Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*

Delia O'Rourke, Dilair Baban, Maria Demidova, et al.

Genome Res. 2006 16: 1005-1016
Access the most recent version at doi:10.1101/gr.50823006

References This article cites 49 articles, 19 of which can be accessed free at:
http://genome.cshlp.org/content/16/8/1005.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Copyright © 2006, Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*

Delia O'Rourke,[1] Dilair Baban,[2] Maria Demidova,[1] Richard Mott,[3] and Jonathan Hodgkin[1,4]

[1]Genetics Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom; [2]MRC Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, Oxford OX1 3QU, United Kingdom; [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom

The interaction between the nematode *Caenorhabditis elegans* and a Gram-positive bacterial pathogen, *Microbacterium nematophilum*, provides a model for an innate immune response in nematodes. This pathogen adheres to the rectal and post-anal cuticle of the worm, causing slowed growth, constipation, and a defensive swelling response of rectal hypodermal cells. To explore the genomic responses that the worm activates after pathogenic attack we used microarray analysis of transcriptional changes induced after 6-h infection, comparing virulent with avirulent infection. We defined 89 genes with statistically significant expression changes of at least twofold, of which 68 were up-regulated and 21 were down-regulated. Among the former, those encoding C-type lectin domains were the most abundant class. Many of the 89 genes exhibit genomic clustering, and we identified one large cluster of 62 genes, of which most were induced in response to infection. We tested 41 of the induced genes for involvement in immunity using mutants or RNAi, finding that six of these are required for the swelling response and five are required more generally for defense. Our results indicate that C-type lectins and other putative pathogen-recognition molecules are important for innate immune defense in *C. elegans.* We also found significant induction of genes encoding lysozymes, proteases, and defense-related proteins, as well as various domains of unknown function. The genes induced during infection by *M. nematophilum* appear largely distinct from genes induced by other pathogens, suggesting that *C. elegans* mounts pathogen-specific responses to infection.

[Supplemental material is available online at www.genome.org. The gene expression microarray data from this study have been submitted to ArrayExpress, under accession no. E-MEXP-696.]

The complete genome sequences of all eukaryotes so far sequenced reveal the presence of many predicted genes that have no known function as yet. In the case of organisms such as *Drosophila, Caenorhabitis elegans*, or mouse, in which efficient genetic analysis is possible, tests of function by means of gene deletion or RNAi knockdown do not greatly improve the situation, because loss of gene function often leads to no obvious mutant phenotype. In some cases this may be due to redundancy, but another explanation is that tests are usually carried out under laboratory conditions, so that many of the functions that are essential for survival in the natural environment of each organism are not examined. Genes required for defense against pathogens are likely to constitute a significant part of the category of "mystery genes."

The nematode *C. elegans* has become established as a very useful model organism for the study of many different processes including development, neurobiology, and apoptosis. More recently, it has been used to study innate immunity and host-pathogen interactions (Millet and Ewbank 2004; Gravato-Nobre and Hodgkin 2005; Mylonakis and Aballay 2005). Under normal laboratory conditions, it is cultured by feeding on *Escherichia coli*, which is not pathogenic except to aging worms (Garsin et al. 2003). In contrast, its natural environment is the soil, which is rich in a great diversity of microbes, many of them potentially pathogenic. It follows that *C. elegans* is likely to possess effective microbial perception and antimicrobial defense mechanisms. Indeed, the genome of *C. elegans* encodes numerous proteins with homology with proteins that perform these perception and defense roles in other organisms (Nicholas and Hodgkin 2004b; Schulenburg et al. 2004). These homologs may fulfill similar roles in *C. elegans,* but as yet there has been limited evidence of their involvement.

A combination of forward genetic screens and reverse-genetic, post-genomic approaches has identified some of the worm genes required for response to pathogenic attack. These genes belong to six signaling cascades, which may be activated in response to infection by a variety of pathogens including Gram-positive and Gram-negative bacteria, fungi, and bacterial toxins. The six pathways are ERK MAP kinase (Nicholas and Hodgkin 2004a), p38 MAP kinase (Kim et al. 2002), TGF-β (Mallo et al. 2002), programmed cell death (Aballay and Ausubel 2001), DAF-2/DAF-16 insulin-like receptor signaling (Garsin et al. 2003), and JNK-like MAP kinase (Huffman et al. 2004).

However, we know little about the *C. elegans* proteins that may perceive pathogens upstream of these pathways or about the downstream proteins that mediate antimicrobial defense. To investigate this question, we have used microarray expression data to analyze whole-genome changes in expression caused by the

[4]Corresponding author.
E-mail jonathan.hodgkin@bioch.ox.ac.uk; fax 44-1865-275318.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.50823006.

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

interaction between *C. elegans* and the Gram-positive bacterium *Microbacteroim nematophilum* (Hodgkin et al. 2000; Gravato-Nobre and Hodgkin 2005). This pathogen adheres to the rectum and post-anal cuticle, causing swelling of the underlying hypodermal cells (the Dar, or Deformed Anal Region response), slowed growth and constipation. We have previously shown that the worm responds to this pathogen through activation of a protective ERK map kinase pathway, which activates swelling and prevents severe constipation (Nicholas and Hodgkin 2004a). Mutants with defects in the MAP kinase *mpk-1*, or other kinases in this pathway, exhibit a Bus (Bacterially Un-Swollen) phenotype and are much more severely affected by the pathogen, indicating that the swelling is protective. Our aim was to identify genes downstream of this and other signal-transduction pathways that may be induced upon infection with *M. nematophilum*. We infected synchronized larval worms in liquid culture to ensure that the population was evenly exposed to the pathogen and that developmental differences between worms were minimized.

We then tested the contribution that up-regulated genes made to the worm's response to *M. nematophilum* infection by using available knockout mutants or RNAi to knock-down expression of selected candidate genes. For most of the 41 tested genes, RNAi had no effect, but for six genes, RNAi resulted in a defective swelling (Dar) response, and for five others, RNAi resulted in much "sicker" (more unhealthy) phenotypes in the presence of pathogen, with exaggerated tail swelling, very severe constipation, growth retardation, or arrest.

We found that C-type lectins and at least one other putative pathogen-recognition molecule were important for innate immune defense in *C. elegans*. We also found induction of a variety of antimicrobial effectors including lysozymes, proteases, and defense-related proteins, and of several proteins of unknown function. Finally, we demonstrate the striking induction of a cluster of 62 genes on chromosome IV, as well as significant clustering of other responding genes. The chromosome IV cluster contains genes encoding numerous putative pathogen-recognition proteins including C-type lectins, galectins, and GPCR's (G-protein coupled receptors).

The majority of *C. elegans* genes induced after infection with *M. nematophilum* appear to be specific to this host–pathogen interaction, because their induction has not been observed in comparable studies using other bacterial pathogens (Mallo et al. 2002; Huffman et al. 2004). This provides further evidence that *C. elegans* can mount pathogen-specific responses. However, there is noticeable overlap in the kinds of protein domains encoded by genes induced by *M. nematophilum* as compared with genes induced by other pathogens, suggesting that although each infection may induce a specific response, the different sets of induced proteins may still share some functionalities.

Results

Genomic transcriptional analysis

C. elegans gene expression was examined after a 6-h infection of synchronized L2/L3 larval stage worms with *M. nematophilum* (Fig. 1; virulent CBX102 or avirulent UV336). RNA was extracted from infected worms and then hybridized to Affymetrix *C. elegans* GeneChips, and gene transcript levels were determined using Gene Chip Operating Software (MAS5.0 and GCOS1.0). Infections with CBX102 were compared with infections with UV336 in biological triplicate, i.e., the entire experiment was performed three times. The 68 genes that were induced on average more than twofold in at least two of three of the replicates are listed in Table 1. This set of 68 genes is strikingly enriched with certain protein domains, gene families, and proteins that have putative roles in defense.

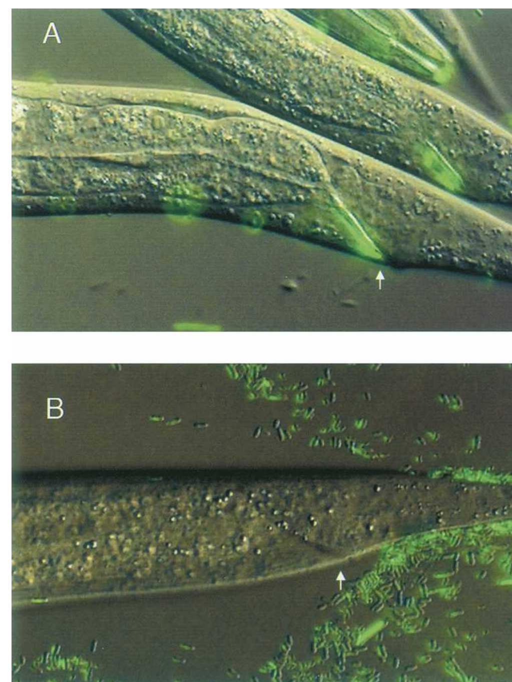

Figure 1. Infection of N2 worms in liquid culture. Synchronized L1 worms were grown in liquid culture with *E. coli* HB101 for 24 h at 25°C before addition of *M. nematophilum*; either virulent CBX102 (*A*) or avirulent UV336 (*B*) was added. Infection was allowed to proceed for 6 h before worms were harvested for RNA extraction and staining adherent bacteria with the nucleic acid dye SYTO 13 (Molecular Probes). Infection (*A*) is apparent from the bacterial staining in the rectum and the associated tail swelling. Arrow indicates anus.

Induced C-Type lectins

Ten of the 68 up-regulated genes encode proteins containing C-type lectin domains (CTLD). This is 6% of the 172 CTLD-containing genes (as predicted by Ensembl: http://www.ensembl.org/index.html) in the genome and 18-fold more than the 0.6 CTLD genes expected to occur by chance. C-type lectins in *C. elegans* have been proposed to act as pathogen-recognition molecules, similar to their role in pathogen-associated molecular pattern (PAMP) recognition in *Drosophila* (Franc and White 2000). However, it is also possible that they function as effectors of the antimicrobial response by binding to carbohydrates on the surface of pathogens, similar to cockroach opsonins (Wilson et al. 1999). Here, their role might be to neutralize and mask virulence factors on the pathogen surface and perhaps interfere with bacterial quorum sensing (Nicholas and Hodgkin 2004b). As *C. elegans* has no cellular immune system, C-type lectins could function as secreted proteins in the intestine and on environmentally exposed surfaces. Two CTLD-containing genes were induced by infection of *C. elegans* with *Serratia marcescens*, (Mallo et al. 2002). They were different from those found in this study, supporting the theory that specific pathogen recognition molecules are induced by particular infections.

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 1. Genes activated over twofold following infection of *C. elegans* with *M. nematophilum*

WormBase Cosmid ID	Gene symbol	Fold induction	Chr	Description
F09E10.11	*tts-1*	7.5	X	Transcribed telomerase-like sequence; noncoding RNA; up-regulated 20-fold in dauer
ZK666.6	*clec-60*	6.6	II	C-type lectin domain; CUB domain; von Willebrand factor type A domain
F53A9.8		5.1	X	Similarity to *Plasmodium* histidine-rich glycoprotein precursor
F21C10.10		4.7	V	Similarity to *Euglena gracilis* L-3-hydroxyacyl-CoA dehydrogenase subunit precursor
Y54G2A.11a		4.4	IV	Similarity to Myb DNA-binding domain
Y46C8AL.3	*clec-70*	4.4	IV	C-type lectin domain short and long forms
F35E12.5		3.8	V	Domain of unknown function DUF141
T07D10.4	*clec-15*	3.7	I	Two C-type lectin domains, CUB domain; coding sequence identical to *clec-13*
H02F09.3		3.7	X	Similarity to *Drosophila* salivary gland secreted protein Sgs1-PA
F01G10.3	*ech-9*	3.7	IV	ECH (hydroxyacyl-CoA dehydrogenase/enoyl-CoA hydratase) family
Y46C8AL.2		3.5	IV	Similarity to *Plasmodium* histidine-rich glycoprotein precursor
T22F3.11		3.4	V	Similarity to Major Facilitator Superfamily and Myb, DNA binding
F49E11.10		3.4	IV	Similarity to defense-related protein SCP domain and testis-specific protein TPX-1
F53B2.5		3.2	IV	Contains armadillo/β-catenin-like repeats
C45G7.2		3.1	IV	Similarity to destabilase/invertebrate lysozyme
C50F4.9		3.1	V	Similarity to PTS system, glucose-specific, IIBC component
C54D1.2	*clec-86*	3	X	C-type lectin domain short and long forms
M28.8		2.9	II	Glutamine amidotransferase, class-II
F28D1.5	*thn-2*	2.9	IV	THN protein; similarity to thaumatin, pathogenesis related
F09B9.1		2.8	X	Acyltransferase family
F54F3.3		2.8	V	Triglyceride lipase-cholesterol esterase, contains α/β hydrolase fold
C49F5.7		2.8	X	Similarity to *Buchnera aphidicola* acyl carrier protein
K01A2.2a	*far-7*	2.8	II	FAR (fatty acid/retinoid binding protein) family
B0024.4		2.7	V	Domain of unknown function DUF274
F22B7.9		2.6	II	Predicted methyltransferase
E02D9.1a		2.6	I	Similarity to protein kinase domains
Y39B6A.1		2.6	V	Similarity to *Plasmodium* histidine-rich glycoprotein precursor
R07C12.1		2.6	IV	7-transmembrane receptor
C17H12.8		2.6	IV	Domain of unknown function DUF141, CUB domain
K12G11.3	*sodh-1*	2.6	V	Sorbitol/alcohol dehydrogenase family class V
T24E12.5		2.6	II	Similarity to *Plasmodium* merozoite surface protein
F28F8.2	*acs-2*	2.5	V	Fatty-acid-CoA synthetase family, AMP-binding enzyme
C17H1.7		2.5	I	Similarity to *Plasmodium* reticulocyte binding-like protein 2b
F56D6.2	*clec-67*	2.5	IV	C-type lectin domain short and long forms
F11A5.10	*glc-1*	2.5	V	Glutamate-gated chloride channel
ZK666.7	*clec-61*	2.4	II	C-type lectin domain; CUB domain; von Willebrand factor type A domain
F14D12.5	*sulp-2*	2.4	X	Sulfate transporter
E03H4.10	*clec-17*	2.4	I	Two C-type lectin domains, CUB domain
W07B8.4		2.4	V	Similarity to cysteine protease domains
Y54G2A.8	*clec-82*	2.4	IV	C-type lectin domain
C49C8.5		2.4	IV	Similarity to ficolin and related extracellular proteins
Y53F4B.14	*asc-1*[a]	2.3	II	ASC (Activating Signal Cointegrator) family
C04G6.1	*mpk-2*	2.3	II	MPK-2 MAP kinase protein
K08D10.10		2.3	IV	Similarity to *Saccharomyces cerevisiae* oligopeptide transporter
K05F1.10		2.3	II	Similarity to trypsin inhibitor-like cysteine-rich domain
C15A11.7		2.3	I	Similarity to amine oxidase
Y5F2A.2		2.3	IV	Transthyretin-like family
C07E3.3		2.3	II	Similarity to *Cryptococcus neoformans* hypothetical protein
F52F10.3		2.3	V	Acyltransferase family
F28F8.2		2.3	V	Similarity to long-chain-fatty-acid-CoA ligase
ZC443.3		2.3	V	
ZC395.5		2.2	III	Similarity to *Saccharomyces cerevisiae* DNA polymerase V
F09F7.6		2.2	III	Similarity to moesin
F49F1.5		2.2	IV	Similarity to ShTK domain
F35C5.5	*clec-62*	2.2	II	C-type lectin domain; CUB domain; von Willebrand factor type A domain
C31A11.5		2.2	V	Acyltransferase family
F54D7.4	*zig-7*	2.2	I	ZIG (two immunoglobulin domains) family
ZK593.6	*lgg-2*	2.2	IV	LGG (microtubule-associated protein) family
K05B2.3	*ifa-4*	2.2	X	IFA (intermediate filament protein) filament family
F56D6.1	*clec-68*	2.2	IV	Lectin C-type domain short and long forms
Y41D4B.16		2.1	IV	Domain of unknown function DUF274
C07A9.8		2.1	III	Putative membrane protein
ZK896.5		2.1	IV	Domain of unknown function DUF141; CUB domain
M60.4b		2.1	X	Similarity to *Drosophila melanogaster* gene CG15211-PA
C02A12.4	*lys-7*	2.1	V	LYS (lysozyme) family
F26F12.3[a]		2.1	V	Similarity to E3 ubiquitin ligase
F45D11.15		2.1	II	Domain of unknown function DUF684
F43C11.3		2	II	

[a]Probe in intron.
A total of 68 genes were activated in *C. elegans* following infection with virulent CBX102 compared with infection with avirulent UV336. These genes were induced over twofold on average and in at least two of the triplicate replicates. The table lists the WormBase Cosmid gene ID, gene symbol, fold induction after 6-h infection, chromosomal location, and description including annotations and homology from WormBase release 144.

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 2. Possible pathogen recognition proteins induced during *M. nematophilum* infection

Gene name	Signal sequence	Domain structure and specificity	Expression pattern	Regulation	Fold Induction
clec-60	Yes	von Willebrand factor A followed by C-type lectin domain and CUB domain	Intestine: int 8 and 9		6.6
clec-86	Yes	C-type lectin domain	unknown		3.0
clec-67	Yes	C-type lectin domain followed by ~300 aa with no similarity to known protein domains	Intestine	TGF-β-regulated (Mochii et al. 1999)	2.5
clec-61	Yes	von Willebrand factor A followed by C-type lectin domain and CUB domain	unknown		2.4
clec-17	Yes	Two C-type lectin domains followed by a CUB domain	unknown		2.4
clec-62	Yes	von Willebrand factor A followed by C-type lectin domain and CUB domain	unknown		2.2
clec-68	Yes	C-type lectin followed by ~300 aa with no similarity to known protein domains.	unknown		2.2
clec-13 clec-15	Yes	Two C-type lectin domains followed by a CUB domain	unknown	DAF-16-regulated (Murphy et al. 2003)	3.7
clec-70	Yes	C-type lectin followed by ~300 aa with no similarity to known protein domains.	unknown		4.4
clec-82	Yes	C-type lectin followed by ~300 aa with no similarity to known protein domains.	unknown		2.4
C49C8.5	Yes	Extracellular ficolin. fibrinogen β and γ chains, C-terminal globular domain	unknown		2.4

Like the majority of proteins containing CTLD in the *C. elegans* genome, all of the lectins that were induced following infection with *M. nematophilum* have a hydrophobic signal peptide sequence at the N terminus and no predicted membrane anchorage sequence, suggesting that they are soluble and secreted (Table 2). Consistent with a role in innate immunity, *clec-67 (F56D6.2)* is regulated in part by the TGF-β pathway (Mochii et al. 1999) and *clec-15 (T07D10.4)* in part by DAF-16 (Murphy et al. 2003). Moreover, *clec-67 (F56D6.2)* is expressed in the intestine (Mochii et al. 1999), where pathogens are likely to be recognized.

We investigated the contribution that induced CTLD proteins made to *C. elegans'* response to *M. nematophilum* infection by using RNAi to knock down expression of some of these genes. RNAi of three C-type lectins, *clec-60 (ZK666.6)*, *clec-17 (E03H4.10)*, and *clec-86 (C54D1.2)* resulted in an altered response to infection. The RNAi-treated worms showed a hyper-Dar response and suffered from severe constipation and are included in the group of genes that when knocked down, cause the worms to be "sick" on the pathogen (Table 3). These lectins are not up-regulated in any of the other published infections so far studied (Mallo et al. 2002) and little has been reported on their regulation or expression patterns.

We therefore constructed a promoter GFP fusion for *clec-60 (ZK666.6)*, the most highly induced lectin at 6 h, and found that expression was evident throughout the intestine in larvae and predominantly in the posterior intestinal cells int8 and int9 in adults (Fig. 2A–C). Quantitative measurement of *clec-60p::GFP* induction following infection was difficult, as the expression levels within a population of worms were variable and appeared to be influenced by starvation or stress as well as by infection.

In order to independently confirm the transcriptional induction of *clec-60 (ZK666.6)* seen in the microarray analysis, we used semiquantitative RT–PCR. This confirmed the induction of *clec-60 (ZK666.6)* and of *F21C10.10*, relative to the control *ama-1* whose transcript levels were unchanged by infection (Fig. 2D).

Induced putative pathogen recognition proteins

Other genes encoding putative pathogen-recognition proteins are found in the up-regulated spatial clusters of the genes described below. Within these clusters we find genes for galectins, (galactose-binding proteins), which were also seen to be induced by infection with *Serratia marcescens* (Mallo et al. 2002) and for proteins from the large family of chemoreceptors, some of which have been suggested to act as pathogen-recognition molecules (Millet and Ewbank 2004). These include the serpentine class H (*srh*) genes *srh-220, srh-222, srh-223, str-122* (olfactory), and four other 7TM genes, i.e., *R07C12.1, R07C12.3, R07C12.4*, and *K08D10.10*. Two other up-regulated genes, *F47C12.1* and *W02C12.1*, encode large EGF-receptor-like multidomain proteins and could also be pathogen-recognition molecules. RNAi of *F47C12.1* resulted in Bus (unswollen) worms in the presence of the pathogen (Table 4).

Table 3. Genes that affect *C. elegans'* defense against *M. nematophilum*

Cosmid ID	Fold induction	Functional domains/ homology/ proposed activity	Experimental evidence
ZK666.6 (*clec-60*)	6.6	von Willebrand factor A followed by C-type lectin domain and CUB domain	RNAi feeding, N2 and *rrf-3*
E03H4.10 (*clec-17*)	2.4	Two C-type lectin domains followed by a CUB domain	RNAi feeding, N2 and *rrf-3*
C54D1.2 (*clec-86*)	3.0	C-type lectin domain	RNAi feeding, *rrf-3*
C45G7.3	4.3 (1 expt.)	Similar to invertebrate-type lysozymes	RNAi feeding, N2 and *rrf-3*
C02A12.4 (*lys-7*)	2.1	Cleavage of bacterial cell-wall peptidoglycan	RNAi feeding, N2 and *rrf-3;* mutant alleles

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

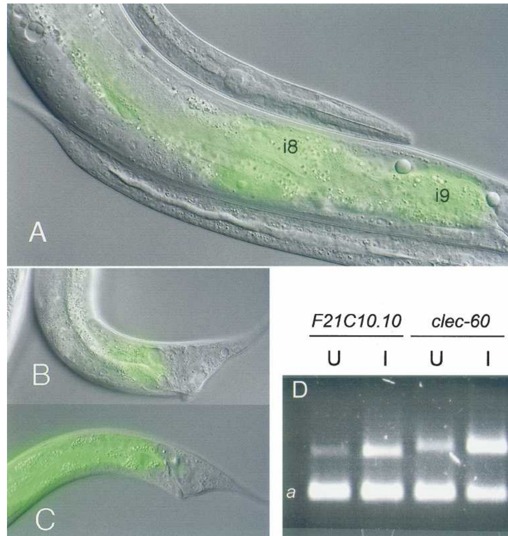

Figure 2. Expression of the C-type lectin gene *clec-60*. A promoter fusion, *clec-60p::GFP*, was expressed in adults in the posterior intestine cells, int-8 (i8) and int-9 (i9) on both OP50 (*A*) and MBL plates (*B*,*C*). The variation in the intensity and extent of the GFP signal within infected worms is evident in *B* and *C*. (*D*) RT–PCR experiment showing *F21C10.10* and *clec-60* (*top* bands) expression compared with *ama-1* (control, *bottom* band a) in uninfected (U) and infected (I) worms.

Induced putative antimicrobial effectors

Once *C. elegans* has perceived the pathogen, it must then activate further defense mechanisms, including the synthesis of antimicrobial proteins. There are many putative antimicrobial effectors encoded in the *C. elegans* genome (Nicholas and Hodgkin 2004b; Schulenburg et al. 2004), and a number have been shown to be induced by infection or to have antimicrobial activity in in vitro assays (Mallo et al. 2002; Couillault et al. 2004). Table 5 lists the putative antimicrobial effectors induced following a 6-h infection with *M. nematophilum* and they are discussed in the following sections.

Lysozymes

Lysozymes play an important role in both vertebrate and invertebrate immunity. In *Entamoeba histolytica* they act in concert with amoebapores to degrade bacteria (Leippe 1999). The *C. elegans* lysozyme gene *lys-7* is induced 2.1-fold after 6-h infection with *M. nematophilum*. This lysozyme is expressed in the intestine (Mallo et al. 2002), which is where a defense against *M. nematophilum* might be mounted, and is regulated by the DAF-16 pathway, (Murphy et al. 2003), which is important for defense against other pathogens. Two other lysozyme genes, *lys-3* and *lys-8*, were also upregulated, although to a lesser degree (1.7-fold for *lys-3* and 1.5-fold for *lys-8*).

We investigated the role of *lys-7* in defense against *M. nematophilum* by growing strains for both available *lys-7* mutant alleles on MBL plates. Infected worms were very slow growing, sick, and constipated, and a large proportion of the population arrested at the L3 larval stage (Fig. 3; Table 3). Both alleles are large deletions and would not be expected to produce a functional protein. We also observed a low-penetrance phenotype of abnormalities in the gonads, including both vacuoles and apparent proliferation of cells. We are currently investigating these unexpected abnormalities further. No severe defects were observed on OP50 plates, although a number of worms exhibited a milder version of the gonad defects observed on MBL plates (data not shown).

A second group of putative lysozymes were also induced in one (*C45G7.3*) or two (*C45G7.2*) of the biological replicates. These proteins were originally defined as destabilases based on homology with proteins from the medicinal leech (*Hirudo medicinalis*) (Zavalova et al. 1996), which had a novel isopeptidase activity. However, more recently, this family of proteins has been shown to be part of a novel family of lysozymes (Zavalova et al. 2000), related to invertebrate-type lysozymes (Bachali et al. 2002). We found that *C45G7.3* is important in the defense against *M. nematophilum*, since knockdown of expression using RNAi made the worms more unhealthy on the pathogen (Table 3). However, RNAi of *C45G7.2* had no obvious effect. These results confirm the importance of lysozymes in the response of *C. elegans* to infection and also suggest a role for invertebrate-type lysozymes in worm defense.

Other antimicrobial effectors

We found specific induction of the lipase gene *F54F3.3* in all three biological replicates. Lipases act directly against microorganisms and are induced in both *Drosophila* (De Gregorio et al. 2001) and *C. elegans* following infection (Mallo et al. 2002), suggesting that these enzymes play an important role in antimicrobial defense. However, neither RNAi nor knockout (*tm1954*) of this gene had any pronounced effect on the response to infection.

There are seven putative antimicrobial genes, *thn-1* to *thn-7*, related to the plant defense thaumatin/PR5 (Kitajima and Sato 1999) and 18 PR-1 homologs in the *C. elegans* genome (as inferred from a BLAST search using *Arabidopsis thaliana* pathogenesis-related protein 1 [PR-1At2g14610]). Both of these families are up-regulated following infection in plants and have been shown to have antifungal activity, although the mechanism of this activity is unknown. Two members of these families are up-regulated in the *M. nematophilum* response, i.e., both *thn-2* and a PR-1 related gene, *F49E11.10*, are up-regulated more than two-fold after 6 h of infection. RNAi of *thn-2* gene confirmed its possible role in defense, but gave puzzling results; when N2 worms

Table 4. Genes that affect the morphogenetic swelling response

Cosmid ID	Fold induction	Protein size and functional domains	Experimental evidence
F47C12.1	1.4	1827 aa; Fibrillin containing Ca2+-binding EGF-like domains; EGF-like receptor	RNAi feeding, N2 and *rrf-3*;RNAi injections
F35E12.5	3.8	343 aa; DUF141 domain	RNAi feeding, N2 and *rrf-3*;RNAi injections
C50F4.8	1.7	300 aa; no identified domains	RNAi feeding, N2 and *rrf-3*
W02C12.3	1.3	361–524 aa; 18 alternatively spliced isoforms, bHLH transcription factor	RNAi feeding, N2 and *rrf-3*
F49F1.1	1.4	265 aa; ShTK domain	RNAi feeding, N2 and *rrf-3*; RNAi injections
F53A9.8	5.1	87 aa; histidine-rich protein	RNAi feeding, N2 and *rrf-3*;RNAi injections

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 5. Putative antimicrobial effectors induced by infection with *M. nematophilum*

Protein family	Induced gene(s)	Induction at 6 h	Proposed activity
Lysozyme	*lys-3*	1.7	Cleavage of bacterial cell-wall peptidoglycan
	lys-7	2.1	
	lys-8	1.5	
Lipase	*F54F3.3*	2.8	Acts directly against microoganisms
Thaumatin/PR-5	*thn-2*	2.9	Antifungal activity mediated by membrane permeabilization
Metridin/ShK domain	*F49F1.5*	2.2	Weak similarity to antimicrobial peptide from *Amaranthus caudatus* suggests antifungal activity
Invertebrate-lysozyme/ destabilase	*C45G7.2*	3.1	Similar to invertebrate-type family of lysozymes/destabilases (Zavalova et al. 2000; Bachali et al 2002)
	C45G7.3	4.3 (1 expt)	
Defense-related protein	*F49E11.10*	3.4	Defense-related protein, contains SCP domain
Zinc metallopeptidases	*F53A9.1*	2.2	Homology to *Drosophilia* transient response gene, induced rapidly following infection (Park et al. 2004)

were treated, they exhibited a severe Dar (swollen tail) phenotype after infection, whereas *rrf-3* worms (RNAi-hypersensitive) were Bus (un-swollen), possibly as a result of a threshold effect.

Other genes that may have an antimicrobial role include those encoding Metridin ShK toxin domains. We found genes containing this domain in the twofold up-regulated data (*F49F1.5*) and in the cluster of up-regulated genes on chromosome IV. RNAi of one of the latter genes containing a Metridin ShK toxin domain, *F49F1.1*, which is induced 1.4-fold, resulted in Bus (un-swollen) worms in the presence of *M. nematophilum* (Table 4).

One of the most highly induced genes, *F53A9.8* (5.1-fold up-regulated, Table 1), encodes a small (87 amino acids) histidine-rich protein and is found in a cluster of induced genes on the X chromosome (Fig. 4). The function of this protein is unknown; it has homology with the histidine-rich glycoprotein precursor of *Plasmodium lophura* and the *Drosophila* protein Catsup-PA, which is involved in catecholamine metabolism and melanization. *Catsup* mutants of *Drosophila* exhibit morphological defects of the cuticle and melanotic pseudotumors (Lemaitre et al. 1995). RNAi knockdown of *F53A9.8* results in Bus (un-swollen) worms in the presence of *M. nematophilum* (Table 4), suggesting that *F53A9.8* could also have a role in cuticle morphology in *C. elegans*, in view of the cuticular deformation involved in the Dar response. Melanization also provides part of the antibacterial defense mechanism in *Drosophila*, although it is not known to have such a function in nematodes.

Induced DUF141 proteins

Three of the 68 induced genes encode proteins that contain domain of unknown function 141 (DUF141), a *C. elegans*-specific domain (now also referred to as a CUB-like domain [PF02408, IPR00366]). This is 17 times more DUF141-containing proteins than would be expected by chance in this data set. A different gene containing this domain (*F55G11.4*) is induced by *S. marcescens* infection in *C. elegans* (Mallo et al. 2002).

The function of DUF141 remains to be elucidated and none of the 52 DUF141 genes in the genome have a phenotype in the RNAi screens conducted so far (WormBase WB144). Recently, however, DUF141-containing proteins have been found to contribute to longevity in *C. elegans* (Hamilton et al. 2005), so it is possible that the domain plays some role in general stress response or in an antibacterial response. RNAi of one DUF141 infection-induced gene, *F35E12.5*, produced Bus (un-swollen) worms in the presence of the pathogen, suggesting that this gene was required for the defensive swelling response (Table 4).

Finally, a gene highly induced in all three experiments was *tts-1* (transcribed telomerase-like sequence); this produces a noncoding RNA, and has been shown to be highly up-regulated when worms enter the dauer alternative developmental pathway under conditions of environmental stress (Jones et al. 2001) . This RNA is also up-regulated in long-lived *daf-2* and starved L1 larvae (Halaschek-Wiener et al. 2005). The biological role of *tts-1* is still unclear; it has been proposed to interact with telomeres or telomere-associated proteins or to have a more general role in altering chromatin structure under stress conditions. We grew the strain VC107, which contains a 295-bp deletion in *tts-1,* on MBL plates. When infected by *M. nematophilum*, the worms appeared to be slightly sicker than control wild-type worms, suggesting that although *tts-1* contributes to the response of *C. elegans* to *M. nematophilum*, it is not essential. Possibly *tts-1* may be induced as part of a general stress response that is activated as a result of infection.

Down-regulated genes

Using the same criteria we used to define induced genes, we found that 21 transcripts were down-regulated twofold or more in at least 2/3 of the biological triplicates (8/21 in all three replicates). Gene Ontology (http://www.ebi.ac.uk/GOA/) predicts that five of these repressed genes encode proteins with oxidoreductase activity (*dhs-3, dhs-25*, *acdh-2*, *F08A8.2*, and *CO2B10.1*). Nine others also include genes that encode proteins involved in metabolism, including an acyl-CoA thioesterase (*C17C3.1*), a 3-oxoacidCoA transferase (*C05C10.3*), an arginine kinase (*F44G3.2*), *hsp-60,* a sodium/phosphate transporter (*C35A5.3*), an acid phosphatase (*C05C10.4*), a folate transporter (*F27B4.7*), a glutathione-S-transferase (*gst-10*), and a copper ion

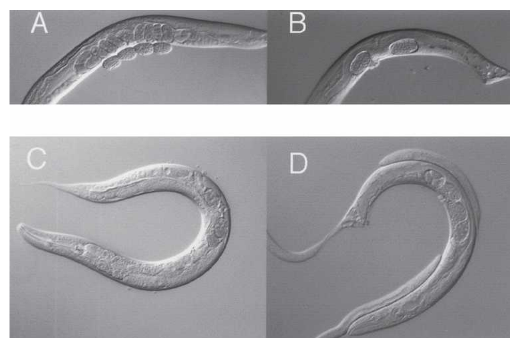

Figure 3. Lys-7 worms grown on MBL are severely constipated and developmentally delayed. Two mutants of *lys-7*, RB1286 *(ok1385)* and RB1285 *(ok1384)* were grown on MBL plates. Both *lys-7* alleles were severely constipated, had a strong Dar response, and a developmental delay, such that most of the population arrested at the L3 larval stage. (*A*) RB1285 on OP50; (*B*) RB1285 on MBL; (*C*) RB1286 on OP50; (*D*) RB1286 on MBL.

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

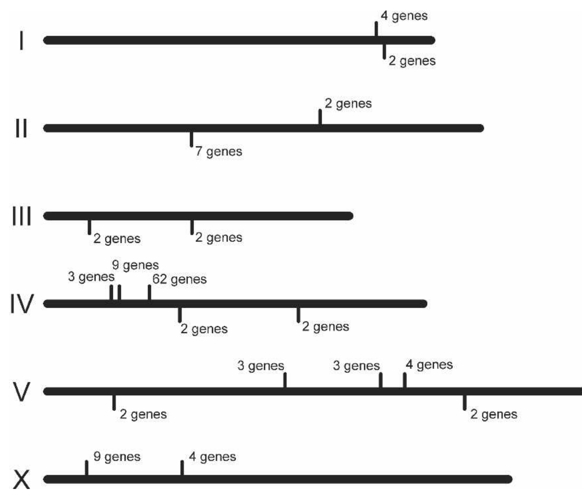

Figure 4. Clusters of genes whose transcript levels are altered following infection with *M.nematophilum*. Chromosomes are represented as black bars and the position and number of up-regulated (above chromosome) or down-regulated genes (below chromosome) are illustrated.

transporter (*F58G6.2*). In other respects, however, no obvious pattern can be seen among these genes. Conceivably, general down-regulation of metabolism may form part of the response to infection.

Genomic clustering of induced genes

We observed that a large proportion of infection-induced genes were found on chromosomes IV and V and that induced genes containing similar domains were often present in clusters of closely related or duplicated sequences. For example, 50 of the 58 genes in the *C. elegans* genome that contain DUF141 domains are found within five groups of five or more genes and these are, with one exception, found exclusively on chromosomes IV and V. A similar pattern was observed for the C-type lectin domain proteins, which are also found in clusters spread across the genome. The clustering of closely related genes may be the result of a selective pressure acting on these groups of genes as a result of coevolving host/pathogen interactions or as a defense against multiple pathogens (Schulenburg et al. 2004). In this study, we found one cluster of DUF141-encoding genes on chromosome V and a number of clusters of CTLD genes across the genome were induced following infection.

We asked whether other induced genes, including those induced to less than twofold, formed statistically significant spatial clusters within the genome. Using the algorithm described in the Methods section, we identified 18 spatial clusters of genes in which the sum of the genes' log ratios was significantly larger than would be expected by chance on a genome-wide basis (Supplemental Data S2; up-regulated and down-regulated clusters).

The analysis identified a number of both up-regulated and down-regulated clusters of genes. Ten clusters of two or more genes were up-regulated and eight clusters were down-regulated, as illustrated in Figure 4. Six of the up-regulated clusters were found on chromosomes IV and V, one was on chromosome I, one was on chromosome II and two were found on the X chromosome. The largest up-regulated cluster defined a 235-kb region on chromosome IV. Of the 62 predicted genes within this interval, 57 genes were covered by 67 microarray probes on the Affymetrix chips. A number of genes have multiple microarray probes, while others have none, and a few microarray probes correspond to predicted introns or intragenic regions (Supplemental Data S2). Of 20 genes that were scored as Present (see Methods), 18 were up-regulated and two were down-regulated. The partial data for the remaining genes indicate a similar pattern (see Supplemental Data 1 and 2), suggesting that almost all of the genes in this interval were up-regulated after infection. The genes in the cluster encoded numerous domains potentially implicated in *C. elegans* antimicrobial defense. For example, it included 13 genes with CTLD, three encoding galectins (galactoside-binding lectins), a number of genes for EGF-like proteins, four *srh* or *str* genes (for seven transmembrane receptors) *srh-220*, *srh-222*, *srh-223*, *str-122*, and four other genes encoding seven-pass *trans*-membrane proteins, *R07C12.1, R07C12.3, R07C12.4,* and *K08D10.10*. It is possible that all of these may play some role in pathogen-recognition and/or antimicrobial defense.

Within the other smaller clusters of up-regulated genes there were genes for other C-type lectins, DUF141 domains, transcription factors, noncoding RNAs and a protease. The induction of some of the genes within these clusters was modest, and a few were not up-regulated at this time point.

We chose a number of genes from the identified clusters to include in the RNAi feeding and injection experiments, testing both those that were highly up-regulated following infection and some that were modestly up-regulated, which does not preclude them from having an important role in the response to infection. For example, *C50F4.8*, encoding a protein with no defined PFAM domain, was induced modestly after 6 h (1.7-fold), but is found in a cluster with two other genes, *C50F4.1* and *C50F4.9*, that were all induced following infection. RNAi of *C50F4.8* and subsequent exposure to *M. nematophilum* resulted in Bus (un-swollen) worms in the presence of the pathogen. This gene was therefore grouped along with others that affect the morphogenetic swelling response in Table 4.

Three noteworthy genes that affect the swelling response are all found in the large cluster of genes on chromosome IV, i.e., *F47C12.1, F49F1.1*, and *W02C12.3*. First, *F47C12.1* encodes a large protein, 1827 amino acids long, and is homologous to fibrillins and related proteins that contain calcium-binding EGF-like domains. It contains a number of PFAM domains that have suggested extracellular functions; BLAST searches indicate strong homology with other larger multidomain proteins in mosquito, *Drosophila*, and human. The worm protein may have a similar function in response to infection since RNAi knock down of this gene resulted in Bus (un-swollen) worms in the presence of the pathogen.

A second up-regulated gene shown to affect the swelling response is *F49F1.1*, which encodes a secreted 265 amino acid protein with Metridin/ShK domains, as discussed above. The domain is found 124 times within the *C. elegans* genome and in three other proteins induced in *M. nematophilum* infection *(F49F1.5, F49F1.6, F49F1.7)*. These genes form a cluster with *F49F1.1* on chromosome IV and are all highly related. While the function of this domain is unknown, it appears to be required for the swelling response, as RNAi of *F49F1.1* resulted in worms that were Bus in the presence of the *M. nematophilum*; the three related genes have not been tested.

Third, *W02C12.3* encodes a predicted basic helix-loop-helix (bHLH) transcription factor, one of 42 such genes in the *C. elegans* genome (Reece-Hoyes et al. 2005). Like many other members of this family, it is alternatively spliced to give, in the case of

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

W02C12.3, 18 different transcripts. Expression studies have demonstrated its expression in multiple tissues (R. Johnsen, pers. comm.), but previous RNAi tests have yielded no obvious phenotype. The overall induction of this gene following 6 h of infection is not great (1.3-fold), but it may be that alternative splice forms are used for different infection-specific immune responses. At least one of these bHLH protein isoforms appears to be important for the Dar response, since RNAi knockdown of *W02C12.3* resulted in Bus worms following infection with *M. nematophilum*.

Coexpression of neighboring genes has been observed in a variety of contexts in humans (Lercher et al. 2002), plants (Williams and Bowles 2004; Ren et al. 2005), *Drosophila* (Spellman and Rubin 2002), *C. elegans* (Lercher et al. 2003), and yeast (Cohen et al. 2000). In *C. elegans*, Roy et al. (2002) first demonstrated the clustering of small numbers of muscle-expressed genes along chromosomes. More recently, 10 neighboring genes were found to be coregulated by VHL-1 (Bishop et al. 2004), while transcripts enriched during spermatogenesis were also found to be spatially clustered (Miller et al. 2004).

These previous studies in *C. elegans* (Roy et al. 2002; Lercher et al. 2003) demonstrated clustering of genes on a scale of <25 kb. In our data we observe one example of a large spatial cluster (>200 kb) as well as a number of smaller clusters (<25 kb). Lercher et al. (2003) have suggested that coexpression of neighboring genes in *C. elegans* is mostly due to operons and duplicated genes that share common regulatory sequences. This explanation may account for some of the coexpression of the smaller clusters that we observed, though most of the genes concerned are not in operons. However, the large cluster on LGIV is similar in size to the clusters of coexpressed genes observed in *Drosophilia* (Spellman and Rubin 2002). This suggests that this cluster may be regulated by reorganization of higher order chromatin structures. Alternatively, genes within the clusters may be directly and coordinately activated by transcription factors that bind to the promoters of these genes.

Overlap with other infections and innate immune pathways

There have been several studies on the transcriptional response of *C. elegans* to infection by a variety of pathogens including the Gram-negative bacterium *S. marcescens* (Mallo et al. 2002), the fungus *Drechmeria coniospora* (Couillault et al. 2004), and the Gram-positive bacterial pore-forming toxin, Cry5B (Huffman et al. 2004). We compared the genes induced by these infections/toxins with our data in order to examine the degree of specificity in the response to *M. nematophilum* infection.

None of the seven genes induced over twofold in the *S. marcescens* infection were induced by infection with *M. nematophilum*. However, there was a striking overlap in domain composition of the induced genes, including C-type lectins, lipases, and DUF141 proteins. We also note that the lysozyme genes *lys-7* and *lys-8* are induced by both types of infection.

We found no induction of the antimicrobial peptides induced by the *Drechmeria coniospora* infection in our data set (Couillault et al. 2004). This included all of the caenacins represented on the chip and the antimicrobials *nlp-29, nlp-31,* and *nlp-33* as well as the *tir-1* and *tol-1* proteins, which were also studied by Couillault et al. (2004) Similarly, we found no overlap between our data and the response of *C. elegans* to the Cry5B toxin (Huffman et al. 2004). This is not entirely surprising, because the p38 MAP kinase signal-transduction pathway, required for defense against Cry5B toxin, is not required for defense against *M. nematophilum* (Nicholas and Hodgkin 2004a).

A number of researchers have suggested that there is an overlap between genes involved in longevity and those that have an antimicrobial function (Kurz and Tan 2004). In *C. elegans* this overlap appears to include genes regulated by the DAF-2/DAF-16 insulin-like receptor pathway (Kurz and Tan 2004). Mutations in this pathway result in long-lived worms that are resistant to the normally fatal Gram-positive pathogens *Enterococcus faecalis* and *Staphylococcus aureus* (Garsin et al. 2003). Furthermore, the *lys-7* gene is derepressed in *daf-2* mutant worms and may contribute to longevity, since RNAi of *lys-7* in a *daf-2* mutant worm results in a reduced life span (Murphy et al. 2003). We examined our set of changing genes for overlap with published lists of DAF-2/DAF-16-regulated genes (Lee et al. 2003; McElwee et al. 2004; Halaschek-Wiener et al. 2005). Although our analysis was not a comparison of whole-genome microarray data sets, but only of the published gene lists, we found only limited overlap in both induced and repressed data sets. This supports our observations (K.J. Yook and J. Hodgkin, in prep.) that *daf-2* and *daf-16* do not play a major role in defense against *M. nematophilum*.

Discussion

This study has shown that a specific and distinctive set of *C. elegans* genes change in their expression levels after infection by a rectal bacterial pathogen. We have applied stringent criteria to define this set of genes, using synchronized populations, threefold replication of a single time point, and comparing the response between exposure to virulent and avirulent bacteria. Undoubtedly, more changes in gene expression could be detected if these criteria were relaxed, for example, by looking at different life-stages, different infection conditions, and by following time-courses. In addition, we focused mainly on genes with more than twofold changes in gene expression as detected in total RNA, so we may have missed strongly responding genes if they are expressed in only a few cells.

Our microarray observations and the subsequent functional tests support the belief that many genes that lack obvious function in global phenotype screens may have specific roles in defense against pathogens. It is noteworthy that so many *clec* genes (encoding C-type lectin domain proteins) are up-regulated in this infection, and also that different *clec* genes have been seen to be up-regulated in other types of infection. These CTLD proteins could be acting as part of a pathogen-detection system, as antibacterial effectors, or as some combination of these functions. The limited functional tests we have carried out so far support the idea that they do indeed contribute to defense. Given the large size of the *clec* gene family (more than 170 members) and the fact that they are predicted to be mostly secreted proteins, we speculate that CTLD proteins may constitute a major effector arm of innate immunity in nematodes and perhaps in other animals too.

A number of other kinds of genes were found to be up-regulated, such as those belonging to families encoding known antimicrobial factors, such as lysozymes and thaumatins, and also gene families of unknown function, such as DUF141 proteins. Our RNAi experiments provide preliminary evidence for their involvement in defense, but proper understanding of what each contributes will require more detailed experimentation on a gene-by-gene basis.

Lastly, our experiments reveal the presence of clusters of

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

genes that change their expression coordinately after infection. The cluster of 62 genes on chromosome IV is particularly striking. Both this and most of the other clusters contain a diversity of genes, so their existence is not simply a consequence of local gene duplications. It remains to be seen whether the clustering is due to shared *cis*-regulatory elements or to longer-range chromatin domain effects, but it does not seem to be associated with organization of responding genes into operons. One can speculate that these clusters represent "resistance islands," which have evolved to create a coordinated defense against this particular infection. Such islands would not have the genomic mobility associated with pathogenicity islands in prokaryotes, but nevertheless may allow more rapid modulation and evolution of defense factors. It is conceivable that other islands of this type will become apparent as more of the genomic responses of *C. elegans* to other kinds of specific infection or stress are examined.

Methods

C. elegans culture conditions

Growth and manipulation of *C. elegans* were as previously described (Brenner 1974; Sulston and Hodgkin 1988).

C. elegans strains

The following three strains were obtained from the *C. elegans* Genetics Center (CGC): VC107 *tts-1(gk105)*, RB1285 *lys-7(ok1384)*, and RB1286 *lys-7(ok1385)*.

A strain carrying a deletion allele *(tm1954)* of the lipase gene *F54F3.3* was obtained from Shohei Mitani (Tokyo Women's University School of Medicine).

Infection in liquid culture

To reduce transcriptional changes caused by exposure to different bacterial food sources, we compared infection with virulent *M. nematophilum* to infection with an avirulent mutant of *M. nematophilum* that fails to induce a Dar response and does not adhere to the worm cuticle (Fig. 1). The avirulent mutant was created using UV mutagenesis (T. Akimkina, K. Yook, S. Curnock, J. Hodgkin, in prep.). To assay the early responses to infection, we chose the earliest time point (6 h) at which infection could be seen in >90% of the population, defined as slight swelling of the tail region and adherence of bacteria to the rectal cuticle. The control and test infections were repeated three times so that the microarray data could be compared in biological triplicate. We performed one additional comparison of infected/uninfected worm RNA isolated from worms infected for 12 h.

To obtain a synchronized culture of worms, gravid N2 worms were treated with alkaline hypochlorite to release embryos. Embryos were hatched to L1 stage by shaking in M9 buffer at 20°C for 18 h. The L1s were then split equally between two 2-L glass flasks containing S-medium and 15 g of *E. coli* HB101/500 mL and grown for 24 h at 25°C to L3 larval stage. After 24 h, 1.5g/500 mL of *M. nematophilum*, either virulent CBX102 or avirulent UV336 were added. Worms were sampled before addition of *M. nematophilum* and again after 6 or 12 h for staining with SYTO13 to monitor infection before harvesting them for RNA extraction.

SYTO 13 staining

Worms (1 mL) from liquid culture were allowed to settle in 9 mL of Tris-buffered saline (TBS) for 60 min to allow for digestion of bacteria in the gut; 60 µL of settled worms were then added to 300 µL of 15 µM SYTO 13 (Molecular Probes) in TBS and left for 30–60 min. Then, 3 µL of settled worms were added to 3 µL of 5% propylene phenoxytol in M9 on a coverslip, inverted, and viewed by Nomarski and UV fluorescence microscopy.

RNA sample preparation

After 6 h of infection, worms were harvested by three centrifugations at 1500 rpm for 2 min at 4°C and resuspended in chilled M9 buffer to remove bacteria; RNA was extracted using Trizol (Invitrogen) following the manufacturer's instructions.

Each infection with virulent *M. nematophilum* (CBX102) and the control avirulent UV336 was done in biological triplicate. Hybridization to Affymetrix *C. elegans* GeneChips was performed in the core facility of the MRC Functional Genetics Unit at the Department of Human Anatomy and Genetics, Oxford, UK following Affymetrix protocols to determine transcript levels.

Microarrays

RNA produced was quantified using NanoDrop ND-1000, and RNA quality was checked using RNA 6000 Nano Assay on Agilent bioanalyzer 2100 (Agilent). We used commercially available high-density oligonucleotide *C. elegans* genome arrays (Affymetrix Cat. No. 900,383, containing 22,500 probe sets against 22,150 unique transcripts). Eight arrays in total were used in this study (three pairs for the 6-h infection time point, and one pair for the 12-h time point). All procedures and hybridizations were performed according to the GeneChip expression technical manual (Affymetrix). In brief, 10 µg of RNA was converted to double-stranded cDNA with superscript II (Invitrogen) using T7-$(dT)_{24}$ primer (Affymetrix). The cDNA was then cleaned on GeneChip sample clean up column (Affymetrix). Then biotin-labeled cRNA probes were prepared using a Enzo Bioarray High yield RNA Transcript labeling kit (Affymetrix). The resulting cRNA was purified using GeneChip sample clean up. After the cRNA passed a quality control check on a bioanalyzer, we fragmented the cRNA to ~200 bp (fragmentation buffer used 200 mM Tris-acetate at pH8.2, 500 mM potassium acetate, 150 mM magnesium acetate at 94°C for 35 min). Fragmented cRNA (15 µg) was used in a 300-µL hybridization cocktail containing spiked controls (Affymetrix), 0.1 mg/mL Herring Sperm DNA (Promega), 0.5 mg/mL Acetylated BSA (Invitrogen). A total of 200 µL of this hybridization cocktail was used on each chip and incubated at 45°C for 16 h in the hybridization oven, rotating at 60 rpm. Following hybridization, the arrays were processed using a GeneChip Fluidics Station 400 according to recommended protocols (EukGE-WS2v4, Affymetrix) of double-staining and post-hybridization washes. Fluorescent images were captured using gene Array Scanner 2500 (Affymetrix). The GeneChip design, using multiple probes with both perfect and imperfect matches to target sequences, allows distinction between members of gene families such as CTLD genes, apart from the single case of *clec-15* and *clec-13*, which are recently duplicated loci with identical coding sequence.

Data analysis and QC

Gene transcript levels were determined from data image files using GeneChip Operating Software (MAS5.0 and GCOS1.0, Affymetrix). Global scaling was performed to compare genes from chip to chip; each chip was normalized to the same target intensity (TGT = 100). Data quality controls met the Affymetrix quality assessment guidelines. For each replicate and each probe, log fold changes between infected and uninfected samples were calculated only using normalized data classified as Present

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

by MAS5.0 software. Primary data have been deposited on ArrayExpress, accession number E-MEXP-696.

Identification of significantly induced genes

The experimental design consisted of three biological replicates, each subdivided into two synchronized worm populations, one of which was infected with virulent *M. nematophilum* (CBX102) and one infected with avirulent *M. nematophilum* (UV336). In total, six hybridizations were performed, and the ratio of the expressions of each tag was computed within each biological replicate, resulting in three replicates of ratio data, on which all analyses were based. Because there were no technical replicates, and in many cases at least one probe was marked as Absent by MAS5.0, the number of observations per probe was often too small to perform analysis of variance, which requires at least three observations per probe. Instead, we performed a conservative nonparametric analysis.

We defined a probe that is induced or repressed at least twofold in a replicate to be changed. We consider the set of probes that, across replicates marked as Present by MAS5.0, were both (1) changed in at least two replicates and (2) changed on average. The second criterion is redundant when all three replicates are changed, but helps control situations where one replicate is markedly different from the others. We show that these probes have a false discovery rate (FDR) of ~5%.

Let x_{it} be the $\log_2$ ratio of infected to uninfected normalized intensities for replicate $i = 1,2,3$ and probe t, calculated as $\log_2$[(infected+20)/(uninfected+20)], the offset 20 being added to reduce the likelihood that weakly expressed genes would give spuriously high ratios by chance. Let $p_i(r)$ be the observed proportion of probes in experiment i for which $x_{it} > r$, and $q_i(r)$ be the observed proportion of probes in experiment i for which $x_{it} < -r$. Thus, if $r = 1$, $p_i(1)$ is the proportion of probes in experiment i that are induced at least twofold, and $q_i(1)$ is the corresponding proportion repressed at least twofold. Then, the probability that a probe is induced by at least r or repressed by at least $-r$ in all three experiments is

$$P_3(r) = p_1(r)p_2(r)p_3(r) + q_1(r)q_2(r)q_3(r)$$

Similarly, the probability that a probe is changed by at least r in exactly two out of three replicates is

$$P_2(r) = p_1(r)p_2(r)(1-p_3(r)) + p_1(r)(1-p_2(r))p_3(r) + (1-p_1(r))p_2(r)p_3(r) + q_1(r)q_2(r)(1-q_3(r)) + q_1(r)(1-q_2(r))q_3(r) + (1-q_1(r))q_2(r)q_3(r)$$

For a given threshold r, the number of probes changed in at least two of three experiments is distributed as a Poisson random variable with mean and variance

$$E(r) = N(P_2(r) + P_3(r)),$$

where $N = 12403$, the number of probes for which at least one replicate was Present according to MAS5.0. Supplemental data S4 gives the raw counts and values for $p_i(1)$ and $q_i(1)$, from which it follows that $NP_3(1) = 0.0192$, $NP_2(1) = 5.510$, and $E(1) = 5.5292$.

We observed 41 probes changed in all three replicates (Supplemental Table S1) and a further 66 changed in two of three replicates (data not shown). Hence, of the total of 107 probes, we expect 5.292 false positives, or a FDR of 4.9%. Note that since the expected number changed in all three replicates is 0.0192, compared with 41 observed, this latter set is extremely significant (see Supplemental data S4). Of the 107 probes, we focus on just the 93 that were also changed on average, which correspond to 89 genes. Of the 93 probes, 71 (68 genes) were induced and 22 (21 genes) repressed. The induced genes are listed in Table 1. The full list is in Supplemental data S3.

There were also 13 changed probes for which only one replicate gave Present data and for which it is not possible to assign statistical significance. There were a further 10 changed probes with more than one Present replicate but with fewer than two significantly changed replicates. All of these probes were omitted from the analysis.

We conducted one additional microarray experiment on RNA isolated from L3 worms after 12-h infection with *M. nematophilum* CBX102 or UV336. Although with only one experiment it is not possible to assign any statistical significance to the data, we found that a similar number of probes were up-regulated over twofold (65) and there was a significant overlap in the identity of the genes up-regulated at 6 and 12 h. Futhermore, we found that the level of up-regulation for the majority of genes induced at 6 h had increased further by 12 h.

Spatial clustering of genes

This analysis uses all probes, not just those classified as changed twofold, and is therefore able to detect regions in which the majority of probes are up- or down-regulated, even if they are not individually significant. We used the ungapped Smith-Waterman algorithm (Smith and Waterman 1981; Karlin and Altschul 1990). This technique has been applied to genome hybridization data to find aneusomic regions (Price et al. 2005). First, all probes with Present data were ordered across the genome. The probe DNA sequences were mapped to Wormpep http://www.sanger.ac.uk/Projects/C_elegans/WORMBASE/current/wormpep.shtml with BLASTP. WormBase was then queried for gene annotations and coordinates. Next, we subtracted the 90th percentile of the probe log ratios from the log ratio for each probe to define a probe score, such that the scores are predominantly negative. Then we identified all high-scoring segments. These are runs of adjacent probes such that the sum of their scores is a local maximum where it is not possible to make a higher-scoring segment by expanding or contracting the segment. The statistical significance of the segments was determined by permutation; the scores were shuffled between the probes, and the highest-scoring segment across the genome determined. This was repeated 10,000 times to give the distribution of the highest scoring segment to be expected by chance genome-wide. We found seven clusters of induced genes at genome-wide $P < 0.05$ and a further three at $P < 0.1$. There were seven clusters of repressed genes with $P < 0.05$ and a further cluster at $P < 0.1$ (Supplemental data S2 clusters).

Semiquantitative RT–PCR

To confirm expression changes in representative up-regulated genes, semiquantitative RT–PCR was carried out on the C-type lectin gene *ZK666.6* (*clec-60*) and another up-regulated gene *F21C10.10*, using SuperScript III kit (Invitrogen). Expression was compared with expression of the standard reference gene *ama-1,* after a 6-h exposure to avirulent or virulent *M. nematophilum*. Samples were taken at 20 and 25 cycles, within the linear range of the amplification. Primers and expected products (size in basepairs) were as follows: (amplified products corresponded to the cDNA size); for *ama-1* 5′-GCCGCTGCGCATTGTCTC-3′ and 5′-CAGAATTTCCAGCACTCGAGGAGCGGA-3′ (gDNA 492bp, cDNA 346bp); for *ZK666.6* 5′-GCAAACAAGGGTACAACCGT-3′ and 5′-GTGGTTGATCCCATACCCAG-3′ (gDNA 646bp, cDNA 552bp); for *F21C10.10* 5′-CGATTCAAGTGTCATCAAGGCTATC-3′ and 5′-TGCTGGAGCATTGAGTGGTGTG-3′ (gDNA 654bp, cDNA 522bp).

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

Reporter constructs/promoter GFP constructs

To examine the expression patterns of the C-type lectin encoded by *ZK666.6* (*clec-60*), a promoter GFP construct was made. The following primers were used to amplify a 1.64-kb promoter fragment of *clec-60*: forward: GCAACAGTTTTCACATCATTTCTAC GAGG; reverse: CATTGTGAGAATTGCAATAAAAGTGAAATCC.

The promoter fragments were fused to GFP from the plasmid pPD95.75 using the protocol of Hobert (2002) The PCR products were injected along with a plasmid carrying the *unc-119(+)* gene into *unc-119* worms. Transformed worms were selected by their ability to move normally before screening for GFP expression. For high-power DIC/UV microscopy, worms were washed off OP50 or MBL plates and then mounted in 5% propylene phenoxytol in M9.

RNAi methods

Where possible, we tested genes for which there were available knockout alleles. Where alleles were not available, we used knockdown of gene expression with RNAi, initially by feeding and then following up suggestive results with RNAi injections. We modified the method of Kamath et al. (2001) to conduct both the feeding screens and RNAi injections. RNAi experiments were done on both N2 worms and the RNAi hypersensitive strain *rrf-3* (Simmer et al. 2003).

Bacterial strains were streaked from the RNAi feeding library (Kamath and Ahringer 2003) onto LB amp plates (ampicillin 50 µg/mL) and allowed to grow at 37°C overnight. Bacteria were then used to inoculate 5 mL of LB-amp (ampicillin 50 µg/mL) and grown at 37°C for 8–12 h before being spread onto dry NGM +1 mM IPTG and 25-µg/mL carbencillin plates. IPTG induction was at 22°C for 18 h.

N2 or *rrf-3* worms were treated with alkaline hypochlorite and embryos allowed to hatch at 15°C for 72 h until they had reached the L3 larval stage. The L3 worms were washed off the plates with M9 and allowed to digest bacteria for 30 min before being picked to NGM + 1 mM IPTG and 25 µg/mL carbencillin with RNAi feeding bacteria for each of the genes to be tested. Each RNAi experiment was performed on six separate plates. After 48 h at 20°C on the RNAi plates, the worms were moved to mixed bacterial lawn (MBL) plates (*E. coli* OP50 and 1% *M. nematophilum* [v/v]) and left overnight at 20°C (allowing the *ts* sterile *rrf-3* to lay eggs) before being moved to 25°C. The progeny of the parent worms were scored on the RNAi feeding plates (pure *E. coli)* for any obvious abnormalities and on the MBL plates for changes in response to infection.

The RNAi injections were as described in Fire et al. (1998). Templates for double-stranded (ds) RNAi were made by PCR with the T7 primer of the L4440-based feeding constructs. RNA was synthesized using the Ambion Megascript kit and dsRNA was made by annealing at 68°C for 10 min, then at 37°C for 30 min. RNA was then phenol/chloroform extracted and precipitated before being resuspended in TE prior to injection. Injected worms were allowed to recover for 4–6 h at 20°C on OP50 plates before being transferred to MBL plates and moved to 20°C overnight, followed by 25°C (for *rrf-3*), or straight to 25°C (for N2). Progeny were scored as for the RNAi feeding experiments. We found RNAi by injection phenotypes were similar to those found in the initial feeding screens.

When conducting the RNAi knockdowns, changes in the Dar response, such as an increase or decrease in swelling, sickness and slowing of growth, larval lethality, and any increase in constipation were all scored. Most of these phenotypes are difficult to score quantitatively; therefore, experiments were scored blind using two independent scorers. Reproducibility in at least two separate experiments was required to define a change in response to infection. Each RNAi feeding experiment for each gene had six replicates and injections were scored on the progeny of five or more injected worms.

We expected to see some reduction in the morphogenetic swelling response if the targeted protein was involved in this process, or to sicker Dar worms, if the genes were involved in an antibacterial response. We also expected to see overlapping functions, e.g., sick Bus worms, similar to those mutants in the ERK MAP kinase pathway (Nicholas and Hodgkin 2004a). No change in response would suggest that the protein was redundant or not critical for response, or else, that the RNAi treatment had been ineffective. In all cases, the response of worms following RNAi was variable, with a low penetrance for the RNAi-induced phenotype. This may be because we were examining a post-embryonic phenotype, for which RNAi has been shown to be less effective.

A further limitation of the RNAi method is the problem of secondary targets where RNAi of one gene, the primary target, may also lead to a reduction in transcript levels from other highly homologous genes, known as secondary targets. Potential secondary targets have between 80% and 95% sequence identity across at least 200 bp of the RNAi probe and are annotated on WormBase. This may be the case for closely related genes involved in antimicrobial defense. For example, the RNAi probe against *clec-13(H16D19.1)* is 100% identical to *clec-15(T07D10.4)* and is expected to affect both transcripts. More rigorous tests will become possible when knock-out alleles of induced genes become available.

Testing alleles

Deletion and point mutation alleles of induced genes were available from the *C. elegans* Genetics Center. These strains were also checked carefully for changes in response to infection. Each strain was tested in triplicate in at least two separate experiments, comparing the response of the allele to that of N2.

Acknowledgments

We thank the *Caenorhabditis* Genetics Center and Shohei Mitani for strains, Maria Gravato-Nobre for advice on RNAi feeding, Jen Taylor for advice on analysis, Jacopo Novelli for injections, and Frederick Partridge, Petros Ligoxygaxis, and Jacopo Novelli for critical reading of the manuscript. M.D. was supported by the Timothy Bailey Trust. This work was funded by the Medical Research Council UK.

References

Aballay, A. and Ausubel, F.M. 2001. Programmed cell death mediated by *ced-3* and *ced-4* protects *Caenorhabditis elegans* from *Salmonella typhimurium*-mediated killing. *Proc. Natl. Acad. Sci.* **98:** 2735–2739.

Bachali, S., Jager, M., Hassanin, A., Schoentgen, F., Jolles, P., Fiala-Medioni, A., and Deutsch, J.S. 2002. Phylogenetic analysis of invertebrate lysozymes and the evolution of lysozyme function. *J. Mol. Evol.* **54:** 652–664.

Bishop, T., Lau, K.W., Epstein, A.C., Kim, S.K., Jiang, M., O'Rourke, D., Pugh, C.W., Gleadle, J.M., Taylor, M.S., Hodgkin, J., et al. 2004. Genetic analysis of pathways regulated by the von Hippel-Lindau tumor suppressor in *Caenorhabditis elegans*. *PLoS Biol.* **2:** e289.

Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77:** 71–94.

Cohen, B.A., Mitra, R.D., Hughes, J.D., and Church, G.M. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* **26:** 183–186.

Couillault, C., Pujol, N., Reboul, J., Sabatier, L., Guichou, J.F., Kohara, Y., and Ewbank, J.J. 2004. TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor

Downloaded from genome.cshlp.org on June 15, 2026 . Published by Cold Spring Harbor Laboratory Press

protein TIR-1, an ortholog of human SARM. *Nat. Immunol.* **5:** 488–494.
De Gregorio, E., Spellman, P.T., Rubin, G.M., and Lemaitre, B. 2001. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc. Natl. Acad. Sci.* **98:** 12590–12595.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391:** 806–811.
Franc, N.C. and White, K. 2000. Innate recognition systems in insect immunity and development: New approaches in *Drosophila*. *Microbes Infect.* **2:** 243–250.
Garsin, D.A., Villanueva, J.M., Begun, J., Kim, D.H., Sifri, C.D., Calderwood, S.B., Ruvkun, G., and Ausubel, F.M. 2003. Long-lived *C. elegans daf-2* mutants are resistant to bacterial pathogens. *Science* **300:** 1921.
Gravato-Nobre, M.J. and Hodgkin, J. 2005. *Caenorhabditis elegans* as a model for innate immunity to pathogens. *Cell. Microbiol.* **7:** 741–751.
Halaschek-Wiener, J., Khattra, J.S., McKay, S., Pouzyrev, A., Stott, J.M., Yang, G.S., Holt, R.A., Jones, S.J., Marra, M.A., Brooks-Wilson, A.R., et al. 2005. Analysis of long-lived *C. elegans daf-2* mutants using serial analysis of gene expression. *Genome Res.* **15:** 603–615.
Hamilton, B., Dong, Y., Shindo, M., Liu, W., Odell, I., Ruvkun, G., and Lee, S.S. 2005. A systematic RNAi screen for longevity genes in *C. elegans*. *Genes & Dev.* **19:** 1544–1555.
Hobert, O. 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques* **32:** 728–730.
Hodgkin, J., Kuwabara, P.E., and Corneliussen, B. 2000. A novel bacterial pathogen, *Microbacterium nematophilum*, induces morphological change in the nematode *C. elegans*. *Curr. Biol.* **10:** 1615–1618.
Huffman, D.L., Abrami, L., Sasik, R., Corbeil, J., van der Goot, F.G., and Aroian, R.V. 2004. Mitogen-activated protein kinase pathways defend against bacterial pore-forming toxins. *Proc. Natl. Acad. Sci.* **101:** 10995–11000.
Jones, S.J., Riddle, D.L., Pouzyrev, A.T., Velculescu, V.E., Hillier, L., Eddy, S.R., Stricklin, S.L., Baillie, D.L., Waterston, R., and Marra, M.A. 2001. Changes in gene expression associated with developmental arrest and longevity in *Caenorhabditis elegans*. *Genome Res.* **11:** 1346–1352.
Kamath, R.S. and Ahringer, J. 2003. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods* **30:** 313–321.
Kamath, R.S., Martinez-Campos, M., Zipperlen, P., Fraser, A.G., and Ahringer, J. 2001. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis ele*gans. Genome Biol. **2:** research0002.
Karlin, S. and Altschul, S.F. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci.* **87:** 2264–2268.
Kim, D.H., Feinbaum, R., Alloing, G., Emerson, F.E., Garsin, D.A., Inoue, H., Tanaka-Hino, M., Hisamoto, N., Matsumoto, K., Tan, M.W., et al. 2002. A conserved p38 MAP kinase pathway in *Caenorhabditis elegans* innate immunity. *Science* **297:** 623–626.
Kitajima, S. and Sato, F. 1999. Plant pathogenesis-related proteins: Molecular mechanisms of gene expression and protein function. *J. Biochem.* **125:** 1–8.
Kurz, C.L. and Tan, M.W. 2004. Regulation of aging and innate immunity in *C. elegans*. *Aging Cell* **3:** 185–193.
Lee, S.S., Kennedy, S., Tolonen, A.C., and Ruvkun, G. 2003. DAF-16 target genes that control *C. elegans* life-span and metabolism. *Science* **300:** 644–647.
Leippe, M. 1999. Antimicrobial and cytolytic polypeptides of amoeboid protozoa–effector molecules of primitive phagocytes. *Dev. Comp. Immunol.* **23:** 267–279.
Lemaitre, B., Meister, M., Govind, S., Georgel, P., Steward, R., Reichhart, J.M., and Hoffmann, J.A. 1995. Functional analysis and regulation of nuclear import of dorsal during the immune response in *Drosophila*. *EMBO J.* **14:** 536–545.
Lercher, M.J., Urrutia, A.O., and Hurst, L.D. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.* **31:** 180–183.
Lercher, M.J., Blumenthal, T., and Hurst, L.D. 2003. Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Res.* **13:** 238–243.
Mallo, G.V., Kurz, C.L., Couillault, C., Pujol, N., Granjeaud, S., Kohara, Y., and Ewbank, J.J. 2002. Inducible antibacterial defense system in *C. elegans*. *Curr. Biol.* **12:** 1209–1214.
McElwee, J.J., Schuster, E., Blanc, E., Thomas, J.H., and Gems, D. 2004. Shared transcriptional signature in *Caenorhabditis elegans* Dauer larvae and long-lived daf-2 mutants implicates detoxification system in longevity assurance. *J. Biol. Chem.* **279:** 44533–44543.
Miller, M.A., Cutter, A.D., Yamamoto, I., Ward, S., and Greenstein, D. 2004. Clustered organization of reproductive genes in the *C. elegans* genome. *Curr. Biol.* **14:** 1284–1290.
Millet, A.C. and Ewbank, J.J. 2004. Immunity in *Caenorhabditis elegans*. *Curr. Opin. Immunol.* **16:** 4–9.
Mochii, M., Yoshida, S., Morita, K., Kohara, Y., and Ueno, N. 1999. Identification of transforming growth factor-β-regulated genes in *Caenorhabditis elegans* by differential hybridization of arrayed cDNAs. *Proc. Natl. Acad. Sci.* **96:** 15020–15025.
Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Li, H., and Kenyon, C. 2003. Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* **424:** 277–283.
Mylonakis, E. and Aballay, A. 2005. Worms and flies as genetically tractable animal models to study host-pathogen interactions. *Infect. Immun.* **73:** 3833–3841.
Nicholas, H.R. and Hodgkin, J. 2004a. The ERK MAP kinase cascade mediates tail swelling and a protective response to rectal infection in *C. elegans*. *Curr. Biol.* **14:** 1256–1261.
———. 2004b. Responses to infection and possible recognition strategies in the innate immune system of *Caenorhabditis elegans*. *Mol. Immunol.* **41:** 479–493.
Park, J.M., Brady, H., Ruocco, M.G., Sun, H., Williams, D., Lee, S.J., Kato Jr., T., Richards, N., Chan, K., Mercurio, F., et al. 2004. Targeting of TAK1 by the NF-κB protein Relish regulates the JNK-medicated immune response in *Drosophila*. Targeting of TAK1 by the NF-κB protein Relish regulates the JNK-mediated immune response in *Drosophila*. *Genes & Dev.* **18:** 584–594.
Price, T.S., Regan, R., Mott, R., Hedman, A., Honey, B., Daniels, R.J., Smith, L., Greenfield, A., Tiganescu, A., Buckle, V., et al. 2005. SW-ARRAY: A dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data. *Nucleic Acids Res.* **33:** 3455–3464.
Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., and Walhout, A.J. 2005. A compendium of *Caenorhabditis elegans* regulatory transcription factors: A resource for mapping transcription regulatory networks. *Genome Biol.* **6:** R110.
Ren, X.Y., Fiers, M.W., Stiekema, W.J., and Nap, J.P. 2005. Local coexpression domains of two to four genes in the genome of *Arabidopsis*. *Plant Physiol.* **138:** 923–934.
Roy, P.J., Stuart, J.M., Lund, J., and Kim, S.K. 2002. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418:** 975–979.
Schulenburg, H., Kurz, C.L., and Ewbank, J.J. 2004. Evolution of the innate immune system: The worm perspective. *Immunol. Rev.* **198:** 36–58.
Simmer, F., Moorman, C., van der Linden, A.M., Kuijk, E., van den Berghe, P.V., Kamath, R.S., Fraser, A.G., Ahringer, J., and Plasterk, R.H. 2003. Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain reveals novel gene functions. *PLoS Biol.* **1:** e12.
Smith, T.F. and Waterman, M.S. 1981. Identification of common molecular subsequences. *J. Mol. Biol.* **147:** 195–197.
Spellman, P.T. and Rubin, G.M. 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1:** 5.
Sulston, J. and Hodgkin, J. 1988. Methods. In *The nematode* Caenorhabditis elegans. (eds. W.B. Wood et al.) pp. 587–606. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Williams, E.J. and Bowles, D.J. 2004. Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res.* **14:** 1060–1067.
Wilson, R., Chen, C., and Ratcliffe, N.A. 1999. Innate immunity in insects: The role of multiple, endogenous serum lectins in the recognition of foreign invaders in the cockroach, *Blaberus discoidalis*. *J. Immunol.* **162:** 1590–1596.
Zavalova, L., Lukyanov, S., Baskova, I., Snezhkov, E., Akopov, S., Berezhnoy, S., Bogdanova, E., Barsova, E., and Sverdlov, E.D. 1996. Genes from the medicinal leech (*Hirudo medicinalis)* coding for unusual enzymes that specifically cleave endo-ε (γ-Glu)-Lys isopeptide bonds and help to dissolve blood clots. *Mol. Gen. Genet.* **253:** 20–25.
Zavalova, L.L., Baskova, I.P., Lukyanov, S.A., Sass, A.V., Snezhkov, E.V., Akopov, S.B., Artamonova, I.I., Archipova, V.S., Nesmeyanov, V.A., Kozlov, D.G., et al. 2000. Destabilase from the medicinal leech is a representative of a novel family of lysozymes. *Biochim. Biophys. Acta* **1478:** 69–77.

Received December 20, 2005; accepted in revised form April 10, 2006.