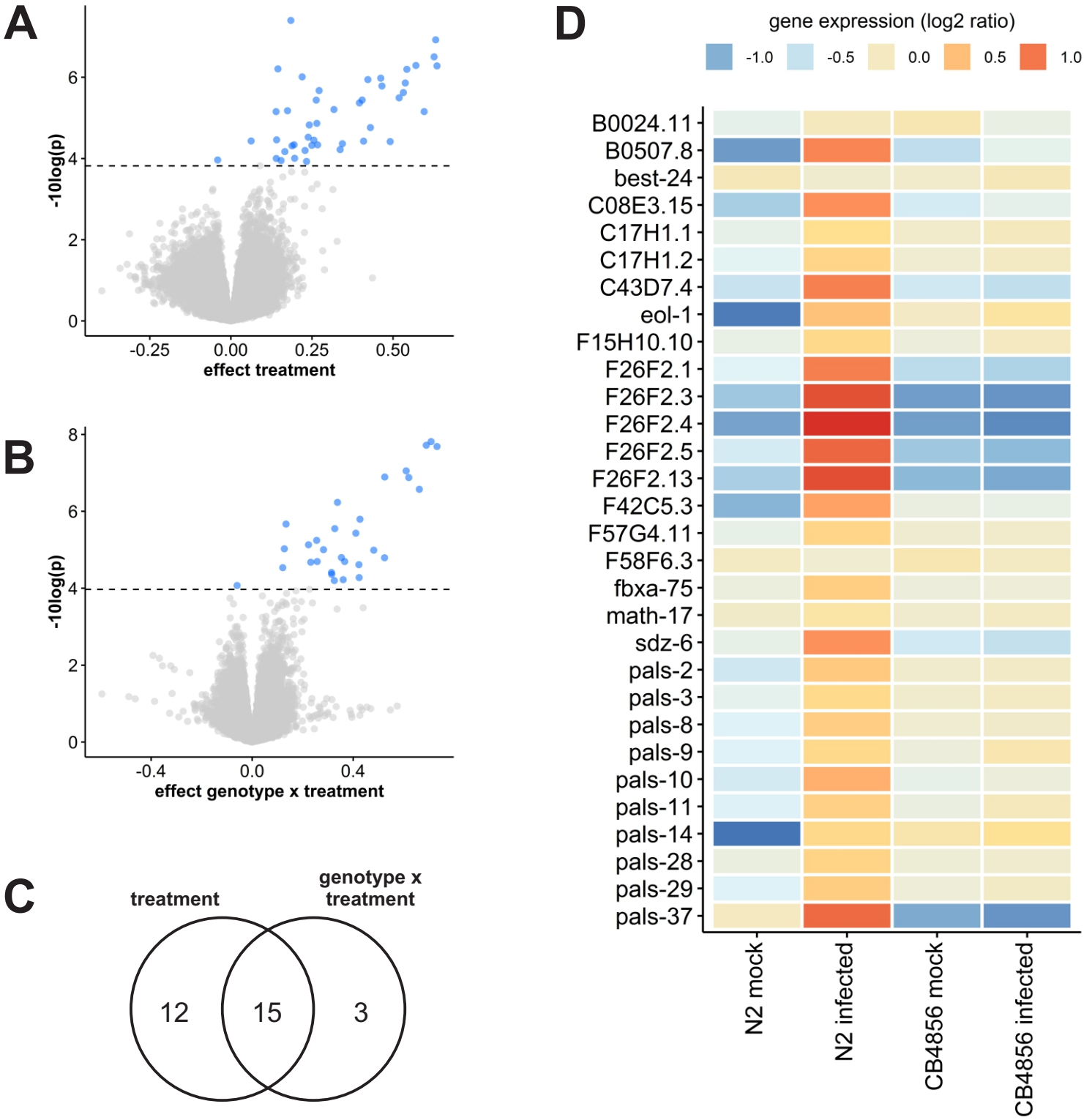
Table S4 Conservation, haplotype number and Tajima's D value per *pals*-gene


A
time after synchronization: 40h
liquid exposure
to OrV for 1 hour
time after synchronization:
1.5h 2h 3h
8h 10h 12h
20.5h 22h 24h
28h 30.5h 32
flash freeze
B
gene expression (log2 intensities)
C08E3.15
F26F2.1
F26F2.3
F26F2.4
pals-11
pals-14
N2 mock
CB4856 mock
JU1580 mock
N2 infected
CB4856 infected
JU1580 infected


A
gene expression (log2 ratio)
-1.0
-0.5
0.0
0.5
pals-2
pals-3
pals-4
pals-5
pals-6
pals-7
pals-8
pals-9
pals-10
pals-11
pals-13
pals-14
pals-22
pals-25
pals-27
pals-28
pals-29
pals-30
pals-31
pals-33
N2 mock
N2 infected
CB4856 mock
CB4856 infected
B
gene expression (log2 intensities)
pals-5
pals-14
pals-22
pals-25
C
number of genetic variants harboured by CB4856
I
III
VI
V
pals-1
pals-2
pals-3
pals-4
pals-5
pals-6
pals-7
pals-8
pals-9
pals-10
pals-11
pals-13
pals-14
pals-16
pals-17
pals-18
pals-19
pals-22
pals-23
pals-24
pals-25
pals-27
pals-28
pals-29
pals-30
pals-31
pals-32
pals-33
pals-36
pals-37
pals-38


A
harboured genetic variants in pals-genes (%)
harboured genetic variants overall (%)
N2
CB4856
JU1580
B
Tajima's D
Genomic location (Mb)
I
II
III
IV
V
X
pals-30
pals-18
pals-19
pals-17
pals-22
pals-24
pals-20
pals-26
pals-15
pals-2
pals-1
pals-8
pals-4
pals-29
pals-32
pals-39
pals-36
pals-37
pals-34
pals-33


A
time after synchronization: 26h
liquid exposure
to OrV for 1 hour
time after synchronization: 56h
flash freeze
B
20µL OrV
50µL OrV
100µL OrV
viral load (log2)
25
20
15
10
5
0
*
*
*
*
N2
JU1580
CB4856
N2
JU1580
CB4856
N2
JU1580
CB4856


A
-10log(p)
effect treatment
B
-10log(p)
effect genotype x treatment
C
treatment
genotype x treatment
12
15
3
D
gene expression (log2 ratio)
-1.0
-0.5
0.0
0.5
1.0
B0024.11
B0507.8
best-24
C08E3.15
C17H1.1
C17H1.2
C43D7.4
eol-1
F15H10.10
F26F2.1
F26F2.3
F26F2.4
F26F2.5
F26F2.13
F42C5.3
F57G4.11
F58F6.3
fbxa-75
math-17
sdz-6
pals-2
pals-3
pals-8
pals-9
pals-10
pals-11
pals-14
pals-28
pals-29
pals-37
N2 mock
N2 infected
CB4856 mock
CB4856 infected