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

Lisa van Sluijs[1,2], Kobus J. 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

Background

Genetic variation in host populations may lead to differential viral susceptibilities. Here, we investigate the role of natural genetic variation present for an antiviral pathway, the Intracellular Pathogen Response (IPR), underlying susceptibility to Orsay virus in the model organism *Caenorhabditis elegans*. The IPR involves transcriptional activity of 80 genes including the *pals*-genes. The *pals*-genes form an expanded gene family which hints they could be shaped by an evolutionary selective pressure. Here we examine the genetic variation in the *pals*-family for traces of selection and explore the molecular and phenotypic effects of having distinct *pals*-gene alleles.

Results

Genetic analysis of 330 world-wide *C. elegans* strains reveals that genetic diversity within the IPR-related *pals*-genes can be categorized in a few haplotypes worldwide. Importantly, two key-IPR regulators, *pals-22* and *pals-25*, are in a genomic region carrying signatures of balancing selection. Therefore, distinct *pals-22/pals-25* alleles have been maintained in *C. elegans* populations over time, which suggests different evolutionary strategies exist in IPR regulation. We investigated the IPR by infecting two *C. elegans* strains that represent distinct *pals-22/pals-25* haplotypes, N2 and CB4856, with Orsay virus to determine their susceptibility and transcriptional response to infection. Our data suggests that regulatory genetic variation underlies constant high activity of IPR genes in CB4856 which could determine the host transcriptional defense. We found that CB4856 shows initially lower viral susceptibility than N2. High basal IPR expression levels might help counteract viral infection directly, whereas N2-like strains that need to activate the IPR genes first may have a slower response. Nevertheless, most wild strains harbor N2-like alleles for the *pals*-genes.

Conclusions

Our work provides evidence for balancing genetic selection of immunity genes in *C. elegans* and illustrated how this may shape the transcriptional defense against pathogens. The transcriptional and genetic data presented in this study therefore provide a novel perspective on the functional diversity that can develop within a main antiviral response in natural host populations.

Background

Viral infections are a key element in all naturally occurring populations. Genetic variation can change host-virus interactions by altering coding sequences of protein products. Moreover, host-virus interactions can also be influenced by genetic variation due to altered gene copy numbers. Structural and regulatory genetic variation may both affect the viral susceptibility after infection, making some individuals within the population more resistant than others [1–4]. Presence of viruses can thereby select for beneficial genetic variants to remain present in the population [5, 6].

The nematode *Caenorhabditis elegans* and its natural pathogen Orsay virus (OrV) are used as a powerful genetic model system to study host-virus interactions [7]. OrV is a positive-sense single-stranded RNA virus infecting *C. elegans* intestinal cells where it causes local disruptions of the cellular structures [7, 8]. Three antiviral responses are known of which RNA interference (RNAi) and uridylation both target viral RNA for degradation [7, 9–11]. The third response, the so-called Intracellular Pathogen Response (IPR), is thought to relieve proteotoxic stress from the infection by the OrV and other intracellular pathogens [12–15]. The 80 genes involved in the IPR pathway are controlled by *pals-22* and *pals-25* that function as a molecular switch between growth and antiviral defense. The gene *pals-22* promotes development and lifespan, whereas *pals-25* stimulates pathogen resistance [13, 14]. Of the 80 IPR genes that become differentially expressed upon infection, 25 genes belong to the *pals*-gene family. Although the function of most *pals*-proteins remains opaque, PALS-22 and PALS-25 physically interact together and are likely to interact with additional PALS-proteins [16]. The total *pals*-gene family contains 39 members mostly found in five genetic clusters on chromosome I, III, and V [17, 18]. Both the antiviral IPR and the antiviral RNAi pathway require presence of *drh-1* that likely functions as a viral sensor [11, 19].

At present, natural populations of *C. elegans* have been isolated worldwide and catalogued into 330 isotypes maintained by the *C. elegans* Natural Diversity Resource (CeNDR) [20]. The collection contains *C. elegans* nematodes from every continent except Antarctica and each isotype in the CeNDR collection has been sequenced. Previous research found that genetic variation in the gene *drh-1* determines susceptibility to the OrV [11], yet most likely additional genetic variants can influence host-virus interactions in nature. The CeNDR database provides an ideal platform to investigate worldwide genetic variation and traces of evolutionary selection in antiviral genes in *C.* elegans.

Current studies investigating the IPR in *C. elegans* have focused on the reference genotype Bristol N2 [13, 14, 16, 19]. Here, we investigated genetic variation in the *pals*-gene family which is specifically expanded

in *C. elegans* (humans and mice only contain a single *pals*-gene ortholog) [18]. Expanded gene families often result from evolutionary selection, which strengthens the hypothesis that genetic variants in the *pals*-family could determine viral susceptibility [21]. We set out to examine if the *pals*-genes experience selective pressure by analyzing the genetic variation in 330 wild strains in the CeNDR database. We found that only a few haplotypes occur worldwide for the *pals*-genes and the some are in regions that are of ancient origin. This indicates that different pools of *pals*-genes have been maintained in *C. elegans* populations: a hallmark of balancing selection. Genetic variation in the *pals*-gene family, and specifically in the IPR-regulators *pals-22* and *pals-25*, may affect susceptibility to viral infection. This hypothesis is explored by infecting two well-studied strains, Bristol N2 and the Hawaiian strain CB4856, representing distinct *pals-22/pals-25* haplotypes. Our data suggest that regulatory genetic variation in *pals-22/pals-25* can determine (basal) IPR gene expression, illustrating that natural genetic variation in IPR genes may influence host-pathogen interactions in wild populations.

Results

Global genetic variation in the pals*-family is shaped by balancing selection*

The Intracellular Pathogen Response (IPR) counteracts viral infection in *Caenorhabditis elegans* and involves activity of 25 *pals*-genes [13, 14]. To examine genetic variation in the *pals*-family genes, we investigated sequence information from the 330 wild strains from the CeNDR database [20]. For each wild strain the genetic variation (compared to the reference strain N2) was summarized for genes in the *pals*-family and for all genes. For 48 wild strains genetic variation (defined by SNPs) in the *pals*-family was higher than expected compared to the overall genetic variation (chi-square test, FDR < 0.0001), but for 204 strains of the 330 analyzed strains the *pals*-gene family contained less variation than the overall genetic variation (chi-square test, FDR < 0.0001) (Fig 1A) (See Material and Methods for details). This indicated that while the *pals*-gene family is an expanded gene family, most wild strains contain relatively little genetic variation in the *pals*-genes compared to N2.

Populations from distinct geographical locations may encounter different selective pressures than others [22, 23]. However, after mapping the amount of natural variation to the geographical location, no clear pattern could be found (Fig S1). Interestingly, some local strains show highly diverging levels of genetic diversity within the *pals*-family. For example, strain WN2002 was isolated in Wageningen (The Netherlands) and contains 3 times more genetic variation in the *pals*-family than the average of other genes. Strain WN2066 was isolated from the same compost heap as WN2002. Yet, compared to N2, WN2066 contains only 0.27% genetic variation in the *pals*-family set against an overall genetic variation of 2.67%. This shows local diversity for the locus that might be retained in the population due to differential microenvironmental pressures.

Next, we tested whether DNA sequence divergence was governed by genetic drift, or that selective forces were possibly acting on the *pals*-family. Although overall Tajima's D (TD) values in *C. elegans* populations are low as a result of overall low genetic diversity (TD_{mean} = -1.08, TD_{median} = -1.12) [24], four *pals*-genes (*pals-17, pals-18, pals-19*, and *pals-30*) showed positive TD values that indicate either balancing selection acts on these genes or that there was a low frequency of rare alleles (Fig 1B). The most extreme example was *pals-30* that had a TD value of 4.8: the highest value of all tested *C. elegans* genes. In total, 11 out of 39 *pals*-genes had values that fall within the 10% highest TD values for *C. elegans* (TD > -0.42) and these top 10% genes included IPR regulators *pals-22* and *pals-25* (Table S1).

Subsequently, we delved into the genetic diversity for each *pals*-gene by investigating the number of variants per *pals*-gene (Fig S2, Table S1). Several *pals*-genes contained hardly any genetic variation and were therefore conserved on a worldwide scale. This conserved group contains the gene *pals-5* which acts

downstream in the IPR [13]. Other *pals*-genes were highly variable, sometimes containing hundreds of polymorphisms (SNPs) in a single gene. Interestingly, for most genes in the diverse group, few alleles exist worldwide. For example, three alleles were found for the gene *pals-25*: strains that harbor an N2-like allele, an allele containing ~30 polymorphisms (the well-studied Hawaiian strain CB4856 belongs to this group) or an allele containing ~95 polymorphisms (illustrated by the strain WN2002 from the Netherlands). In total, 19 out of 24 highly variable *pals*-genes show a clear grouping within 2 or 3 haplotypes suggesting that these haplotypes are actively maintained in the populations which supports that balancing selection could be acting on these genes (Fig S2, Table S1).

Manual inspection of the mapped reads in the 330 CeNDR strains showed evidence for extensive polymorphisms in the *pals-22*/*pals-25* locus that regulate the IPR transcriptional response [13, 14]. In total, we found three major *pals-22/pals-25* haplotypes (N2-like, CB4856-like, and WN2002-like) that occur globally (Fig 2, Table S2). The highest local genetic diversity was found on the Hawaiian islands and pacific region where *C. elegans* could originate from [25]. The genetic variation in the region where *pals-22* and *pals-25* are located is estimated to have diverged 10^6 generations ago [26]. Notably, *pals-22* and/or *pals-25* are both thought to have early stop codons in CB4856 and WN2002 that could change or disrupt their functioning [18, 20]. Yet, poor mapping to the reference genome in these highly diverse regions hampers the reliability of exact variant calling; in 24 wild strains most of the intron sequence was not covered by reads at all. The latter suggests that the genetic sequence in those strains is highly polymorphic. In conclusion, within the *pals*-gene family we found genes with either globally conserved or a few genetically distinct alleles. 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. Together, our findings indicate that the *pals*-genes have been experiencing evolutionary pressure that resulted in long-term balancing selection of these genes.

Basal pals-*gene expression differs between N2 and CB4856*

Distinct *pals*-gene profiles may result in distinct IPR activity between wild isolates. To study whether the *pals-22/pals-25* haplotypes (N2-like, 268 strains, CB4856-like, 31 strains, or WN2002-like, 24 strains) show differential gene expression for these two genes, we mined DNA and RNA hybridization data from [23]. We found that observed gene expression differences for WN2002 were mainly due to hybridization errors. However, for CB4856 hybridization was similar to N2 and shows *pals-22* expression was significantly lower than in N2 (linear model, $p < 0.001$) (Table S3). As the strains N2 and CB4856 are well-studied and their transcriptional

responses to different types of stress have been described [27–30], we continued to characterize the IPR in these two strains. We measured their transcriptional profile under standard conditions and after exposure to the OrV. These nematodes were 56h old and received either a mock or OrV infection treatment when they were in the L2 stage.

First, we focused on transcriptional differences because of genetic variation between both wild types. Expression patterns of the full dataset were analyzed 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, see for example [23, 27–29] (Fig S3). Among the 6383 genes (represented by 9379 spots) that were differentially expressed between N2 and CB4856 (under mock conditions) were 131 genes known to be involved in OrV infection (Table S4, S5A). These include eleven IPR genes of which nine show higher expression in CB4856. Moreover, six *pals*-genes showed higher basal expression in CB4856 compared to N2 (Fig 3A) (FDR < 0.05). For *pals-22* and *pals-25*, expression is higher in N2 than in CB4856 (Fig 3A) (FDR < 0.05, Table S5A), which may determine gene expression of other IPR members (Fig 3B) [14].

The strain CB4856 harbors much of the total genetic variation known to occur in nature for the genomic *pals*-clusters on chromosome III and V. Given that our experiment used microarrays, we accounted for hybridization errors, therefore we have excluded 17 IPR genes from the analysis based on probable hybridization differences (Table S6A-C). Specifically, for at least 20 out of 31 investigated *pals*-genes (including *pals-14*, *pals-22*, and *pals-25*) expression differences were likely genuine and not due to hybridization differences, because they had at least 95% overlap between probe and genome sequence and similar hybridization in N2 and CB4856 (Table S6D, E). The remaining 11 *pals*-genes with poor alignment scores were not taken along in the analyses. Furthermore, expression patterns measured by RT-qPCR confirmed the microarray data for *pals-6*, *pals-14*, and *pals-22*, although the slight difference in *pals-25* expression on the microarrays was not replicated (Fig S4). Besides, lower expression of *pals-22* in CB4856 compared to N2 was also observed in an RNA-Seq study [31].

Most of the genetically diverse *pals*-genes on chromosome III and V have previously been shown to display local regulation of gene expression (*cis*-quantitative trait locus; *cis*-eQTL) (Table S7). Moreover, at least 10 genes across different *pals*-clusters were regulated by genes elsewhere in the genome (*trans*-eQTL) (Table S7). The established IPR regulators *pals-22* and *pals-25* are likely candidates for this regulatory role. Most of the expression QTL were consistently found across multiple studies, environmental conditions and labs [10, 27, 29,

30, 32–34]. Therefore, we conclude that regulatory genetic variation of *pals*-genes could underlie observed differences in basal expression of IPR genes between the strains N2 and CB4856.

Stress-related phenotypes in the N2 and CB4856 strain

Multiple (a)biotic stressors, including heat stress and viral exposure, activate the IPR pathway [13]. Differences in IPR and *pals*-gene activity between N2 and CB4856 may determine the phenotype of stressed nematodes which was investigated here. The response of CB4856 to the OrV has so far only been measured in one experiment. This research indicated that mixed-stage populations of N2 and CB4856 that are continuously exposed to OrV for four days have comparable viral susceptibility [11]. We could confirm these findings (Fig S5), but also found that young adult CB4856 nematodes collected for our transcriptional analysis were less susceptible than N2 nematodes (Fig 4). As a control, we took along the susceptible strain JU1580, which indeed was more susceptible than the strain N2 for both assays [7, 10] (Fig 4B, Fig S5).

We investigated the cause of difference in susceptibility in N2 and CB4856 nematodes 30h post infection. As earlier studies have shown that CB4856 nematodes feed at least as much as N2 nematodes [35, 36], the initially lower susceptibility of CB4856 nematodes could not be explained by reduced food intake. Therefore, another explanation would be a different number of infected individuals or cells. Using Fluorescent *in situ* Hybridization (FISH) staining of OrV we could only visualize a few infected cells (infection detected in less than 1% of the animals), despite viral infection could be confirmed by RT-qPCR (Table S8). Contrary, almost half the nematodes showed fluorescence in a control sample of plate-infected JU1580 validating the sensitivity of the assay [11] (Table S8). Furthermore, we used FISH on plate-infected IPR reporter strains (*pals-5::*GFP and F26F2.1::GFP in an N2 background) and observed that most nematodes that showed OrV FISH signal also showed IPR gene expression (Fig S6). About 43% of the individual nematodes in the population showed IPR expression (for both reporter strains). But many individuals that showed IPR gene expression did not show OrV FISH signal (Fig S6). More specifically, 21% of the animals that showed *pals-5* expression did not show OrV stained cells (Table S8). For the F26F2.1 reporter strain even 95% of the animals showed F26F2.1 expression but no OrV stained cells (Table S8). Together, this suggests that the OrV does not reach high levels of infection in both N2 and CB4856 30h post infection, therefore low levels of OrV might already activate IPR members.

IPR genes also become differentially expressed upon prolonged heat stress providing thermotolerance in N2 [13, 14, 16]. Although the response of N2 and CB4856 to different temperatures has been well investigated, the role of IPR genes in heat stress has not been specifically compared between N2 and CB4856.

Both strains have a different temperature response as CB4856 nematodes die more often when experiencing severe heat stress, yet surviving nematodes live longer than those under standard conditions [37, 38]. Additionally, CB4856 and N2 naturally prefer different environmental temperatures and contrary to N2 ,where this preference can be conditioned, this preference is fixed in CB4856 [39–41].

The transcriptional response to IPR impulses in N2 and CB4856

We proceeded by investigating the transcriptional IPR response to OrV infection and heat stress in N2 and CB4856. The transcriptional response of age-synchronized CB4856 and N2 to heat stress was previously described by [37, 42] and we re-analyzed the data with a focus on the *pals*-genes. The transcriptomes of heat shocked and recovering nematodes were compared to the control treatment. We found that IPR genes became differentially expressed after heat-shock and during recovery in both N2 and CB4856 (Fig S7B). The mean basal IPR expression was higher in CB4856 than in N2 under control, heat shock and recovery conditions.

Subsequently, we analyzed the transcriptomes of nematodes collected 30 hours post infection (Fig 3). As expected, based on the PCA relatively few genes did respond to the OrV in our experiment (Fig S3) (Table S5B). Gene expression analysis by a linear model showed that 27 genes (represented by 57 spots) were differentially expressed upon infection by OrV (FDR < 0.1) (Fig S8A) and 18 genes (represented by 44 spots) were differentially expressed by a combination of both treatment and genotype (FDR < 0.1) (Fig S8B). These two groups of genes were largely overlapping (Fig S8C) and most of these genes only respond to infection in the genotype N2 (Fig 3B). Many of the *pals*-gene family members became higher expressed after infection in N2, but not in the strain CB4856 (Fig 3A). Some of these difference between N2 and CB4856 are due to hybridization errors in CB4856 (Table SA-E), but also 11 genes with correctly aligned probes in CB4856 did not show a response (Fig 3A, B). 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 [13, 18].

Compared to previous studies, which found 48 or 320 genes responding to OrV infection, fewer differentially expressed genes are found in this set-up which may be a result of shorter exposure to the OrV (1h in liquid here compared to constant exposure on the plate in other assays) [17, 43]. Because of the minor effect of OrV infection on transcriptional activity, a relaxed false discovery rate (FDR < 0.1) was used to analyze the data. As all genes discovered using this threshold were IPR genes that are previously described by others, these were probably true positive hits [17, 43].

Thus, N2 showed a weak, but expected IPR to OrV infection, however we did not detect increased expression of IPR genes in CB4856 nematodes. On the other hand, *pals*-genes could be activated by heat stress in CB4856. The lack of transcriptional response 30h post infection could result from early activation of IPR genes or the infection stress being too mild stress to trigger the IPR. We tested the first hypothesis by measuring IPR activity over a 30-hour time-course. This did not show evidence for an earlier IPR in CB4856 than N2 although we noticed that OrV responsive genes in CB4856 were more dynamic than in N2 (Fig S9A) (Table S5C). Because we measured higher viral loads for plate-infected CB4856 (Fig S5), we measured gene expression of *pals-6, pals-14, pals-22*, and *pals-25*, by RT-qPCR for these populations. We found that *pals-6* and *pals-14* were upregulated in CB4856, suggesting that higher pathogen pressure is sufficient to activate the IPR in this strain (Fig S9B).

Discussion

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

We have studied the effect of genetic diversity in relation to natural viral infection of the nematode *C. elegans*. Our findings show that genetic variation in *C. elegans* affects the Intracellular Pathogen Response (IPR): a transcriptional response that counteracts pathogens by increased proteostasis and in which at least 27 *pals*-genes are involved [13, 14]. The 39 members of the expanded *pals*-gene family are mostly conserved within the *C. elegans* species and the *pals*-genes for which divergent alleles do occur can be clustered into a few distinct haplotypes.

Population genetic analyses showed that IPR genes are experiencing selective pressure which could be a result of balancing selection, population bottlenecks or presence of rare genetics variants. We argue that balancing selection is the most likely cause for three reasons. First, *pals-22* and *pals-25* were experimentally validated to regulate the IPR and to balance growth and immunity [14]. Second, we observed that few major haplotypes occur for the gene-pair and most other *pals*-genes. Manual inspection of the *pals*-22/*pals-25* locus did not suggest presence of rare variants, rather the presence of a highly divergent region of ancient origin [26]. Third, presence of the *pals-22/pals-25* divergent region did not correlate with overall genetic variation, hence is unlikely to be the result of a bottleneck.

Besides *pals-22* and *pals-25*, multiple other *pals*-genes studied here show signs of balancing selection (high Tajima's D values) [44], in particular the genes on the first and second cluster on chromosome III (0.1 and 1.4Mb). In contrast, most of the genes in *C. elegans* show negative Tajima's D values due to a recent selective sweep affecting chromosome I, IV, V, and X. This greatly reduces the genetic variation within the species [24]. The *pals*-genes with relatively high Tajima's D values on chromosome III are located in a region that has diverged early in the natural history of *C. elegans* [26]. Despite this ancient divergence, few haplotypes occur for this region. Also *pals-22* and *pals-25* fall within this region and are therefore expected to generally show similar transcription patterns for the IPR pathway for most *C. elegans* strains. This is in line with our results that show only a minority of strains, including CB4856, carry distinct regulatory genetic variants.

Genetic variation in the *pals*-gene family regulates an evolutionary important transcriptional response to environmental stress, which include pathogens. Given the minor effect of OrV on fecundity [7, 11], it seems unlikely that the OrV is one of the pathogens that exert selection pressure underlying the balancing selection. However, immunity responses upon microsporidia and oomycete infection are also mediated by the IPR [12–15].

As these pathogens are lethal [45], we think that it is possible that these classes of pathogens underlie could underlie maintenance of different IPR haplotypes in natural populations. A recent example shows balancing selection in the plant genus *Capsella* also results in maintenance of ancestral genetic variation in immunity genes. The two *Capsella* species studied retained genetic variably at immunity loci, despite a recent population bottleneck and reproduction by selfing that together reduced overall genetic variation. Here, parallels can be drawn to *C. elegans*, a species that also mainly reproduces by selfing and has experienced loss of global genetic diversity as a result of a selective sweep [24]. Together, these studies show that within natural populations immunity-related genetic variation can be retained by balancing selection. Yet, also selective pressure of abiotic factors may result in balancing selection, for example in sunflower species that adapt to different environments [46]. As the IPR also responds to heat stress, abiotic factors may also contribute to balancing selection of the IPR.

Transcriptional activation of the IPR in two genetically diverse strains

Although our data shows that the strain CB4856 has multiple IPR genes with higher basal expression than in the strain N2 it is unclear whether this is linked to its response to OrV infection. We only observe lower viral accumulation in CB4856 compared to N2 during the first 30 hours of infection. During this initial period of viral infection, we did not detect upregulation of IPR genes in CB4856 compared to its basal expression. Yet, after a longer period of viral exposure, CB4856 accumulates as much virus in the population as N2 and subsequently IPR genes were also upregulated in CB4856. Our results provide novel insights in natural genetic variation in *C. elegans* immunity genes and measure the transcriptional response upon Orsay virus in two strains with genetically distinct *pals* alleles.

CB4856 shows a high basal expression of multiple IPR genes probably due to regulatory genetic variation in the *pals*-genes. Longer exposure to OrV or heat-stress is sufficient to activate the IPR in both the N2 and CB4856 strain. Therefore, a higher threshold for viral accumulation or cellular stress may prevent IPR activation in CB4856. Transgene-based expression of OrV under promotors with different expression levels may help to uncover which level of OrV is necessary for IPR activation [19, 47]. We could also show that a difference in timing of the response was unlikely by performing a time-course experiment. Because Orsay virus infection does not infect all individuals within the population and its cellular tropism is limited to the intestinal cells [7, 8], we cannot rule out there may be local changes in transcription in the intestinal cells or for certain individuals of CB4856. Transcriptional techniques, such as single-cell RNA-seq or TOMO-seq [48, 49], that

identify gene expression within infected cells could provide more details about the transcriptional response in the cells at stake.

The role of the IPR in fitness upon OrV infection remains elusive as findings in N2 show that there is a balance in spending energy on antiviral defense versus growth, it could be that laboratory selection on growth in the N2 strain has affected parts of the regulatory pathway thus trading-off constant high expression of IPR genes against traits beneficially under laboratory conditions [14, 50]. It is therefore not clear if the N2 strain that carries the world-wide most common allele, would be representative of the typical response of that allele. Still, having a plastic- instead of a fixed IPR response may be a beneficial strategy in nature. In that case, the fixed genetic response of CB4856 could be explained by strong selection for stress resistance. The hypothesis that the IPR determines immunity and growth in diverse genetic backgrounds can be tested by comparing IPR gene expression, development, and pathogen susceptibility of additional wild-strains that have an 'N2-like', 'CB4856-like', or 'WN2002-like' haplotype for the IPR genes, in particular *pals-22* and *pals-25*.

Are there alternative IPR strategies?

Strains potentially harbor regulatory genetic variation tailored to specific environments. In a harsh environment constant activity of the IPR may be preferred over low expression. 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 [13, 14]. Despite the increasing amount of ecological data for both *C. elegans* [20] and its pathogens [45, 51, 52], 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 minority genotypes are outcompeted in the absence of the intracellular pathogen *Nematocida parisii*, but perform better in the presence of the pathogen [51]. Perhaps this can also help explain our observations of divergent *pals22/pals-25* haplotypes in strains found at the same site. Moreover, experimental evolution experiments hold the potential to bridge to gap between the lab and the field by investigating if the presence of intracellular pathogens invokes any genetic and transcriptional changes within the *pals*-family [53, 54].

The 'slow' transcriptional response in the CB4856 strain may have different biological reasons. As CB4856 populations showed lower viral loads, the lack of a transcriptional response could simply result from

fewer infected nematodes or fewer infected cells per animal. Fluorescent staining of infected cells showed that only a few cells per nematode population (containing hundreds of individuals) could be visualized to be infected with the OrV, making a comparison between the two strains not trustworthy. Yet the number of plate-infected individuals that were stained was comparable to literature [43] and we could reliably detect the OrV by RT-qPCR in infected N2 and CB4856 populations. This indicates that the infection assay works properly and suggests that the viral levels per cell remained under the detection limit. Perhaps our data represents only the first cycle of viral replication as previous findings indicate that viral loads reached a plateau 30 hours post exposure with this method [10]. Later on, viral reproduction may increase further in line with observations that plate-infected N2 and CB4856 populations were equally infected. Thus, CB4856 nematodes may only be capable of limiting viral infection temporally.

Conclusions

This study provides insights into the natural context of the evolutionary conserved genetic and the plastic, transcriptional response after infection. We show that relatively little genetic diversity is found worldwide within clusters of *pals*-genes that regulate the IPR transcriptional response. In addition, the genetic diversity that exists is captured by only a few highly divergent haplotypes occurring worldwide. Therefore, we suggest that genes that function in the IPR transcriptional response could be under balancing selection, possibly from intracellular pathogens. Our results show that genetic variation within wild *C. elegans* strains could shape the basal expression of IPR genes and this may determine the transcriptional response after infection. Thus, this study provides new insights into the diversity of ways that hosts can develop both genetic and transcriptional responses to protect themselves from harmful infections.

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 [55]. The strains ERT54 (jyIs8[*pals-5p::*GFP, *myo-2p::*mCherry] X) and ERT71 (jyIs15[F26F2.1p::GFP; *myo-2*::mCherry]) were kind gifts from Emily Troemel [12–14]. 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 [55]. 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.

Short-term Orsay virus infection assay (exposure in liquid)

Orsay virus stocks were prepared according to the protocol described before [7]. After bleaching, nematodes were infected using 20, 50, or 100 µL Orsay virus/500 µL infection solution as previously described [10]. Mock infections were performed by adding M9 buffer instead of Orsay virus stock [55]. For each strain the maximum viral load was determined (4 biological replicates). The maximum viral load is the highest viral load that can be obtained for a strain and is reached when increasing amounts of virus do not significantly affect the viral load anymore (t-test, $p > 0.05$). For N2 and JU1580 20 µL of OrV sufficed to reach the maximum viral load. For CB4856 at least 50 µL OrV was needed to maximize the viral load. Hence, using 50µL OrV/500 µL infection solution the maximum viral load was obtained for all three strains which was therefore used in subsequent experiments (Fig 1B). Two virus stocks were used for these experiments: one for the first four biological replicates and one for the remaining four replicates.

The samples for the viral load and transcriptional analysis were infected in Eppendorf tubes with 50 µL 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 50 µL Orsay virus/500 µL infection solution at 40 hours post bleaching (L3-stage). These strains were infected in the L3 stage to obtain high RNA concentrations for microarray analysis also in the early samples. 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 [10]. A single Orsay virus stock was used for this experiment.

Long-term Orsay virus infection assay (exposure on plate)

Orsay virus stocks were obtained as described previously [7]. The long-term exposure assay was based on previous experiments by [7, 11]. Three young adult nematodes were placed on a plate with 20, 50, or 100 µL Orsay virus that was added to the plate shortly before transfer. M9 was added to mock-treated plates instead of Orsay virus stock. Two days after incubation part of the population was transferred to a fresh plate to prevent starvation. Four days (96 hours) after placing the first nematodes, populations were collected for RNA isolation. Viral loads of the samples were determined by RT-qPCR as described by [10]. A single Orsay virus stock was used for this experiment.

Fluorescent in situ *hybridization (FISH) of infected nematodes*

Custom Stellaris FISH Probes were designed against OrV RNA1 by utilizing the Stellaris RNA FISH Probe Designer (Biosearch Technologies, Inc., Petaluma, CA) available online at www.biosearchtech.com/stellarisdesigner. The mock-treated or infected nematodes were hybridized with the Stellaris RNA FISH Probe set labeled with CAL Fluor® Red 590 Dye (Biosearch Technologies, Inc.), following the manufacturer's instructions available online at www.biosearchtech.com/stellarisprotocols based on protocols by Raj *et al*. [56–58].

N2 and CB4856 populations were fixed 30h after infection (according to the short-term infection assay). Half of the nematodes were flash frozen to determine the viral load in the populations [10] and the other half were used in the FISH procedure. Eight biological replicates were performed for this assay. The strains JU1580, ERT54 and ERT71 were mock-treated or OrV infected by chunking nematodes from a starved to a fresh plate and adding either 50 µL M9 or OrV. These nematodes were fixed for FISH 48 hours post mock-treatment or infection. For the reporter strains (ERT54 and ERT71) three biological replicates were performed and JU1580 nematodes were infected once.

RNA isolation

The RNA of the samples in the transcriptional analysis (infected 26 hours post bleaching and collected 56 hours post bleaching) 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 package was used. The data was not background corrected before normalization (as recommended by [59]). Within-array normalization was done with the Loess method and between-array normalization was done with the Quantile method [60]. The obtained single channel normalized intensities were $\log_2$ transformed and the transcriptional response data (infected 26 hours post bleaching) 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 [61].

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) for sample j, and *y* is the intensity (not the $\log_2$-transformed intensity) of spot i for sample 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 models

The $\log_2$ intensity data of the nematodes that were mock-treated 26 hours post bleaching and collected 56 hours post bleaching was analyzed using the linear model

$$Y_i = T + \varepsilon$$

with Y being the $\log_2$ normalised intensity of spot i (1, 2, ..., 45220). Y was explained over genotype (G; either N2 or CB4856) 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) [62]. The analyzed dataset is part of the dataset containing mock-treated and OrV infected samples.

The $\log_2$ intensity data of the nematodes that were either mock-treated or infected 26 hours post bleaching and collected 56 hours post bleaching was analyzed using the linear model

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

with Y being the $\log_2$ normalized 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) [62].

The $\log_2$ intensity data for samples of the time series was analyzed 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) [62]. For the samples in the timeseries a correlation coefficient (r) was obtained by calculating the slope of gene expression over time.

The $\log_2$ intensity data of the nematodes that were exposed to heat shock (obtained from [42]) was analyzed 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 control, heat shock or recovery), 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 < 0.05) [62].

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) [63, 64]; the MODENCODE release 32 transcription factor binding sites (www.modencode.org) [65], which were mapped to transcription start sites (as described by [66]). Furthermore, a comparison with previously identified genes involved in OrV infection was made using a custom-made database (Table S4).

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 and IPR-genes (*C. elegans* (V2) Gene Expression Microarray 4X44K slides, Agilent) were aligned to the genome sequence of CB4856 (PRJNA275000) using command-line BLAST, using blastn with standard settings (Blast command line application; v2.2.28) (https://git.wur.nl/mark_sterken/Orsay_transcriptomics) [20, 26, 67]. We also compared the *pals*-genes probes to a differential hybridization experiment, to see if differences in DNA sequence explain the mRNA-based hybridization differences. Therefore, we obtained data from Volkers *et al*., 2013 (normalized data, E-MTAB-8126) and corelated the gene-expression differences with the hybridization differences [23].

Gene expression measurements by RT-qPCR

Gene expression measurements were performed on the cDNA of each of the 32 samples used in the N2 and CB4856 gene expression analysis of 30 hours of mock-treated orinfection (8 biological replicates) and on the cDNA of four N2 and CB4856 mock-treated or infected samples exposed to 50µL on plate. Gene expression was quantified by RT-qPCR using custom designed primers (*pals-6* forward 5'-TGGGTTCTGGATCAAGCAAAT-3', *pals-6* reverse 5'-TGTTCTAGAGCTGCCTGTCTCTG-3', *pals-14* forward 5'-TCGGGAAAGCATCAATGAACTGC-3', *pals-14* reverse 5'-TGTTGTGCCTCTCCTCTGCC-3', *pals-22* forward 5'-TTTTAATCTTGAAAGTGACCGCTGGG-3', *pals-22* reverse 5'-ACTCTCTGTTGTCGTCTTGCAAAATT-3', *pals-25* forward 5'-TGCAATCCGAAGATTGGTGA-3', *pals-25* reverse 5'-AAATTCTAACTTGCTCAGCATGGA-3') that overlap at least one exon-exon border to prevent unintended amplification of any remaining DNA. RT-qPCR was performed on the MyIQ using iQ SYBR Green Supermix (Biorad) and the recommended protocol. Gene expression in each sample was quantified for the gene of interest and two reference genes (Y37E3.8 and *rpl-6*) [10] in duplo.

To determine the relative gene expression, we normalized the data as in [10]. In short, we normalized the *pals*-gene expression based on the two reference genes using

$$E = \frac{Q_G}{((Q_{rpl6} / \bar{Q}_{rpl6}) + (Q_{Y37E3.8} / \bar{Q}_{Y37E3.8}))}$$

where E is the normalized gene expression, Q_G is the expression of the gene of interest, $Q_{rpl\text{-}6}$ is the expression of the reference gene *rpl-6* and $Q_{Y37E3.3}$ is the expression of the reference gene Y37E3.8.

Genetic variation analysis

Genetic data on *C. elegans* wild strains were obtained from the CeDNR website (release 20180527) [20]. 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 compared the total natural variation observed. Next, we also manually inspected the *pals-22/pals-25* locus of each of the 330 isolates via the Variant Browser tool on the CeNDR website (www.elegansvariation.org) [20]. The *pals-22/pals-25* locus could be classified in three major groups based on structural variation observed in the bam-files.

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 [68]. The number of polymorphisms within the *pals*-gene family were compared to the geographical origin of the strain obtained from the CeDNR database [20]. The data were visualized using the packages 'maps' (3.3.0) and 'rworldmap' (1.3-6) using custom written scripts (https://git.wur.nl/mark_sterken/Orsay_transcriptomics).

eQTL data analysis

The eQTL data was mined from http://www.bioinformatics.nl/EleQTL [69].

Abbreviations

FDR: False Discovery Rate

FISH: Fluorescent *In Situ* Hybridization

IPR: Intracellular Pathogen Response

OrV: Orsay virus

PCA: Principal Component Analysis

QTL: Quantitative Trait Locus

RNAi: RNA interference

RT-qPCR: Reverse Transcription-quantitative PCR

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

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).

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). NIH grant 1R01AA 026658-01 to JEK. The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. M.G.S. was supported by NWO domain Applied and Engineering Sciences VENI grant (17282).

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. Moreover, we thank Emily Troemel for generously sharing the IPR reporter strains ERT54 and ERT71. The authors like to thank Sanne van Hamond, Sophie Vromans and Marèl Scholten for their RT-qPCR primer designs and efficiency tests. We thank WormBase for being a versatile and accessible resource.

References

1. Wang L, Pittman KJ, Barker JR, Salinas RE, Stanaway IB, Williams GD, et al. An Atlas of Genetic Variation Linking Pathogen-Induced Cellular Traits to Human Disease. Cell Host Microbe. 2018;24:308-323.e6. doi:10.1016/j.chom.2018.07.007.

2. Franco LM, Bucasas KL, Wells JM, Niño D, Wang X, Zapata GE, et al. Integrative genomic analysis of the human immune response to influenza vaccination. Elife. 2013;2:e.00299. doi:10.7554/eLife.00299.

3. Piasecka B, Duffy D, Urrutia A, Quach H, Patin E, Posseme C, et al. Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. Proc Natl Acad Sci. 2018;115:E488–97. doi:10.1073/pnas.1714765115.

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

5. Enard D, Cai L, Gwennap C, Petrov DA. Viruses are a dominant driver of protein adaptation in mammals. Elife. 2016;5:e12469. doi:10.7554/eLife.12469.

6. Wilke CO, Sawyer SL. Adaptive evolution: At the mercy of viruses. Elife. 2016;5. doi:10.7554/eLife.16758.

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

8. Franz CJ, Renshaw H, Frezal L, Jiang Y, Félix M-A, Wang D. Orsay, Santeuil and Le Blanc viruses primarily infect intestinal cells in *Caenorhabditis* nematodes. Virology. 2013;448:255–64. doi:10.1016/j.virol.2013.09.024.

9. Le Pen J, Jiang H, Di Domenico T, Kneuss E, Kosałka J, Leung C, et al. Terminal uridylyltransferases target RNA viruses as part of the innate immune system. Nat Struct Mol Biol. 2018;25:778–86. doi:10.1038/s41594-018-0106-9.

10. Sterken MG, Snoek LB, Bosman KJ, Daamen J, Riksen JAG, Bakker J, et al. A heritable antiviral RNAi response limits orsay virus infection in Caenorhabditis elegans N2. PLoS One. 2014;9:e89760. doi:10.1371/journal.pone.0089760.

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

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

13. Reddy KC, Dror T, Sowa JN, Panek J, Chen K, Lim ES, et al. An Intracellular Pathogen Response Pathway Promotes Proteostasis in C. elegans. Curr Biol. 2017;27:3544-3553.e5. doi:10.1016/j.cub.2017.10.009.

14. Reddy KC, Dror T, Underwood RS, Osman GA, Desjardins CA, Cuomo CA, et al. Antagonistic paralogs control a switch between growth and pathogen resistance in C. elegans. Plos Pathog. 2019;15:e1007528. doi:10.1371/journal.ppat.1007528.

15. Osman GA, Fasseas MK, Koneru SL, Essmann CL, Kyrou K, Srinivasan MA, et al. Natural Infection of C. elegans by an Oomycete Reveals a New Pathogen-Specific Immune Response. Curr Biol. 2018;28:640-648.e5. doi:10.1016/j.cub.2018.01.029.

16. Panek J, Reddy KC, Luallen RJ, Fulzele A, Bennett EJ, Troemel ER. A newly defined cullin-RING ubiquitin ligase promotes thermotolerance as part of the Intracellular Pathogen Response in C. elegans. Proc Natl Acad Sci. 2020;:586834.

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

18. Leyva-Díaz E, Stefanakis N, Carrera I, Glenwinkel L, Wang G, Driscoll M, et al. Silencing of repetitive DNA is controlled by a member of an unusual Caenorhabditis elegans gene family. Genetics. 2017;207:529–45. doi:10.1534/genetics.117.300134.

19. Sowa JN, Jiang H, Somasundaram L, Tecle E, Xu G, Wang D, et al. The *C. elegans* RIG-I homolog DRH-1 mediates the Intracellular Pathogen Response upon viral infection. J Virol. 2019; October.

20. Cook DE, Zdraljevic S, Roberts JP, Andersen EC. CeNDR, the *Caenorhabditis elegans* natural diversity resource. Nucleic Acids Res. 2017;45:D650–7. doi:10.1093/nar/gkw893.

21. Thomas JH. Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. Genome Res. 2006;16:1017–30.

22. Sivasundar A, Hey J. Sampling from natural populations with RNAi reveals high outcrossing and population structure in Caenorhabditis elegans. Curr Biol. 2005;15:1598–602.

23. Volkers RJ, Snoek L, Hubar CJ van H, Coopman R, Chen W, Yang W, et al. Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild *Caenorhabditis elegans* populations. BMC Biol. 2013;11:93. doi:10.1186/1741-7007-11-93.

24. Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, et al. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. Nat Genet. 2012;44:285–90. doi:10.1038/ng.1050.

25. Crombie TA, Zdraljevic S, Cook DE, Tanny RE, Brady SC, Wang Y, et al. Deep sampling of Hawaiian Caenorhabditis elegans reveals high genetic diversity and admixture with global populations. Elife. 2019;8:e50465.

26. Thompson OA, Snoek LB, Nijveen H, Sterken MG, Volkers RJMM, Brenchley R, et al. Remarkably Divergent Regions Punctuate the Genome Assembly of the *Caenorhabditis elegans* Hawaiian Strain CB4856. Genetics. 2015;200:975–89. doi:10.1534/genetics.115.175950.

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

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

29. Li Y, Álvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JAG, et al. Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. PLoS Genet. 2006;2:2155–61. doi:10.1371/journal.pgen.0020222.

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

31. Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, Pajkic D, et al. Natural Variation in Gene Expression Modulates the Severity of Mutant Phenotypes. Cell. 2015;162:391–402. doi:10.1016/j.cell.2015.06.037.

32. Rockman M V., Skrovanek SS, Kruglyak L. Selection at Linked Sites Shapes Heritable Phenotypic Variation in *C. elegans*. Science (80-). 2010;330:372–6. doi:10.1126/science.1194208.

33. Li Y, Breitling R, Snoek LB, Van Der Velde KJ, Swertz MA, Riksen J, et al. Global genetic robustness of the alternative splicing machinery in Caenorhabditis elegans. Genetics. 2010;186:405–10. doi:10.1534/genetics.110.119677.

34. Viñuela A, Snoek LB, Riksen JAG, Kammenga JE. Genome-wide gene expression regulation as a function of genotype and age in C. elegans. Genome Res. 2010;20:929–37. doi:10.1101/gr.102160.109.

35. Wang YA, Snoek BL, Sterken MG, Riksen JAG, Stastna JJ, Kammenga JE, et al. Genetic background modifies phenotypic and transcriptional responses in a C. elegans model of α-synuclein toxicity. BMC Genomics. 2019;20:1–12. doi:10.1186/s12864-019-5597-1.

36. Balla KM, Andersen EC, Kruglyak L, Troemel ER. A Wild C. Elegans Strain Has Enhanced Epithelial Immunity to a Natural Microsporidian Parasite. PLoS Pathog. 2015;11:e1004583. doi:10.1371/journal.ppat.1004583.

37. Jovic K, Grilli J, Sterken MG, Snoek BL, Riksen JAG, Allesina S, et al. Transcriptome resilience predicts thermotolerance in Caenorhabditis elegans. BMC Biol. 2019;17:1–12.

38. Rodriguez M, Snoek LB, Riksen JAG, Bevers RP, Kammenga JE. Genetic variation for stress-response hormesis in C. elegans lifespan. Exp Gerontol. 2012;47:581–7. doi:10.1016/j.exger.2012.05.005.

39. Gaertner BE, Parmenter MD, Rockman M V., Kruglyak L, Phillips PC. More than the sum of its parts: A complex epistatic network underlies natural variation in thermal preference behavior in Caenorhabditis elegans. Genetics. 2012;192:1533–42.

40. Jurado P, Kodama E, Tanizawa Y, Mori I. Distinct thermal migration behaviors in response to different thermal gradients in Caenorhabditis elegans. Genes, Brain Behav. 2010;9:120–7.

41. Anderson JL, Albergotti L, Proulx S, Peden C, Huey RB, Phillips PC. Thermal preference of Caenorhabditis elegans: A null model and empirical tests. J Exp Biol. 2007;210:3107–16.

42. Jovic K, Sterken MG, Grilli J, Bevers RPJ, Rodriguez M, Riksen JAG, et al. Temporal dynamics of gene expression in heat-stressed Caenorhabditis elegans. PLoS One. 2017;12:1–16.

43. Sarkies P, Ashe A, Le Pen J, McKie MA, Miska EA. Competition between virus-derived and endogenous small RNAs regulates gene expression in *Caenorhabditis elegans*. Genome Res. 2013;23:1258–70. doi:10.1101/gr.153296.112.

44. Tajima F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics. 1989;123:585–95.

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

46. Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature. 2020. doi:10.1101/790279.

47. Jiang H, Chen K, Sandoval LE, Leung C, Wang D. An Evolutionarily Conserved Pathway Essential for Orsay Virus Infection of Caenorhabditis elegans. BMC. 2017;8:1–16.

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

49. Trapnell C, Furlan SN, Waterston RH, Huynh C, Cao J, Adey A, et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science (80-). 2017;357:661–7. doi:10.1126/science.aam8940.

50. Sterken MG, Snoek LB, Kammenga JE, Andersen EC. The laboratory domestication of Caenorhabditis elegans. Trends Genet. 2015;31:224–31. doi:10.1016/j.tig.2015.02.009.

51. Richaud A, Zhang G, Lee D, Lee J, Félix M-AA. The Local Coexistence Pattern of Selfing Genotypes in Natural Metapopulations. Genetics. 2018;208 February:807–21. doi:10.1534/genetics.117.300564/-/DC1.1.

52. Frézal L, Jung H, Tahan S, Wang D, Félix M-A. Noda-Like RNA Viruses Infecting *Caenorhabditis* Nematodes: Sympatry, Diversity, and Reassortment. J Virol. 2019;93.

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

54. Teotonio H, Estes S, Phillips PC, Baer CF. Experimental Evolution with *Caenorhabditis* Nematodes. Genetics. 2017;206:691–716.

55. Brenner S. The Genetics of Caenorhabditis elegans. Genetics. 1974;77:71–94.

56. Femino AM, Fay FS, Fogarty K, Singer RH. Visualization of Single RNA Transcripts in Situ. Science (80-). 1998;280 April:585–90.

57. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008;5:877–9.

58. Raj A, Tyagi S. Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. 1st edition. Elsevier Inc.; 2010. doi:10.1016/S0076-6879(10)72004-8.

59. Zahurak M, Parmigiani G, Yu W, Scharpf RB, Berman D, Schaeffer E, et al. Pre-processing Agilent microarray data. BMC Bioinformatics. 2007;8:1–13. doi:10.1186/1471-2105-8-142.

60. Smyth GK, Speed T. Normalization of cDNA microarray data. Methods. 2003;31:265–73. doi:10.1016/S1046-2023(03)00155-5.

61. Snoek LB, Sterken MG, Volkers RJM, Klatter M, Bosman KJ, Bevers RPJ, et al. A rapid and massive gene expression shift marking adolescent transition in *C. elegans*. Sci Rep. 2015;4:3912. doi:10.1038/srep03912.

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

63. Lee RYN, Howe KL, Harris TW, Arnaboldi V, Cain S, Chan J, et al. WormBase 2017: Molting into a new stage. Nucleic Acids Res. 2018;46:D869–74. doi:10.1093/nar/gkx998.

64. Stein L, Sternberg PW, Durbin R, Thierry-Mieg J, Spieth J. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. Nucleic Acids Res. 2002;29:82–6. doi:10.1093/nar/29.1.82.

65. Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, et al. Integrative Analysis of the *Caenorhabditis* elegans Genome by the modENCODE Project. Science (80-). 2000;330:1775 LP – 1787.

66. Tepper RG, Ashraf J, Kaletsky R, Kleemann G, Murphy CT, Bussemaker HJ. PQM-1 complements DAF-16 as a key transcriptional regulator of DAF-2-mediated development and longevity. Cell. 2013;154:676–90. doi:10.1016/j.cell.2013.07.006.

67. Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990;215:403–10. doi:https://doi.org/10.1016/S0022-2836(05)80360-2.

68. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: An efficient swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014;31:1929–36. doi:10.1093/molbev/msu136.

69. Snoek BL, Sterken MG, Hartanto M, van Zuilichem AJ, Kammenga JE, de Ridder D, et al. WormQTL2: an interactive platform for systems genetics in Caenorhabditis elegans. Database (Oxford). 2020;2020:1–17.

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

71. Sandoval LE, Jiang H, Wang D. The Dietary Restriction-Like Gene drl-1 , Which Encodes a Putative Serine/Threonine Kinase, Is Essential for Orsay Virus Infection in Caenorhabditis elegans. J Virol. 2019;93:e01400-18.

Figure legends

Figure 1 – Natural variation in the *C. elegans pals*-gene family worldwide – A) The percentage of genetic variants (defined by SNPs) in the *pals*-gene family compared to the overall natural variation for each of the 330 wild isotypes [20]. For axis the amount of SNPs is relative to the reference strain N2. 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 strains in the CeNDR database [20]. Blue dots indicate Tajima's D values for *pals*-genes.

Figure 2 Worldwide haplotype diversity found for *pals-22* and *pals-25* – A) Three distinct haplotypes were found at the *pals-22/pals-25* locus, here represented by an illustration of the read coverage at the locus. The most common haplotype (N2-like) is found in 268 stains and shows low coverage of the second intron of *pals-22*. The second common haplotype is CB4856-like and was found in 31 strains. For these strains coverage indicates strong structural variation in the introns of both genes, as well as larger insertions/deletions. Then, the CX11307-like haplotype was found in 28 strains and consists of very extensive structural variation at the locus, where almost the entire intron structure is not covered by reads. B) A geographical representation of the *pals-22/pals-25* haplotypes found worldwide. C) Zoomed in representation of Figure S10B of the strains collected in Europe. D) Zoomed in representation of Figure S10B of the strains collected on Hawaii.

Figure 3 Gene expression of IPR and *pals*-genes in *C. elegans* N2 and CB4856 under control and OrV-infected conditions – A) Heat-map showing the $\log_2$ intensities of *pals*-genes in N2 mock, N2 infected, CB4856 mock and CB856 infected conditions B) Heatmap showing the expression of IPR genes in $\log_2$ intensities in N2 mock, N2 infected, CB4856 mock and CB856 infected conditions. Underlined genes showed significant (basal) expression differences based on genotype (FDR < 0.05), whereas squares indicated the genes where treatment- or the combination of treatment and genotype had a significant effect (FDR < 0.1) (Table S5). Log_2 ratios are based on the average expression of the gene of interest in the overall dataset. Therefore, the $\log_2$ ratios per experimental group indicate the deviation from the average value. Please note, a subset of the *pals*-genes, namely the *pals*-genes that are also IPR genes (defined in [14]) are depicted twice. This allows for direct comparison to other *pals*-genes that do not become differentially expressed upon infection (like *pals-22* and *pals-25*) and to non-*pals* IPR genes.

Figure 4 Viral susceptibility of *C. elegans* N2, CB4856 and JU1580 upon initial infection – 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) as determined by RT-qPCR for the genotypes N2, CB4856 and JU1580 after exposure to 20, 50 or 100µL OrV/500µL infection solution (student *t*-test; * $p < 0.05$) (4 biological replicates containing roughly 400 nematodes per sample). Boxplots show the division of the data into four quantiles with the middle bar indicating the median.

Supplementary figures

Figure S1 Geographical distribution of natural variation within the *C. elegans pals*-gene family – A) Location of CeNDR strains worldwide. The amount of natural variation (%) within the *pals*-pathway is indicated by the color of the dot. All natural strains 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 S1A of the strains collected in Europe. C) Zoomed in representation of Figure S1A of the strains collected on Hawaii.

Figure S2 Genetic variation per *pals*-gene per *C. elegans* strain – The total number of SNPs within the *pals*-gene family is plotted against the number of known SNPs per *pals*-gene. Each dot represents a strain of the CeNDR database and the color 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 S3 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. The six PC axes that explain at least 5% of the total variation are shown, numbered from 1 (explaining 36.1% of variance) to 6 (explaining 5.5% of variance). Each dot resembles the location of a sample on these axes. The genotype (N2 or CB4856) is indicated by color and the treatment (mock or OrV infection) is indicated by shape.

Figure S4 Gene expression of *pals-6*, *pals-14*, *pals-22*, and *pals-25* in the 30 hours exposure assay determined via RT-qPCR – Box-plots of relative gene expression patterns for *pals-6*, *pals-14*, *pals-22*, and *pals-25* which were determined using RT-qPCR. Measurements were performed on the same 32 samples that were used in the microarray analysis of (mock) infected N2 and CB4856. Each dot represents the expression within a sample. The expression of N2 mock was compared to N2 virus, CB4856 mock and CB4856 virus measurement per gene using a student t-test (*p < 0.05).

Figure S5 Viral loads of *C. elegans* N2, CB4856, and JU1580 after 4 days of exposure. Strains were infected with OrV by adding 20, 50, or 100µL OrV to an NGM plate with three young adults followed by 4 days of incubation (4 biological replicates). Viral susceptibility was determined by measuring viral loads (log_2) by RT-

qPCR (student *t*-test; * $p < 0.05$ compared to N2). Boxplots show the division of the data into four quantiles with the middle bar indicating the median.

Figure S6 Fluorescent *in situ* hybridization (FISH) of ERT54 – Examples of different phenotypes observed after OrV FISH staining in the ERT54 reporter (*pals-5*::GFP) strain. After staining plate-infected ERT54 nematodes, individuals expressing *pals-5* and containing OrV infected cells were observed. These stainings could show in the same individual but were also observed separately. Different fluorescent channels show DAPI nuclear staining (blue), (intestinal) OrV RNA1 signal as stained by FISH (red), expression of the reporter *pals-5*::GFP (green). The exposure time in the red channel was fixed to 1s, for other channels the exposure times were set automatically. All nematodes show red fluorescence due to the *myo-2::*mCherry reporter in the pharynx.

Figure S7 Gene expression of *pals*-genes and IPR genes before and after heat-shock – A) Heat-map showing the $\log_2$ intensities of *pals*-genes in N2 control, N2 heat-shock, N2 recovery, CB4856 control, CB4856 heat-shock and CB4856 recovery conditions B) Heatmap showing the expression of IPR genes in $\log_2$ intensities in N2 control, N2 heat-shock, N2 recovery, CB4856 control, CB4856 heat-shock and CB4856 recovery conditions. This dataset is described by [27, 37]. Underlined genes showed significant (basal) expression differences based on genotype (FDR < 0.05), whereas squares indicated the genes where treatment had a significant effect (FDR < 0.1). Log_2 ratios are based on the average expression of the gene of interest in the overall dataset. Therefore, the $\log_2$ ratios per experimental group indicate the deviation from the average value. A subset of the *pals*-genes, namely the *pals*-genes that are also IPR genes (defined in [14]) is depicted twice to facilitate direct comparison to non-IPR *pals*-genes and non-*pals* IPR genes.

Figure S8 Differentially expressed genes upon OrV infection – A) Volcano plot showing the effect of treatment (OrV infection) on global gene expression patterns. Microarray spots (blue) above the FDR threshold (dotted line) were considered differentially expressed. B) Volcano plot showing the effect of the interaction between treatment (OrV infection) and genotype (N2 or CB4856) on global gene expression patterns. Microarray spots (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 (OrV infection) and/or genotype x treatment.

Figure S9 Analysis of *C. elegans* expression patterns for OrV-response genes over time and long-term pathogenic exposure – 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). The lines represent the fit through the data (by loess) and the grey area around a curve represents the 95% confidence interval, genotypes are indicated by colors and mock infection is indicated by dashed lines. 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. B) Box-plots of relative gene expression patterns for *pals-6*, *pals-14*, *pals-22*, and *pals-25* measured with RT-qPCR after exposure to OrV. Measurements were performed on the same 4 N2 and CB4856 samples that were obtained by continuous exposure to either mock-conditions or 50µL OrV. Each dot represents the expression within a sample. The expression of N2 mock was compared to N2 virus and the expression of CB4856 mock was compared to the CB4856 virus measurement per gene using a student t-test (*$p < 0.05$).

Supplementary tables

Table S1 Population genetic properties per *pals*-gene – Conservation (genes are considered conserved when the vast majority of CeNDR strains contains < 25 polymorphisms), haplotype number and Tajima's D value per *pals*-gene. Genes that are upregulated upon infection in [17] are indicated.

Table S2 Haplotypes of *pals-22* and *pals-25* per CeNDR strain – Haplotypes were manually inspected based for sequence coverage of *pals-22* and *pals-25* in the CeNDR database. The haplotype groups were noted for both genes per strain.

Table S3 Hybridization data from Volkers *et al.*, 2013 - N2, WN2002, and CB4856 DNA (**A**) and RNA hybridization (**B**) to *pals-22* and *pals-25*. The $\log_2$ intensities for the probes are reported. Note that the samples were different for the two experiments. The full dataset was previously published in [23] and can be found online (ArrayExpress E-MTAB-8126).

Table S4 Genes involved during OrV infection – Overview of *C. elegans* genes involved in OrV infection described in literature. Data is obtained from [7, 9, 47, 52, 70, 71, 10–14, 17, 19, 43].

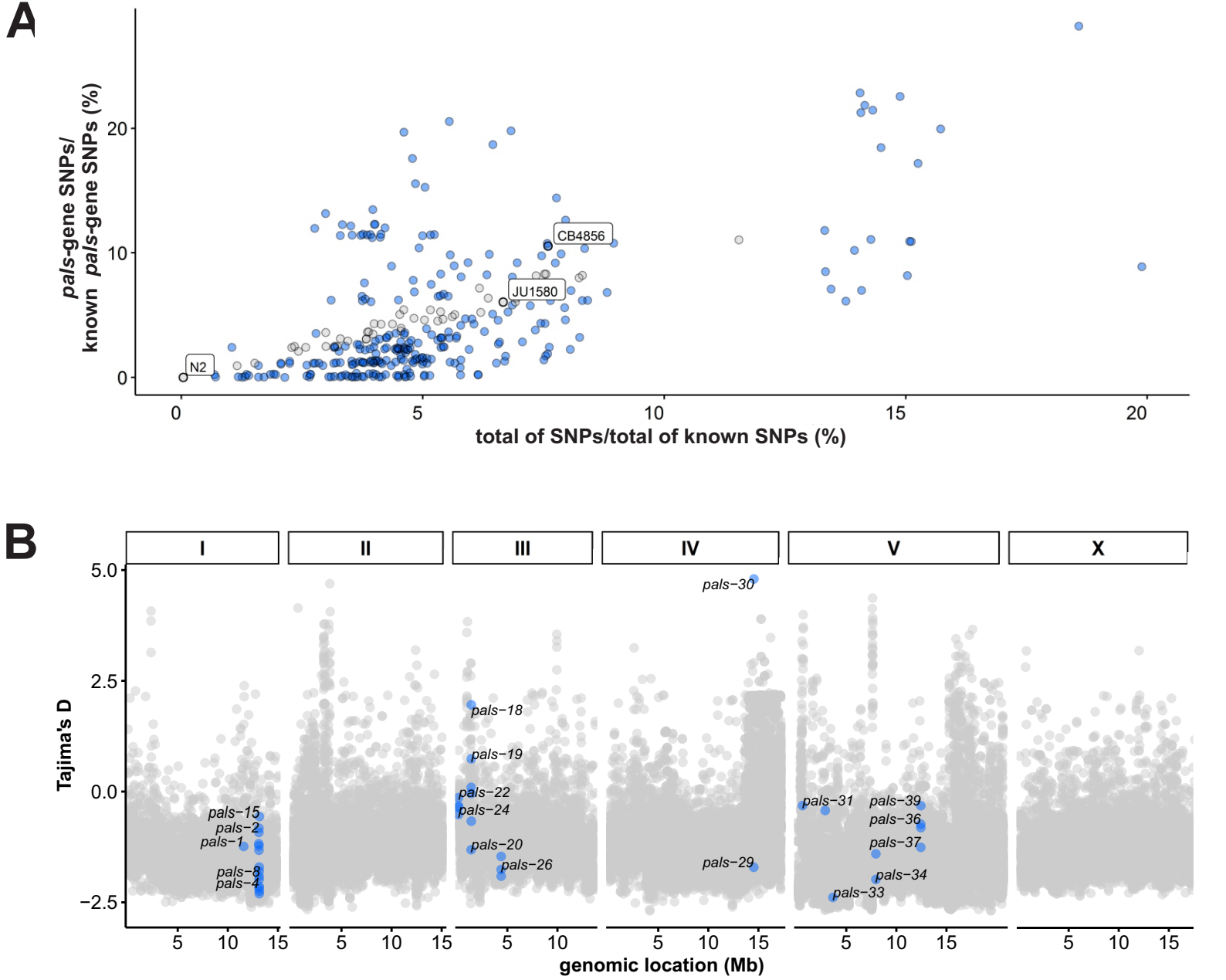
Table S5 Output linear models – A) The linear correlation between genotype, treatment, the interaction between genotype and treatment and gene expression for the N2 and CB4856 30 hours post (mock) infection. B) The linear correlation between viral load and gene expression within infected samples per genotype for the N2 and CB4856 30 hours post (mock) infection. C) The linear correlation between development, genotype, treatment and gene expression for N2, JU1580 and CB4856 infected samples that were isolated 1.5-30 hours post infection.

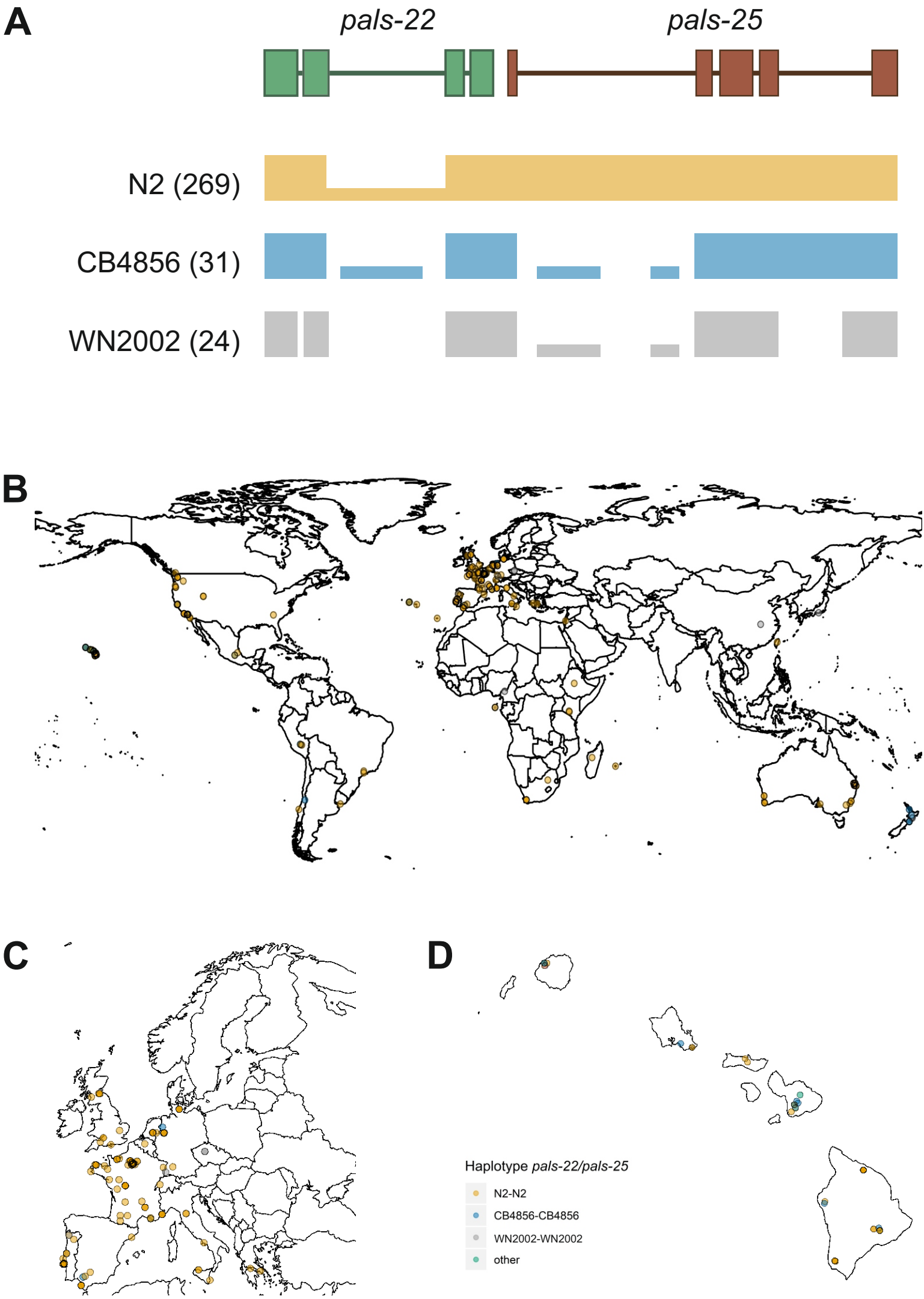
Table S6 Microarray probe binding to the CB4856 genome – A) BLAST alignment of *C. elegans* CB4856 sequence (PJRNA275000) to microarray probes (*C. elegans* (V2) Gene Expression Microarray 4X44K slides) that detect the *pals*-genes. Probes are considered to align correctly when the there is more than 95% overlap on the correct chromosome. B) N2 and CB4856 DNA hybridization to the microarray probes genes that are differentially expressed based on genotype, treatment or their interaction. A ratio of 0 indicates there is

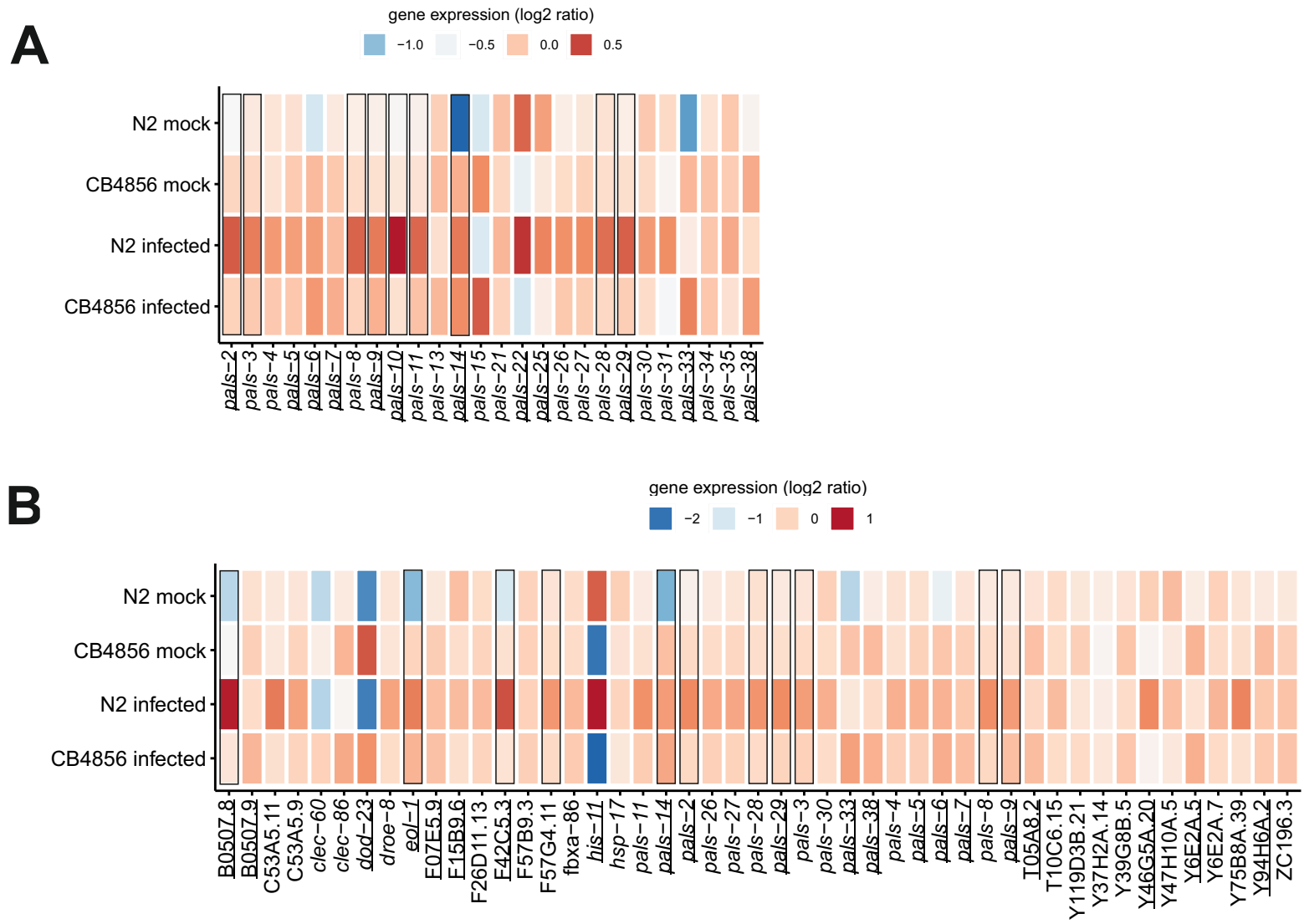
absolutely no difference between N2 and CB4856 hybridization. The full dataset was previously published in [23] and can be found online (ArrayExpress E-MTAB-8126). C) BLAST alignment of *C. elegans* CB4856 sequence (PJRNA275000) to microarray probes (*C. elegans* (V2) Gene Expression Microarray 4X44K slides) detect the IPR genes. Probes are considered to align correctly when the there is more than 95% overlap on the correct chromosome. D) BLAST alignment of *C. elegans* CB4856 sequence (PJRNA275000) to microarray probes (*C. elegans* (V2) Gene Expression Microarray 4X44K slides) that detect the differentially expressed genes from the linear model investigating the terms genotype, treatment and genotype x treatment. Probes are considered to align correctly when the there is more than 95% overlap on the correct chromosome. E) N2 and CB4856 DNA hybridization to the microarray probes genes that are differentially expressed based on genotype, treatment or their interaction. A ratio of 0 indicates there is absolutely no difference between N2 and CB4856 hybridization. The full dataset was previously published in [23] and can be found online (ArrayExpress E-MTAB-8126).

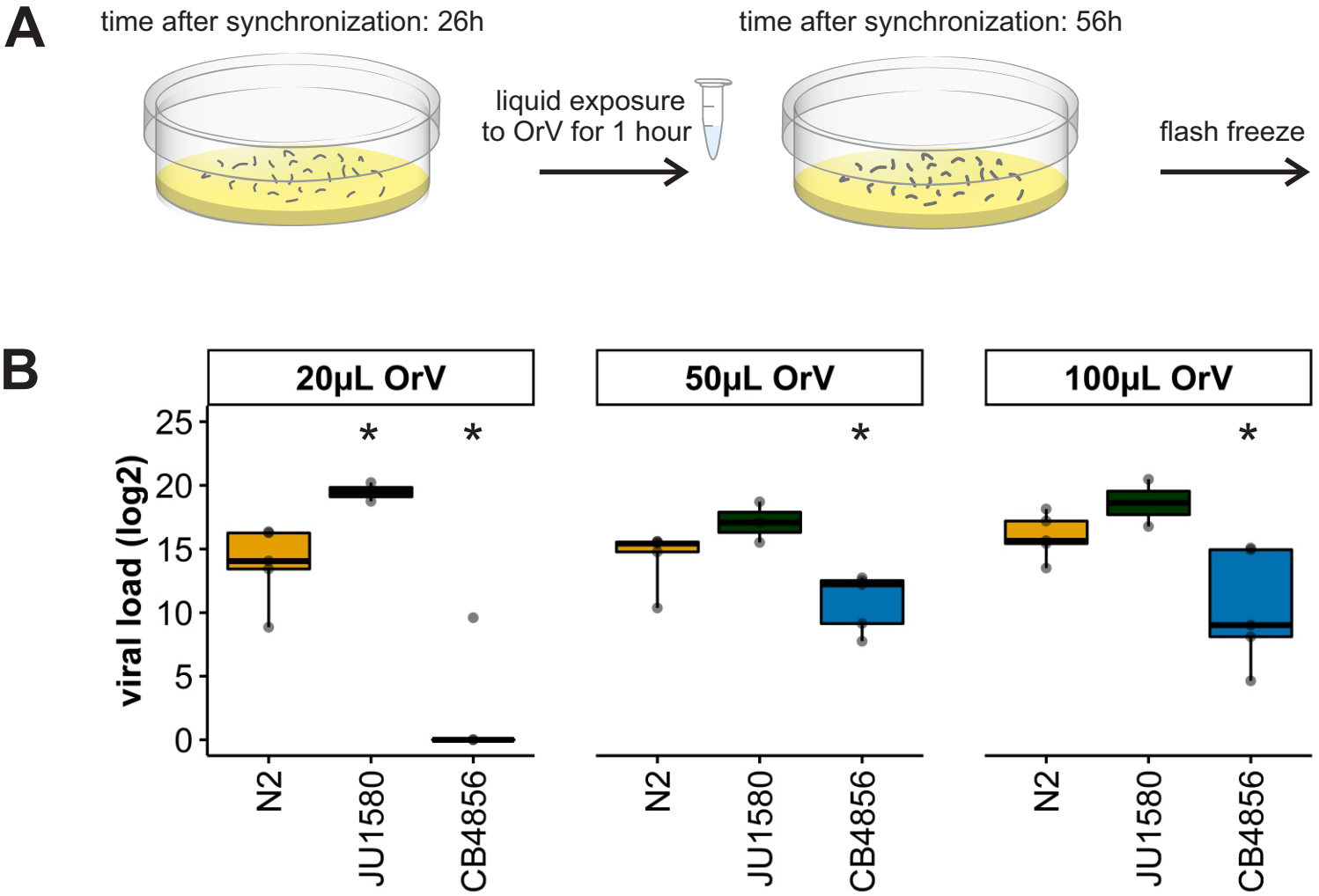
Table S7 eQTL found for the *pals*-genes – Summary of *cis*- and *trans*-eQTL found for the *pals*-genes in previous eQTL studies.

Table S8 Fluorescent signals observed in infected nematodes – Summary of the phenotypes that were observed after OrV FISH staining in the strains N2, CB4856, ERT54, ERT71, and JU1580. The total number of inspected nematodes is shown and the number of nematodes showing either a green fluorescent signal (IPR reporter gene + GFP), red fluorescent signal (OrV RNA1.1) or both signals.


A
pals-gene SNPs/
known pals-gene SNPs (%)
total of SNPs/total of known SNPs (%)
N2
CB4856
JU1580
B
Tajima's D
genomic location (Mb)
I
II
III
IV
V
X
pals−30
pals−18
pals−19
pals−22
pals−24
pals−20
pals−26
pals−15
pals−2
pals−1
pals−8
pals−4
pals−29
pals−31
pals−39
pals−36
pals−37
pals−34
pals−33


A
pals-22
pals-25
N2 (269)
CB4856 (31)
WN2002 (24)
B
C
D
Haplotype pals-22/pals-25
N2-N2
CB4856-CB4856
WN2002-WN2002
other


A
gene expression (log2 ratio)
−1.0
−0.5
0.0
0.5
N2 mock
CB4856 mock
N2 infected
CB4856 infected
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−15
pals−21
pals−22
pals−25
pals−26
pals−27
pals−28
pals−29
pals−30
pals−31
pals−33
pals−34
pals−35
pals−38
B
gene expression (log2 ratio)
−2
−1
0
1
N2 mock
CB4856 mock
N2 infected
CB4856 infected
B0507.8
B0507.9
C53A5.11
C53A5.9
clec−60
clec−86
dod−23
droe−8
eol−1
E07E5.9
E15B9.6
F26D11.13
F42C5.3
F57B9.3
F57G4.11
fbxa−86
his−11
hsp−17
pals−11
pals−14
pals−2
pals−26
pals−27
pals−28
pals−29
pals−3
pals−30
pals−33
pals−38
pals−4
pals−5
pals−6
pals−7
pals−8
pals−9
T05A8.2
T10C6.15
Y119D3B.21
Y37H2A.14
Y39G8B.5
Y46G5A.20
Y47H10A.5
Y6E2A.5
Y6E2A.7
Y75B8A.39
Y94H6A.2
ZC196.3


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