BMC Biology

BioMed Central
The Open Access Publisher

Gene-environment and protein degradation signatures characterize genomic and phenotypic diversity in wild Caenorhabditis elegans populations

BMC Biology 2013, **11**:93 doi:10.1186/1741-7007-11-93

Rita JM Volkers (Rita.Volkers@wur.nl)
L Basten Snoek (Basten.Snoek@wur.nl)
Caspara J van Hellenberg Hubar (Caspara.vanhellenberghubar@wur.nl)
Renata Coopman (Renata.Coopman@UGent.be)
Wei Chen (chenwei.yl@hotmail.com)
Wentao Yang (wyang@zoologie.uni-kiel.de)
Mark G Sterken (Mark.Sterken@wur.nl)
Hinrich Schulenburg (hschulenburg@zoologie.uni-kiel.de)
Bart P Braeckman (Bart.Braeckman@ugent.be)
Jan E Kammenga (Jan.Kammenga@wur.nl)

ISSN 1741-7007

Article type Research article

Submission date 3 July 2013

Acceptance date 14 August 2013

Publication date 19 August 2013

Article URL http://www.biomedcentral.com/1741-7007/11/93

Gene-environment and protein degradation signatures characterize genomic and phenotypic diversity in wild *Caenorhabditis elegans* populations

Rita JM Volkers[1,†]
Email: Rita.Volkers@wur.nl

L Basten Snoek[1,†]
Email: Basten.Snoek@wur.nl

Caspara J van Hellenberg Hubar[1]
Email: Caspara.vanhellenberghubar@wur.nl

Renata Coopman[2]
Email: Renata.Coopman@UGent.be

Wei Chen[3]
Email: chenwei.yl@hotmail.com

Wentao Yang[3]
Email: wyang@zoologie.uni-kiel.de

Mark G Sterken[1]
Email: Mark.Sterken@wur.nl

Hinrich Schulenburg[3]
Email: hschulenburg@zoologie.uni-kiel.de

Bart P Braeckman[2]
Email: Bart.Braeckman@ugent.be

Jan E Kammenga[1*]
[*] Corresponding author
Email: Jan.Kammenga@wur.nl

[1] Laboratory of Nematology, Wageningen University, Droevendaalsesteeg 1, Wageningen 6708PB, The Netherlands

[2] Biology Department, Ghent University, Proeftuinstraat 86 N1, B-9000 Gent, Belgium

[3] Department of Evolutionary Ecology and Genetics, Zoological Institute, Christian Albrechts-Universitaet zu Kiel, Am Botanischen Garten 1-9, Kiel 24118, Germany

[†] Equal contributors.

Abstract

Background

Analyzing and understanding the relationship between genotypes and phenotypes is at the heart of genetics. Research in the nematode *Caenorhabditis elegans* has been instrumental for unravelling genotype-phenotype relations and has important implications for understanding the biology of mammals. But almost all studies, including forward and reverse genetic screens, are limited by investigations in only one canonical genotype. This hampers the detection and functional analysis of allelic variants which play a key role in controlling many complex traits. It is therefore essential to explore the full potential of the natural genetic variation and evolutionary context of the genotype-phenotype map in wild *C. elegans* populations.

Results

We used multiple wild *C. elegans* populations freshly isolated from local sites to investigate the gene sequence polymorphisms and a multitude of phenotypes including the transcriptome, fitness and behavioural traits. The genotype, transcriptome and a number of fitness traits showed a direct link with the original site of the strains. The separation between the isolation sites was prevalent on all chromosomes however chromosome *V* contributed the most to this variation. These results were supported by a differential food preference of the wild isolates for naturally co-existing bacterial species. Comparing polymorphic genes between the populations to a set of genes extracted from 19 different studies on gene expression in *C. elegans* exposed to biotic and abiotic factors, such as bacteria, osmotic pressure and temperature, revealed a significant enrichment for genes involved in gene-environment interactions and protein degradation.

Conclusions

We show that wild C. elegans populations are characterized by gene-environment signatures and have unlocked a wealth of genotype-phenotype relations for the first time. Studying natural isolates provides a treasure trove in addition to current research in C. elegans which covers only a diminutive part of the myriad of genotype-phenotype relations which are present in the wild.

Keywords

Gene-environment interactions, Genotype-phenotype relations, Wild *C. elegans* strains, Transcriptomic diversity

Background

The nematode *Caenorhabditis elegans* is a widely used model species in contemporary biological research covering different disciplines like developmental biology, genetics and evolutionary biology. Many investigations have been of paramount importance for understanding biology of mammals. But almost all studies in *C. elegans*, including forward (knocking out genes by mutation) and reverse (knocking down genes using RNAi) genetic

screens, have been conducted in a few strains of which the canonical strain Bristol N2 is the most thoroughly studied. This severely constrains the detection and functional analysis of allelic variants which play a key role in controlling many complex traits. It is therefore essential to explore the full potential of the natural genetic variation and evolutionary context of the genotype-phenotype map in wild *C. elegans* populations. Moreover, the widely used strains, like N2 and CB4856 have often been kept under controlled laboratory conditions for decades. Also handling, storage and maintenance of worms is standardized. Such artificial regimes very likely create multiple bottle-necks which can affect the genotype-phenotype relationship. For instance, a genetic analysis of wild *C. elegans* strains revealed that the N2 alleles of *npr-1* and *glb-5* originated most likely as an adaptation to laboratory conditions [1]. Genotype-phenotype relations are studied in many species including model organisms, like *Arabidopsis* [2,3], *Drosophila* [4], and *C. elegans* [5,6]. For full appreciation and functional characterization of genes and the genotype-phenotype relations, it is essential to consider the natural context of the species, including analysis of natural isolates and its interaction with natural challenges. Previous studies on *C. elegans* have investigated the response to a wide range of different environmental factors like exposure to different bacteria [7-9], pH [9], osmotic pressure [9,10], chemicals [11,12], temperature [9,12-16], and many others. As yet however, these responses have not been tested in natural populations.

C. elegans is an androdioecious species with a low outcrossing rate leading to homozygous strains in natural isolates [17]. These strains can therefore be treated as haplotypes. Here we studied variation in genotype-phenotype relations for a total of 48 strains of which 41 were freshly isolated from two different sites in France, 20 strains from a woodland area in Santeuil (S) from rotting hogweed stems and 21 from an orchard in Orsay (O) from rotting apples. As an out-group three strains freshly isolated from sites in The Netherlands and two strains previously isolated from France were used. Lastly the genotypically most diverse laboratory kept strains CB4856 and the canonical strain Bristol N2 were added (Additional file 1 (worksheet A)) [9,18-22]. These last two strains were used in many studies to uncover genotype-phenotype relations both by comparing strains or using some type of quantitative trait locus approach [1,18,23-30].

In this study we provide insight into the genotype-phenotype relations in natural *C. elegans* populations by analysing genomic and transcriptomic variation. We found that local genetic diversity reflects site specific signatures of environmental sensing, protein regulation and the immune defence system. Our results indicate that exploring natural isolates in *C. elegans* leads to finding key components of genotype-phenotype relations as opposed to studies which are limited to the canonical strain Bristol N2.

Results and discussion

Local *C. elegans* populations are genotypically separable

Previous investigations have studied population genetics and genomic diversity in *C. elegans* focusing on global [31-35] or local [17,36] populations. Cutter showed that there is a lack of geographic distribution of *C. elegans* genome sequences [31] and Andersen *et al*. reported that chromosome-scale selective sweeps have acted to reduce genetic variation and have shaped the global *C. elegans* population structure in recent history [37]. Barrière and Félix concluded that local diversity is high [17]. In all of these papers, diversity in *C. elegans* was measured as genetic diversity. Until now, very few papers have been published concerning

phenotypic variation in wild isolates (by this we mean isolates which have not been maintained in the lab for a long time) and they have studied only a small number of isolates [38-40]. To our knowledge, no studies have been reported on genotype-phenotype relations in wild populations.

The wild strains from Santeuil and Orsay and the out-group strains (mentioned in the Background paragraph) were genotyped based on the hybridisation of genomic *C. elegans* DNA to microarrays (see Methods section for details). This resulted in the identification of 6368 polymorphic genes with an absolute ratio of 0.5 with the mean hybridization intensity (Figure 1A). Most of these (~66%) were found in only one to three strains (Figure 1B) showing that between-strain variation is more abundant than between-site variation. Major hotspots of polymorphic genes were found on chromosome *II* and *V* and minor on the other chromosomes (Figure 1C; Additional files 1 (worksheet B), 2, 3, and 4). The hotspots of polymorphic genes co-locate with the c-type lectin, nhr, math domain and chemoreceptor gene clusters [41]. In addition, microsatellite loci were used to determine the population structure [36] of the Orsay and Santeuil strains (Additional file 1 (worksheet C), Additional file 5 (panel A)).

Figure 1 Polymorphic genes in wild isolates of *C. elegans*. A. Number of polymorphic genes per strain. Bars above zero indicate the polymorphic genes with a positive ratio (higher intensity than the mean), bars below zero indicate the polymorphic genes with a negative ratio. The wide part of the bars refers to the genes with a ratio of > 1 or < -1. The narrow part of the bars indicates genes with a ratio of > 0.5 or < -0.5. Strains from Orsay in orange, from Santeuil in green and the out-group in purple. **B.** Frequency of occurrence of polymorphic genes. Number of strains indicated at the left, percentage of total shown at the right. For example polymorphic genes only found in one strain make up 33.2% of the total number of polymorphic genes. **C.** Distribution of the polymorphic genes in 48 different *C. elegans* strains. Genomic position on the x-axis, number of polymorphic genes on the y-axis. Chromosomes are shown in different panels, chromosome names are given on the left in the grey boxes. The black bars indicate the total number of polymorphic genes per 100 kb. The lines show the number of genes with a ratio < -0.5 for three different groups of *C. elegans* strains, whereby orange indicates strains from Orsay, green those from Santeuil and purple the out-group strains. For example, the large number of polymorphic genes at the beginning (left arm) of chromosome *II* is mostly caused by the many genes that are very polymorphic or absent in the out-group lines (purple line is high).

The two isolation sites were genotypically separable. This was shown by analysing presence of gene polymorphisms using principal component analysis (PCA) (Figure 2A), a distance matrix visualised by an unrooted neighbour joining (NJ) tree (Figure 2B), and a minimum spanning network of the microsatellite data (Additional file 5 (panel A)). The minimum spanning network, PCA and NJ showed a clear distinction between the Santeuil strains and the Orsay strains, with for both isolation sites one large genetic group and several smaller genetic groups. In the PCA the first two principal components capture ~75% of the variation in DNA hybridisations. N2 is in the far right top corner indicating its genetic difference from all other strains (Figure 2A). Moreover the NJ tree shows that the Santeuil strains (groups S1, S2 and S3) and Orsay strains are different from N2 and CB4856. In both the PCA and NJ analyses the Orsay group (group O) is genetically less diverse than the Santeuil group. Within the main Santeuil group (group S1: all Santeuil strains except JU1924, JU1925, JU1926, JU1934, JU1935, JU1936) diversity is slightly larger. Furthermore in the NJ tree two small genotypic groups separate from the main Orsay and Santeuil groups (group S2: JU1924,

JU1925, JU1926; and group S3: JU1511, JU1934, JU1935, JU1936). The strains within these separate groups are from the Santeuil site except for JU1511 which is from the Orsay site. The strains from Santeuil in group S2 was isolated from one hogweed stem as were those from S3 from their own one hogweed stem (Additional file 1 (worksheet A)). Other strains were found on different hogweed stems. Of these, strains isolated from one hogweed stem also group close together but were not found to form their own separate genotypic groups. The Dutch strains were isolated from two isolation sites, WN2001 from one and WN2002 and WN2003 from the other. The latter two strains group together in the NJ tree. One of the French out-group strains, JU396, groups with the Santeuil strains and the other, JU314, is different from the rest of the strains. N2 and CB4856 are as diverse from the other out-group members as is the whole out-group compared to Orsay or Santeuil. By comparing the four genetic groups (O, S1, S2 and S3) to the out-group, the genes that were polymorphic were identified (Additional file 5 (panel B)). Group S1 appeared to be the most divergent from the out-group with 3181 genes that differed significantly (FDR = 0.05).

Figure 2 The Orsay and Santeuil populations are genotypically separable based on genomic DNA analysis with microarrays. A. Principle component analysis (PCA) plot: PC1 on the x-axis separates the main Santeuil group (green) from all other strains, and PC2 on the y-axis separates the Santeuil and out-group strains (purple) from the Orsay strains (yellow). **B.** NJ tree made with the same genetic data as was used for the PCA plot. Orsay strains are shown in orange, Santeuil strains in green, and the out-group strains in purple**. C.** NJ tree based on the RNA hybridisation data. The average log2 ratio per probe (with the mean) per genotype was used. Only probes with a maximum absolute ratio of >0.5 were used. The 192 genes which' expression level were solely influenced by DNA polymorphisms were not included. Orsay strains are shown in orange, Santeuil strains in green, and the out-group strains in purple**.**

The genetic separation between the Orsay and Santeuil populations was prevalent on all chromosomes (Figure 3). From the scale of the axes it can be seen that most chromosomes contribute to the separation between the two isolation sites and the out-group, except for chromosome *II*. On this chromosome the Santeuil and Orsay lines form one group that separates only from the out-group. Chromosome *V* contributes the most to the variation between Orsay and Santeuil, most likely because of the generally higher level of variation among the strains (also see Additional file 4). Of the ~2500 genes different between S1, S2 and S3, ~42% is located on chromosome *V*. Of all the genes on chromosome *V*, ~20% is polymorphic between S1, S2 and S3. This is a significant enrichment ($p < 1e\text{-}76$) when compared to the other chromosomes of which 8 to 10% of the genes are polymorphic. Chromosomes *I* and *X* are under-represented with polymorphic genes (both ~8% and $p < 1e\text{-}12$).

Figure 3 Neighbour Joining tree of a distance matrix of the genetic polymorphisms for each chromosome. Upper panels: chromosomes *I* to *III*; lower panels: chromosomes *IV* to *X*. Orsay strains are shown in orange, Santeuil strains in green, and the out-group strains in purple.

The detected genotypic diversity between sites is in line with genotyping results from 31 markers using amplified fragment length polymorphisms (AFLP), as well as with microsatellite results from 2 loci in other local *C. elegans* populations [17]. Selective sequencing using restriction-site–associated DNA (RAD) tags however did not reveal significant local diversity, possibly due to the limited number of different genotypes per

location [37]. The four genetic groups (O, S1, S2 and S3) identified by analysis of the ratio intensities were used as input to search for all the genes that are linked to isolation site. This allowed us to identify polymorphic genes by minor hybridization differences (abs ratio < 0.5), beyond those 6386 found by major hybridization differences (abs ratio > 0.5). In this way we identified 3742 genes (FDR = 0.05) that were linked to isolation site (Additional file 6 (panel A)). Of these 3742 genes, 2403 were already identified as highly polymorphic in the initial analysis on major hybridization differences and an additional 1339 genes associated with isolation were found by only minor hybridization differences. Of the genes with major hybridization differences ~62%, could not be linked to an isolation site, again showing that between-strain variation is more abundant than between-site variation.

Genes linked to isolation site are enriched for Gene Classes fbox, math, bath, btb, C-type lectin (clec), serpentine chemoreceptor, and nuclear hormone receptor

To investigate whether specific types of genes were overrepresented in the group of genes that could be linked to isolation site (group of 3742) or in the polymorphic genes not linked to isolation site (group of 3965), enrichment analyses were performed (Table 1 and Additional file 1 (worksheets D and E)). The linked and not-linked groups were analysed using three types of annotations, Gene Class, Anatomy Terms and GO terms, to investigate whether certain types of genes were enriched. For the Gene Classes we found that fbox-, math-, bath-, btb-, C-type lectin (clec)-, serpentine chemoreceptor- and nuclear hormone receptor genes were enriched in the group of genes linked to isolation site (Table 1, Additional files 4 (worksheets F, G and H)) and 6 (panel B). Enrichment analyses were also performed for the genes that were significantly linked to the genetic groups O, S1, S2 and S3, revealing the same gene classes as above (Additional file 1 (worksheet I)). The enrichment analysis of Anatomy Terms or GO terms did not identify a clear pattern linked to isolation site or genetic group (Additional file 1 (worksheets J and K)).

Table 1 Enrichment (based on DNA-array data) of gene classes

			Isolation site		Polymorphic by ratio	
Gene group	**Gene Class**	**Group Size**	**Overlap**	**Significance**	**Overlap**	**Significance**
Serpentine receptors						
Superfamily Str	srh	289	118	**20.3**	40	0.0
Superfamily Str	str	219	68	**6.4**	39	0.2
Superfamily Str	sri	76	28	**4.7**	12	0.1
Superfamily Str	srj	45	19	**4.5**	7	0.2
Solo	srz	104	43	**8.4**	27	1.4
Solo	srw	145	53	**7.8**	20	0.0
Solo	srbc	84	32	**5.6**	18	0.6
Solo	srr	10	5	**2.5**	3	1.0
Superfamily Sra	srab	27	11	**2.9**	5	0.4
Superfamily Srg	srt	72	29	**5.8**	11	0.1
C-type lectins	clec	260	72	**4.7**	44	0.1
F-box	fbxa	220	115	**31.5**	38	0.1
F-box	fbxb	113	43	**7.1**	24	0.6
F-box	fbxc	49	14	1.7	8	0.2
Math,bath,btb	math	50	41	**23.6**	4	0.0
Math,bath,btb	bath	44	26	**9.9**	9	0.5
Math,bath,btb	btb	21	8	2.2	4	0.4
Nuclear hormone receptor	nhr	282	71	**3.2**	57	0.5
Pharyngeal gland toxin-related	phat	6	5	**4.5**	0	NA
Scramblase (phospholipid scramblase)	scrm	8	6	**4.4**	0	NA

Enrichment (based on DNA-array data) of Gene Classes in group of genes with variation linked to isolation site (3742 genes) and in group of polymorphic genes that are not linked to isolation site (3965 genes). Significance in –log10(p), only gene classes with group sizes ≥ 6, overlap ≥ 3 genes, and significance ≥ 2.5 are shown in bold.

Local *C. elegans* populations are separable on the basis of their transcriptomes

Next, the influence of natural genetic variation on gene expression was studied, by measuring the transcript levels of all genes of all strains, corrected for differential hybridization. A NJ tree was constructed based on the RNA hybridisation data (Figure 2C). This tree shows that the genetic groups O, S1, S2 and S3, as well as the out-group, are also separable based on gene expression level. Again, CB4856 and N2 differ from most of the other natural strains. Isolation site and genetic group influenced the variation in RNA levels of 6930 and 7996 genes respectively (Additional file 7). Most of these genes (77% and 78%) were not influenced by DNA polymorphisms (cause of variation in 2330 genes) and genotype (affecting 773 or 1336 genes depending on genetic group or isolation site being incorporated in the model).

Expressed genes linked to isolation site are enriched for Gene Classes clec, fbxa, bath, and nuclear hormone receptor

Enrichment analyses were performed for the genes which RNA levels were influenced by isolation site or genetic group (Additional file 1 (worksheets L, M and N)). The Gene Classes clec, fbxa, bath, and nuclear hormone receptors were significantly enriched, thus yielding similar results to our DNA-level enrichment analyses. In addition, several nematode specific peptide families were also enriched.

Together, these results show that at the genomic level variation between local populations is enriched for gene classes fbox, math, bath, btb, c-type lectin, serpentine chemoreceptor and nuclear hormone receptor, many of which are involved in gene-environment interactions [42-45]. Interestingly, we found that Gene Classes clec, fbxa, bath, and nuclear hormone receptor were also enriched with variation linked to isolation site on the transcriptional level, while the strains that originated from different sites were cultured under the same conditions. Many of these groups of genes have been shown to be differentially expressed after pathogen exposure and thus could be involved in the immune response. For instance, C-type lectin domain containing proteins (CTLD proteins, Gene Class clec) have been repeatedly proposed to contribute to nematode immunity (reviewed in [46]). Their immune function is supported by their specific upregulation in infected *C. elegans* [43,44,47-50] and also by reduced immune phenotypes after RNAi knockdown of *clec-70*, *clec-17*, *clec-60* or *clec-86* [7,51]. Furthermore, F-box proteins (Gene Class fbxa) are part of the protein degradation pathway [52]. In this pathway, substrates for degradation are ubiquinated to be recognisable by the 26S proteasome. Taken together, we show that local genetic diversity reflects site specific signatures of immune response and protein degradation pathways in *C. elegans*. We also show that, besides genotypes, transcript profiles can very well be used to distinguish between local *C. elegans* populations and may point to functional importance of the identified genes or gene classes in different environments [53].

Polymorphic genes are enriched for genes involved in gene-environment interactions

Polymorphic genes between the populations were compared to a set of genes extracted from 19 different studies on gene expression in *C. elegans* exposed to biotic and abiotic factors (Additional file 1 (worksheet S). In the wild, *C. elegans* is exposed to many different bacteria. In studies concerning the effect on gene expression of various bacteria, such as *L. rhamnosus* [54], *M. nematophilum* [7], *D. coniospora* [55], *S. marcescens* [8], *X. nematophila* [8], and *P. aeruginosa* [56], c-type lectins are always found to be differentially expressed, as are in most cases the F-box protein genes. Receptors that are used to sense the environment, such as nuclear hormone receptors and serpentine receptors also are frequently differentially expressed when *C. elegans* is exposed to different bacteria. In response to abiotic factors such as temperature [12], osmotic stress [10] or ions [57,58], c-type lectins and F-box protein genes are also always differentially expressed. C-type lectins, F-box protein genes and receptor genes are furthermore differentially expressed in the presence of various other substances that can be encountered by wild *C. elegans* strains: tryptophan [59], β-naphthoflavone [60], H_2S [61], fluoranthene [62], hormones [63], sediment [64], humic substances [65] and pesticides [12,66,67]. The other gene classes (bath, math and btb) that are of importance for the variation between the locations at which the wild *C. elegans* strains were isolated were found to be differentially expressed in several of the abovementioned environmental studies as well. Altogether, the differential expression of genes in

environmental studies indicates that the genes that are important for the variation between local populations of *C. elegans* are indeed of significance for interactions with the environment.

Local populations are separable for some fitness traits

The next question was if the genetic polymorphisms among strains could influence fitness trait variation. *C. elegans* strains varied significantly in all traits, except population size on *E. coli* OP50 (Table 2). As all tests were done under standardized lab conditions and the variation between strains can be attributed to the genotype this shows that most phenotypic variation has a genetic basis. A genetic determinant has been found for some of these traits [5,13]. A significant influence of the genetic groups was found on the population size on *Bacillus thuringiensis* NRRL B-18BT247 and on the length/width ratio (Additional file 1 (worksheet O)). We additionally reconstructed an NJ tree using phenotypic trait variation, however phenotypic variation did not separate the two isolation sites or any of the four genetic groups. Nevertheless, some phenotypes were specific to an isolation site or to certain genetic groups. Even though the two strains with the largest length/width ratio are from Santeuil, most worms from Santeuil are significantly shorter, have a significantly smaller length/width ratio and so are more stout than worms from Orsay (Table 2 and Additional file 8). Also, the generation time of worms from Santeuil was significantly shorter (Table 2). More details can be found in Additional file 1 (worksheet O).

Table 2 Analysis of phenotypic variation among strains

Phenotype	N (populations per genotype)	Mean (sd)	Anova (Strain)	Mean (sd) [Orsay]	Mean (sd) [Santeuil]	*T*-test
Pop. size *E. coli*	6	3136 (695)	0.3602	3278 (670)	2988 (707)	0.19
Pop. size DSM[a]	6	3385 (750)	**4.83e-4**	3369 (702)	3402 (815)	0.89
Pop. size BT247[b]	6	44 (39)	**< 2.2e-16**	53 (49)	34 (21)	0.12[c]
Development time (days)	2-5	1.79 (0.07)	**4.16e-4**	1.79 (0.07)	1.79 (0.07)	0.74
Generation time (days)	2-5	1.98 (0.08)	**3.13e-6**	1.98 (0.077)	1.97 (0.77)	**0.019**
Embryogenesis (hours)	2-5	4.60(0.87)	**NA**	4.35 (0.70)	4.85 (0.97)	**0.031**
Length (μm)	2-6	1089 (58)	**4.83e-5**	1107 (33)	1070 (72)	**0.023**
Width (μm)	2-6	43.96 (2.63)	**1.33e-9**	44.45 (1.99)	43.46 (3.13)	0.35
Volume (nl)	2-6	1.67 (0.24)	**1.12e-6**	1.73 (0.19)	1.61 (0.28)	0.99
L/W ratio	2-6	24.79 (0.82)	**1.19e-6**	24.94 (0.79)	24.64 (0.83)	**1.40e-5**

Analysis of phenotypic variation among strains (anova) and among isolation sites. Significant values are printed in boldface. N describes the number of replicate populations of worms (each population more than 100 worms) which were tested per genotype, the number of N per isolation-site therefore is 20 times higher.

a) DSM = Non-nematocidal *Bacillus thuringiensis* DSM-350.

b) BT247 = Nematocidal *Bacillus thuringiensis* NRRL B-18247.

c) When all observations per genotype are used instead of the mean per genotype the populations from the different isolations sites are significantly different ($p < 0.0014$).

Local populations are separable with regard to food preference

We then investigated if the wild strains differed among each other in their food preference behaviour for naturally co-existing bacteria and *E. coli*. Under the used laboratory conditions, the worms preferred *E. coli* OP50 over all other bacteria, followed by *Erwinia rhapontici*, *Sphingobacterium* sp., *Rhodococcus erythropolis* and *Lactococcus lactis* (Figure 4, Additional file 1 (worksheet P)). Worms from Santeuil preferred *E. rhapontici* (isolated from Santeuil) equally as *E. coli*. Worms from Orsay preferred *E. coli* over *E. rhapontici*. This suggests that Santeuil worms could have a slight preference for the bacterium species they are likely familiar with. For an overview of the average preference of all strains, see Additional file 9B. The complete dataset can be found in Additional file 1 (worksheet Q) (Wormcount and Choice Index) and Additional file 1 (worksheet R) (Significances). The most significant differences found between the Orsay and Santeuil strains were for the bacteria combinations *E. coli* OP50/*E. rhapontici*, *E. coli* OP50/*R. erythropolis* and *E. rhapontici*/*Sphingobacterium* sp. (Figure 4).

Figure 4 Preference of the strains (rectangles on the left) from the different origins for the different bacteria (right ellipses). Bacteria in green were isolated in Santeuil, bacteria in orange were isolated in Orsay, bacterium in purple is standard lab food OP50.

The canonical strains Bristol N2 and CB4856 are genetic outliers

We finally wondered in how far the canonical wild types Bristol N2 and CB4856 relate to the recently isolated natural strains. Both genotypic and transcriptomic analyses identified them to differ clearly from the standing genetic variation of the wild isolated strains. This also applies when we compare N2 and CB4856 to some of the other out-group strains. It is assumed that many alleles in CB4856 and N2 are lab-derived because both N2 and CB4856 went through multiple phenotypic and genetic bottlenecks over the past decades of lab maintenance [1,68]. Together with our results this shows that the genotype-phenotypes relations in N2, but also CB4856, are likely to be obscured by a number of lab derived alleles of large effect. This might impede the detection and functional analysis of many other genes and alleles which, by themselves have small effects on phenotypes, but together they might have a strong effect.

Conclusions

We have measured a large variety of phenotypes, including the transcriptome, for multiple wild genotypes in *C. elegans* collected from different locally separated sites, under constant and the same laboratory conditions. The wild genotypes could be classified according to their site based on genotypic and transcriptome analyses. The differences were also reflected in several fitness traits, however due to the limited number of populations sampled we have not been able to associate fitness traits to the different sites. Yet, our data provide the basis for uncovering site-specific genotypic and phenotypic signature . This future work should aim to provide insight considering genetic drift or adaptation as the major attribute shaping *C. elegans* local evolution. Most likely both processes play a role depending on the gene or genetic element in question. However for some gene classes, like the chemoreceptors it is tempting to think they are polymorphic due to adaptations to specific habitats. In summary, we have unlocked a wealth of genotype-phenotype relations indicating that the canonical wild

type is a genetic outlier and that its genotype-phenotype characteristics represent a diminutive part of the myriad of interactions present in the wild.

Methods

Nematode- and bacterial strains

The main set of strains of *C. elegans* comprised forty-one new strains that were isolated by Marie-Anne Félix from two different locations in France (Orsay and Santeuil). As an out-group, three new strains isolated in the Netherlands, two strains previously isolated in France and the most diverse canonical strains N2 (Bristol) and CB4856 (Hawaii) were used [16,18-21,23-29,67]. See Additional file 1 (worksheet A) for details. All strains were routinely maintained on NGM with *Escherichia coli* OP50 as a food source [69]. *E. coli* OP50 was used in all experiments, except for the population growth experiment in which *Bacillus thuringiensis* NRRL B-18247 and *B. thuringiensis* DSM-350 were used next to it [70]. In the food preference experiment, besides OP50, *Erwinia rhapontici* and *Rhodococcus erythropolis* (both isolated from and unique for Santeuil), and *Lactococcus lactis* and *Sphingobacterium* sp. (both isolated from and unique for Orsay) that were isolated and identified by M-A. Félix and Buck Samuel, were used.

Genomic DNA analysis: worm culturing, DNA isolation, DNA-microarrays and statistical analysis

Gene expression microarrays were used to co-hybridize N2 vs wild type DNA allowing for analysing population differences based on gene polymorphisms. Fresh populations of mixed stages were cultured for 96 hours at 20°C before sampling. The microarrays used were *C. elegans* (V2) Gene Expression Microarray 4X44K slides, manufactured by Agilent Technologies (Santa Clara, CA, USA). All procedures were performed as recommended by Agilent in the 'Oligonucleotide Array-Based CGH for Genomic DNA Analysis; Enzymatic Labeling for Blood, Cells or Tissues (with a High Throughput option)' –protocol, version 6.3. Genomic DNA isolation was performed with the NucleoSpin Tissue Kit from Machery-Nagel (Düren, Germany). For processing the data of the DNA microarrays the "Limma" package for the "R" environment was used [71]. Background correction was done by using the Substract method. Loess within-array normalisation and Scale between-array normalisation were used to process the raw intensity values [72]. Genotypes were compared by calculating, per spot, the ratio of the intensities of each strain with the mean intensity over all strains. The genes with a ratio of >0.5 or < -0.5 were considered polymorphic. Principal component analysis (PCA) was done using the polymorphic genes from all strains. The un-rooted neighbour joining (NJ) tree was made from a distance matrix made from the ratios of the polymorphic genes with the "R" package "Phangorn" [73]. Linear models were used to calculate the significance of the variation in DNA hybridisation intensities linked to isolation-sites and the identified genetic groups. The model used to determine linkage to isolation-site was "DNA hybridization intensity ~ isolation site (out, Orsay, Santeuil) + error". For linkage to Santeuil we took –log10(p) of 2.3 as threshold, for linkage to Orsay we took –log10(p) of 2.7 as threshold. For linkage to genetic group we used the model "DNA hybridization intensity ~ genetic group (out, O, S1, S2, S3) + error". The thresholds used were –log10(p) 2.5, 2.3, 3.2, 3.3 for O, S1, S2, and S3 respectively. For the number of genes per genetic group and overlapping genes see Additional file 5B. The significance thresholds, adjusted for multiple testing, were determined by permutation, for which the same model was used with

the spot intensities randomly distributed over the genotypes (a p value which gave a ratio of false positives/true positives of < 0.05 was used).

mRNA analysis: culturing, isolation, RNA-microarrays and statistical analysis

For the mRNA microarrays, any few males were discarded and only hermaphrodites grown on *E. coli* OP50 were used. Two independent replicates of each strain (synchronised late L4 larvae) were analysed. For mRNA isolation, an RNEasy Micro Kit from Qiagen (Hilden, Germany) was used, following the 'Purification of Total RNA from Animal and Human Tissues' protocol provided with the kit, with modified lysing procedure, see Additional file 10. The microarrays used were *C. elegans* (V2) Gene Expression Microarray 4X44K slides, manufactured by Agilent. All procedures were performed as recommended by Agilent. For processing the data of the RNA microarrays the "Limma" package for the "R" environment was used. No background correction was performed as recommended [71]. For within-array normalisation the Loess method was used and for between-array normalisation the Quantile method was used. Expression variation was determined by linear models. The variation in intensities was explained by batch, DNA hybridisation, genetic group and genotype (See also the part about statistics in the Genomic DNA analysis paragraph). Significance thresholds, adjusted for multiple testing were determined by permutations of all spots on the array. In the permutations the RNA hybridisation intensities were randomly distributed over the genotypes and batches (the p value which gave a ratio of false positives/true positives < 0.05 was used).

Enrichment analysis

All enrichment analyses were done using a hyper-geometric test. The number of genes selected by a criterion in this paper (e.g. linked to a genetic group) were compared to the genes with a specific annotation (e.g. c-type lectin). The chance that a number of genes will be overlapping depends on the total group size, the number of genes selected and the number of genes with a specific annotation. This, together with the number of overlapping genes can be used in a hyper-geometric test. Annotation groups were considered enriched when the overlap was >3 genes and the significance $-\log10P > 2.5$.

Polymorphic genes between populations were compared to a set of differentially expressed genes extracted from a diverse set of gene-environment interaction studies in *C. elegans*. All enrichment analyses were done using a hyper-geometric test.

Phenotypic assays

Developmental time and generation time

L1 juveniles fed with *E. coli* OP50 were incubated at 24°C and observed at regular time intervals. Developmental time is the period between worm inoculation and the moment at which the first worms with open vulva were observed. Generation time is the period between inoculation and the first appearance of eggs.

Length and width

Analysis of length and width of young gravid worms was performed with a RapidVue particle analyzer (Beckman Coulter). 2000 worms per strain were measured.

Population growth

10 single L4 worms were placed onto a bacterial lawn and cultured at 20°C. After 96 h, the number of worms on the plate was counted.

Food preference assay

To test the food preference of the worms, 5 μl drops of two different bacteria were placed on NGM in each well of a 12 wells plate (Additional file 1, worksheet P). A drop with juveniles until L2 was added to each well and the plate was incubated overnight at 20°C. After this, the worms on each bacterium were counted and the Choice Index was calculated [74].

Statistics

We used an anova to calculate the influence of strain/genotype on the phenotypic variation, where we regressed the individual measurements over the strains/genotypes. We used a 2-sided *t*-test assuming unequal variance to determine if phenotypes were significantly different between isolation sites. An anova was used to determine if phenotypes were significantly different between genetic groups.

Microsatellite analysis

Population genetic differentiation was assessed using six microsatellite loci (Additional file 1 (worksheet C)), which we previously identified to be highly variable for natural and experimental *C. elegans* populations [36]. See [36] and Additional file 10 for details.

Data storage

Micro-array data (both RNA and DNA) can be found at www.WormQTL.org [23].

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

RJMV carried out the RNA- and DNA arrays and drafted the manuscript. LBS designed and carried out all statistical analyses except for the microsatellite study and drafted the manuscript. CJHH carried out the food preference assays. RC carried out the developmental time, generation time, length and width measurements. WC carried out the population growth measurements and generated the microsatellite data. WY helped with the enrichment analysis of the studies on gene expression. MGS helped designing the statistical analyses and helped to draft the manuscript. HS carried out the statistical analysis of the microsatellite study. HS, BPB and JEK conceived of the study and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Marie-Anne Félix (Paris, France) for providing the nematode strains from France; Buck Samuel (Boston, US) for providing the wild bacteria; Joost Riksen (Wageningen, The Netherlands) and Janina Brakel (Kiel, Germany) for technical assistance; and K. Joeri van der Velde and Morris A. Swertz (Groningen Bioinformatics Centre, Groningen, The Netherlands) for assistance with WormQTL. We also would like to thank WormBase for being a versatile and important resource for *C. elegans* research. RJMV was funded by the NWO-ALW (project 855.01.151), RC and BPB were funded by ESF-EEFG (09-EuroEEFG-FP-002/G.0998.10N). JEK and LBS were funded by the ERASysbio-plus ZonMW project GRAPPLE (project nr. 90201066). MGS was supported by Graduate School Production Ecology & Resource Conservation. WC, WY and HS were funded by NEMADAPT (DFG grant SCHU 1415/11-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. McGrath PT, Rockman MV, Zimmer M, Jang H, Macosko EZ, Kruglyak L, Bargmann CI: **Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors.** *Neuron* 2009, **61:**692–699.

2. Snoek LB, Terpstra IR, Dekter R, Van den Ackerveken G, Peeters AJ: **Genetical Genomics Reveals Large Scale Genotype-By-Environment Interactions in *Arabidopsis thaliana*.** *Front Gen* 2012, **3:**317.

3. Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC: **Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci.** *Proc Natl Acad Sci U S A* 2007, **104:**1708–1713.

4. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu DH, Casillas S, Han Y, Magwire MM, Cridland JM, *et al*: **The *Drosophila melanogaster* genetic reference panel.** *Nature* 2012, **482:**173–178.

5. Gutteling EW, Doroszuk A, Riksen JA, Prokop Z, Reszka J, Kammenga JE: **Environmental influence on the genetic correlations between life-history traits in *Caenorhabditis elegans*.** *Heredity (Edinb)* 2007, **98:**206–213.

6. Palopoli MF, Rockman MV, TinMaung A, Ramsay C, Curwen S, Aduna A, Laurita J, Kruglyak L: **Molecular basis of the copulatory plug polymorphism in *Caenorhabditis elegans*.** *Nature* 2008, **454:**1019–U1066.

7. O'Rourke D, Baban D, Demidova M, Mott R, Hodgkin J: **Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*.** *Genome Res* 2006, **16:**1005–1016.

8. Sinha A, Rae R, Iatsenko I, Sommer RJ: **System wide analysis of the evolution of innate immunity in the nematode model species *Caenorhabditis elegans* and *Pristionchus pacificus*.** *PloS one* 2012, **7:**e44255.

9. Grishkevich V, Ben-Elazar S, Hashimshony T, Hunter CP, Yanai I: **A genomic bias for genotype-environment interactions in *C. elegans*.** *Mol Syst Biol* 2012, **8:**587.

10. Rohlfing AK, Miteva Y, Hannenhalli S, Lamitina T: **Genetic and physiological activation of osmosensitive gene expression mimics transcriptional signatures of pathogen infection in *C. elegans*.** *PloS one* 2010, **5:**e9010.

11. Jonker MJ, Sweijen RAJC, Kammenga JE: **Toxicity of simple mixtures to the nematode caenorhabditis elegans in relation to soil sorption.** *Environ Toxicol Chem* 2004, **23:**480–488.

12. Viñuela A, Snoek LB, Riksen JA, Kammenga JE: **Gene expression modifications by temperature-toxicants interactions in *Caenorhabditis elegans*.** *PloS one* 2011, **6:**e24676.

13. Gutteling EW, Riksen JA, Bakker J, Kammenga JE: **Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*.** *Heredity (Edinb)* 2007, **98:**28–37.

14. Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JA, Hazendonk E, Prins P, Plasterk RH, Jansen RC, *et al*: **Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*.** *PLoS genetics* 2006, **2:**e222.

15. Harvey SC, Viney ME: **Thermal variation reveals natural variation between isolates of *Caenorhabditis elegans*.** *J Exp Zool B Mol Dev Evol* 2007, **308:**409–416.

16. Gaertner BE, Parmenter MD, Rockman MV, 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:**1533–1542.

17. Barriere A, Félix MA: **High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations.** *Curr Biol* 2005, **15:**1176–1184.

18. Viñuela A, Snoek LB, Riksen JA, Kammenga JE: **Aging uncouples heritability and expression-QTL in *Caenorhabditis elegans*.** *G3* 2012, **2:**597–605.

19. Capra EJ, Skrovanek SM, Kruglyak L: **Comparative developmental expression profiling of two *C. elegans* isolates.** *PloS one* 2008, **3:**e4055.

20. Seidel HS, Ailion M, Li J, van Oudenaarden A, Rockman MV, Kruglyak L: **A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*.** *PLoS biology* 2011, **9:**e1001115.

21. Rockman MV, Kruglyak L: **Recombinational landscape and population genomics of *Caenorhabditis elegans*.** *PLoS Genet* 2009, **5:**e1000419.

22. Kammenga JE, Doroszuk A, Riksen JAG, Hazendonk E, Spiridon L, Petrescu AJ, Tijsterman M, Plasterk RHA, Bakker J: **A *Caenorhabditis elegans* wild type defies the temperature-size rule owing to a single nucleotide polymorphism in *tra-3*.** *Plos Genetics* 2007, **3:**358–366.

23. Snoek LB, Van der Velde KJ, Arends D, Li Y, Beyer A, Elvin M, Fisher J, Hajnal A, Hengartner MO, Poulin GB, *et al*: **WormQTL–public archive and analysis web portal for natural variation data in *Caenorhabditis spp*.** *Nucleic Acids Res* 2013, **41:**D738–D743.

24. Green JW, Snoek LB, Kammenga JE, Harvey SC: **Genetic mapping of variation in dauer larvae development in growing populations of *Caenorhabditis elegans*.** *Heredity* 2013. doi:10.1038/hdy.2013.50.

25. Rodriguez M, Snoek LB, Riksen JA, Bevers RP, Kammenga JE: **Genetic variation for stress-response hormesis in *C. elegans* lifespan.** *Exp Gerontol* 2012, **47:**581–587.

26. Elvin M, Snoek LB, Frejno M, Klemstein U, Kammenga JE, Poulin GB: **A fitness assay for comparing RNAi effects across multiple *C. elegans* genotypes.** *BMC genomics* 2011, **12:**510.

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

28. Doroszuk A, Snoek LB, Fradin E, Riksen J, Kammenga J: **A genome-wide library of CB4856/N2 introgression lines of *Caenorhabditis elegans*.** *Nucleic Acids Res* 2009, **37:**e110.

29. Rockman MV, Skrovanek SS, Kruglyak L: **Selection at linked sites shapes heritable phenotypic variation in *C. elegans*.** *Science* 2010, **330:**372–376.

30. Kammenga JE, Phillips PC, De Bono M, Doroszuk A: **Beyond induced mutants: using worms to study natural variation in genetic pathways.** *Trends Gen* 2008, **24:**178–185.

31. Cutter AD: **Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*.** *Genetics* 2006, **172:**171–184.

32. Koch R, van Luenen HG, van der Horst M, Thijssen KL, Plasterk RH: **Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*.** *Genome Res* 2000, **10:**1690–1696.

33. Maydan JS, Lorch A, Edgley ML, Flibotte S, Moerman DG: **Copy number variation in the genomes of twelve natural isolates of *Caenorhabditis elegans*.** *BMC Genomics* 2010, **11:**62.

34. Schulenburg H, Muller S: **Natural variation in the response of *Caenorhabditis elegans* towards *Bacillus thuringiensis*.** *Parasitology* 2004, **128:**433–443.

35. Sivasundar A, Hey J: **Population genetics of *Caenorhabditis elegans*: the paradox of low polymorphism in a widespread species.** *Genetics* 2003, **163:**147–157.

36. Haber M, Schungel M, Putz A, Muller S, Hasert B, Schulenburg H: **Evolutionary history of *Caenorhabditis elegans* inferred from microsatellites: evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding.** *Mol Biol Evol* 2005, **22:**160–173.

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

38. Schulenburg H, Ewbank JJ: **Diversity and specificity in the interaction between *Caenorhabditis elegans* and the pathogen *Serratia marcescens*.** *BMC Evol Biol* 2004, **4:**49.

39. Salomon MP, Ostrow D, Phillips N, Blanton D, Bour W, Keller TE, Levy L, Sylvestre T, Upadhyay A, Baer CF: **Comparing mutational and standing genetic variability for fitness and size in *Caenorhabditis briggsae* and *C. elegans*.** *Genetics* 2009, **183:**685–692. 681SI-619SI.

40. Braendle C, Baer CF, Félix MA: **Bias and evolution of the mutationally accessible phenotypic space in a developmental system.** *PLoS Genet* 2010, **6:**e1000877.

41. Thomas JH: **Analysis of homologous gene clusters in *Caenorhabditis elegans* reveals striking regional cluster domains.** *Genetics* 2006, **172:**127–143.

42. Thomas JH, Robertson HM: **The *Caenorhabditis* chemoreceptor gene families.** *BMC Biol* 2008, **6:**42.

43. Jeba Mercy G, Pandian SK, Balamurugan K: **Changes in *Caenorhabditis elegans* life span and selective innate immune genes during *Staphylococcus aureus* infection.** *Folia microbiologica* 2011, **56:**373–380.

44. Sivamaruthi BS, Ganguli A, Kumar M, Bhaviya S, Pandian SK, Balamurugan K: ***Caenorhabditis elegans* as a model for studying *Cronobacter sakazakii* ATCC BAA-894 pathogenesis.** *J Basic Microbiol* 2011, **51:**540–549.

45. Taubert S, Ward JD, Yamamoto KR: **Nuclear hormone receptors in nematodes: evolution and function.** *Mol Cell Endocrinol* 2011, **334:**49–55.

46. Schulenburg H, Hoeppner MP, Weiner J 3rd, Bornberg-Bauer E: **Specificity of the innate immune system and diversity of C-type lectin domain (CTLD) proteins in the nematode *Caenorhabditis elegans*.** *Immunobiol* 2008, **213:**237–250.

47. Jeba Mercy G, Balamurugan K: **Effects of subsequent infections in *Caenorhabditis elegans* with *Staphylococcus aureus* and *Proteus mirabilis*.** *Microbiol Immunol* 2012, **56:**825–835.

48. Kesika P, Karutha Pandian S, Balamurugan K: **Analysis of *Shigella flexneri*-mediated infections in model organism *Caenorhabditis elegans*.** *Scand J Infect Dis* 2011, **43:**286–295.

49. Bogaerts A, Temmerman L, Boerjan B, Husson SJ, Schoofs L, Verleyen P: **A differential proteomics study of *Caenorhabditis elegans* infected with *Aeromonas hydrophila*.** *Developmental and comparative immunology* 2010, **34:**690–698.

50. Sahu SN, Lewis J, Patel I, Bozdag S, Lee JH, LeClerc JE, Cinar HN: **Genomic analysis of immune response against *Vibrio cholerae* hemolysin in *Caenorhabditis elegans*.** *PloS one* 2012, **7:**e38200.

51. Irazoqui JE, Troemel ER, Feinbaum RL, Luhachack LG, Cezairliyan BO, Ausubel FM: **Distinct pathogenesis and host responses during infection of *C. elegans* by *P. aeruginosa* and *S. aureus*.** *PLoS pathogens* 2010, **6:**e1000982.

52. Ho MS, Tsai PI, Chien CT: **F-box proteins: the key to protein degradation.** *J Biomed Sci* 2006, **13:**181–191.

53. Coolon JD, Jones KL, Todd TC, Carr BC, Herman MA: ***Caenorhabditis elegans* genomic response to soil bacteria predicts environment-specific genetic effects on life history traits.** *PLoS Genet* 2009, **5:**e1000503.

54. Grompone G, Martorell P, Llopis S, Gonzalez N, Genoves S, Mulet AP, Fernandez-Calero T, Tiscornia I, Bollati-Fogolin M, Chambaud I, *et al*: **Anti-inflammatory *Lactobacillus rhamnosus* CNCM I-3690 strain protects against oxidative stress and increases lifespan in *Caenorhabditis elegans*.** *PloS one* 2012, **7:**e52493.

55. Pujol N, Zugasti O, Wong D, Couillault C, Kurz CL, Schulenburg H, Ewbank JJ: **Anti-fungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides.** *PLoS pathogens* 2008, **4:**e1000105.

56. Troemel ER, Chu SW, Reinke V, Lee SS, Ausubel FM, Kim DH: **p38 MAPK regulates expression of immune response genes and contributes to longevity in *C. elegans*.** *PLoS Genet* 2006, **2:**e183.

57. Cui Y, McBride SJ, Boyd WA, Alper S, Freedman JH: **Toxicogenomic analysis of *Caenorhabditis elegans* reveals novel genes and pathways involved in the resistance to cadmium toxicity.** *Genome biology* 2007, **8:**R122.

58. Hattori A, Mizuno T, Akamatsu M, Hisamoto N, Matsumoto K: **The *Caenorhabditis elegans* JNK signaling pathway activates expression of stress response genes by derepressing the Fos/HDAC repressor complex.** *PLoS Genet* 2013, **9:**e1003315.

59. Gracida X, Eckmann CR: **Fertility and germline stem cell maintenance under different diets requires nhr-114/HNF4 in *C. elegans*.** *Curr Biol* 2013, **23:**607–613.

60. Menzel R, Bogaert T, Achazi R: **A systematic gene expression screen of *Caenorhabditis elegans* cytochrome P450 genes reveals CYP35 as strongly xenobiotic inducible.** *Arch Biochem Biophys* 2001, **395:**158–168.

61. Miller DL, Budde MW, Roth MB: **HIF-1 and SKN-1 coordinate the transcriptional response to hydrogen sulfide in *Caenorhabditis elegans*.** *PloS one* 2011, **6:**e25476.

62. Swain S, Wren JF, Sturzenbaum SR, Kille P, Morgan AJ, Jager T, Jonker MJ, Hankard PK, Svendsen C, Owen J, *et al*: **Linking toxicant physiological mode of action with induced gene expression changes in *Caenorhabditis elegans*.** *BMC Syst Biol* 2010, **4:**32.

63. Custodia N, Won SJ, Novillo A, Wieland M, Li C, Callard IP: ***Caenorhabditis elegans*** **as an environmental monitor using DNA microarray analysis.** *Ann N Y Acad Sci* 2001, **948:**32–42.

64. Menzel R, Swain SC, Hoess S, Claus E, Menzel S, Steinberg CE, Reifferscheid G, Sturzenbaum SR: **Gene expression profiling to characterize sediment toxicity–a pilot study using *Caenorhabditis elegans* whole genome microarrays.** *BMC Genomics* 2009, **10:**160.

65. Menzel R, Menzel S, Swain SC, Pietsch K, Tiedt S, Witczak J, Sturzenbaum SR, Steinberg CE: **The nematode *Caenorhabditis elegans*, stress and aging: identifying the complex interplay of genetic pathways following the treatment with humic substances.** *Front Gen* 2012, **3:**50.

66. Lewis JA, Szilagyi M, Gehman E, Dennis WE, Jackson DA: **Distinct patterns of gene and protein expression elicited by organophosphorus pesticides in *Caenorhabditis elegans*.** *BMC Genomics* 2009, **10:**202.

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

68. Weber KP, De S, Kozarewa I, Turner DJ, Babu MM, de Bono M: **Whole genome sequencing highlights genetic changes associated with laboratory domestication of *C. elegans*.** *PloS one* 2010, **5:**e13922.

69. Sulston JE, Hodgkin J: **Methods**. In *The nematode Caenorhabditis elegans.* Edited by Wood WB. New York: Cold Spring Harbor Laboratory; 1988:587–606.

70. Boehnisch C, Wong D, Habig M, Isermann K, Michiels NK, Roeder T, May RC, Schulenburg H: **Protist-type lysozymes of the nematode *Caenorhabditis elegans* contribute to resistance against pathogenic *Bacillus thuringiensis*.** *PloS one* 2011, **6:**e24619.

71. Smyth GK: **Limma: linear models for microarray data**. In *Bioinformatics and computational biology solutions using R and Bioconductor.* Edited by Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S. New york: Springer Science + Business Media; 2005:397–420.

72. Smyth GK, Speed T: **Normalization of cDNA microarray data.** *Methods* 2003, **31:**265–273.

73. Schliep KP: **phangorn: phylogenetic analysis in R.** *Bioinformatics* 2011, **27:**592–593.

74. Abada EA, Sung H, Dwivedi M, Park BJ, Lee SK, Ahnn J: ***C. elegans*** **behavior of preference choice on bacterial food.** *Mole Cells* 2009, **28:**209–213.

Additional files

Additional_file_1 as XLSX

Additional file 1 Various supplementary datasets.

Additional_file_2 as PDF

Additional file 2 etailed overview of DNA hybridization differences. Chromosome number on top of each page. In orange, wild-isolates from Orsay, in green, wild-isolates from Santeuil and in purple the genotypes from the out-group. On the y-axis, the log2 ratio of the individual lines with the mean of all lines per microarray probe is shown by the dots. The moving average (9 probes) is shown by the lines. Threshold for moving-average are shown by the horizontal red lines. Probe positions are indicated by the triangles on the x-axis, gene names of genes with a ratio outside the thresholds are shown in the figure. The lines are drawn to the start of the gene on the genome.

Additional_file_3 as PDF

Additional file 3 Genome wide overview of DNA hybridization differences per chromosome. See Additional file 2 for legend.

Additional_file_4 as PDF

Additional file 4 Number of polymorphic genes per chromosome. The wide part of the bars shows the number of genes with an absolute ratio > 1, the narrow part shows the number of genes with an absolute ratio of >0.5. Total number of genes per chromosome (with percentage of polymorphic genes (ratio >0.5) in parentheses) I:2969 (28%); II:3588 (32%); III:2680 (28%); IV: 3435 (30%); V: 5400 (35%); X: 2809 (24%).

Additional_file_5 as PDF

Additional file 5 A Minimum spanning network constructed using microsatellite data. O and red colour refers to Orsay, S and blue to Santeuil. Circle size is proportional to the number of strains with a particular genotype. The solid lines show the main relationships among genotypes, the dotted lines alternative connections. Line length correlates with the inferred number of evolutionary differences. The minimum spanning network was reconstructed with the program Arlequin. **B** Venn-diagram of genes for which DNA hybridisation intensity per genetic group is significantly different from the out-group. Orsay: 1933 genes in total; Santeuil 1: 3181 genes in total; Santeuil 2: 737 genes in total; Santeuil 3: 567 genes in total. Group S1 appeared to be the most divergent from the out-group with 3181 genes that differed significantly. A large part of these genes (803) is also shared with the Orsay group. The amount of significantly different genes that are the same for S3 and S2 or O (4 in both cases) and for S2 and O (1 gene) is remarkably low. S2 also shares a low amount of the same significantly different genes with O (19).

Additional_file_6 as PDF

Additional file 6 A. Schematic overview of the groups of polymorphic genes based on DNA hybridization data. **B**. Percentage of all genes detected using hybridisation of genomic *C. elegans* DNA on microarrays that were linked or not linked to the isolation sites. Together the gene classes Serpentine receptors, F-box, math, bath, btb, clec, and nhr composed almost 25% of the polymorphic genes significantly linked to isolation site. These same gene classes made up less than 10% out of the genes that could not be linked to isolation site.

Additional_file_7 as PDF

Additional file 7 Venn diagram of the genes that showed expression differences due to DNA polymorphisms and genotype in combination with either genetic group (O, S1, S2 and S3) or isolation site (Santeuil and Orsay). Total amounts of genes: genotype left diagram 773; genotype right diagram 1336; DNA 2230; genetic group 7996; isolation site 6930.

Additional_file_8 as PDF

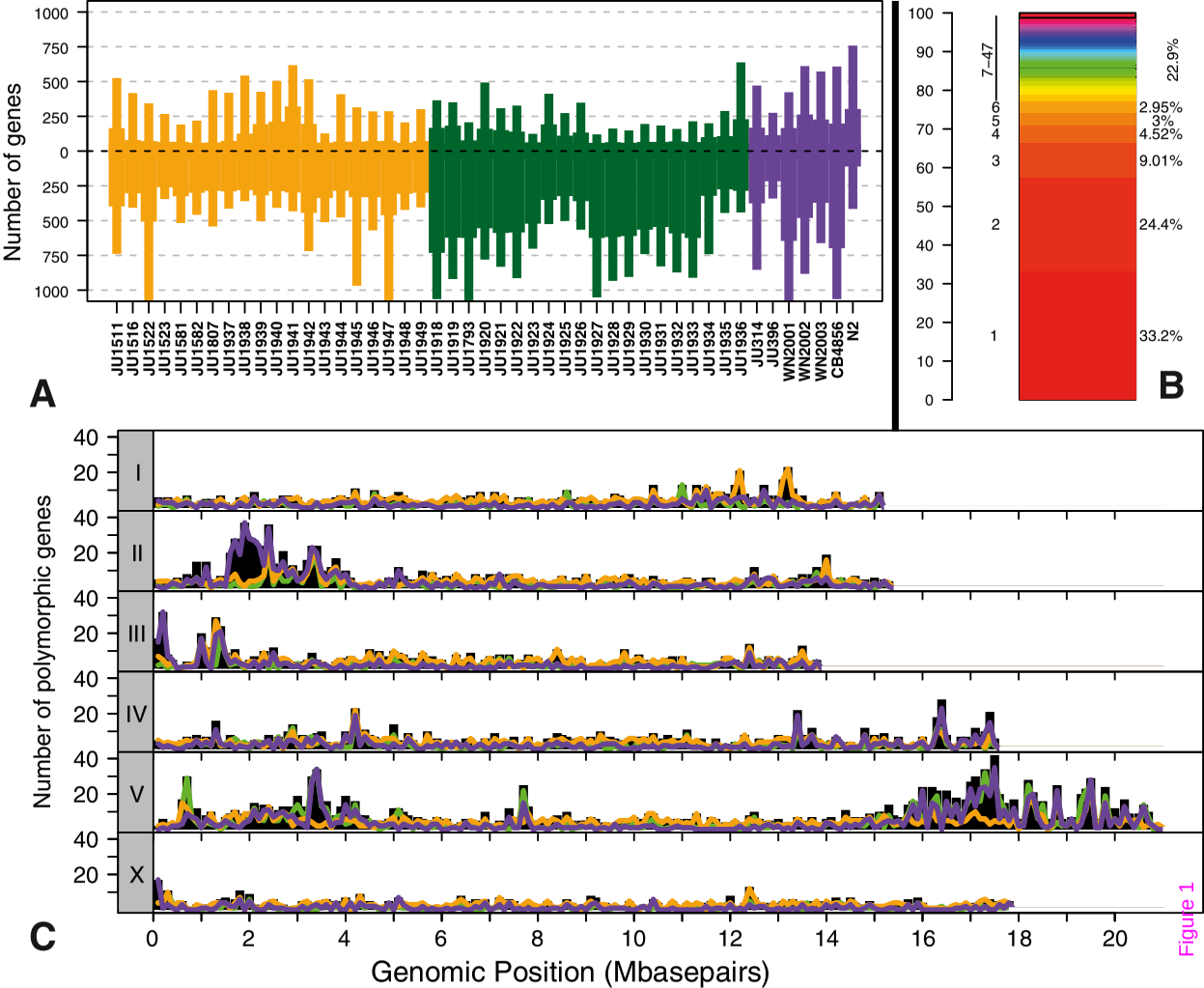
Additional file 8 Phenotypes of the wild-isolates. Strains from Orsay shown in orange, strains from Santeuil shown in green, out group strains in purple. The right panel shows the statistics, mean and standard deviation (sd) as well as the p value of the *t*-test on the phenotypic difference between the Orsay and Santeuil groups with and without outliers removed. When applicable an anova on strain was done and the p-value is shown. Lastly the heritability (H2) was calculated. Labels on the y-axis refer to the phenotypes described in Table 2.

Additional_file_9 as PDF

Additional file 9 Food preference assay **A**. Set-up of the food preference assay and the calculation of the Choice Index (CI). A = bacterium A; B = bacterium B. **B**. Schematic overview of the results of the food preference assay. Bacteria in green were isolated in Santeuil, bacteria in orange were isolated in Orsay, bacterium in purple is standard lab food OP50. All numbers are percentages and are the average of all strains in the experiment. The percentages near the bacteria indicate the fraction of worms that prefer the particular bacterium when tested together with the bacterium at the opposite end of the line. The percentages in or near the yellow circles indicate the fraction of worms that did not choose between the two bacteria. An example: when offering a choice between *Sphingobacterium* and *L. lactis*, on average 85.7% of all worms of all strains prefers *Sphingobacterium*, 5.5% prefers *L. lactis*, and 8.8% did not make a choice between these bacteria.

Additional_file_10 as DOCX

Additional file 10 Detailed description of Methods.


Number of genes
1000
750
500
250
0
250
500
750
1000
JU1511
JU1516
JU1522
JU1523
JU1581
JU1582
JU1807
JU1937
JU1938
JU1939
JU1940
JU1941
JU1942
JU1943
JU1944
JU1945
JU1946
JU1947
JU1948
JU1949
JU1918
JU1919
JU1793
JU1920
JU1921
JU1922
JU1923
JU1924
JU1925
JU1926
JU1927
JU1928
JU1929
JU1930
JU1931
JU1932
JU1933
JU1934
JU1935
JU1936
JU314
JU396
WN2001
WN2002
WN2003
CB4856
N2
A
100
90
80
70
60
50
40
30
20
10
0
7–47
6
5
4
3
2
1
22.9%
2.95%
3%
4.52%
9.01%
24.4%
33.2%
B
Number of polymorphic genes
40
20
I
II
III
IV
V
X
C
0
2
4
6
8
10
12
14
16
18
20
Genomic Position (Mbasepairs)
Figure 1

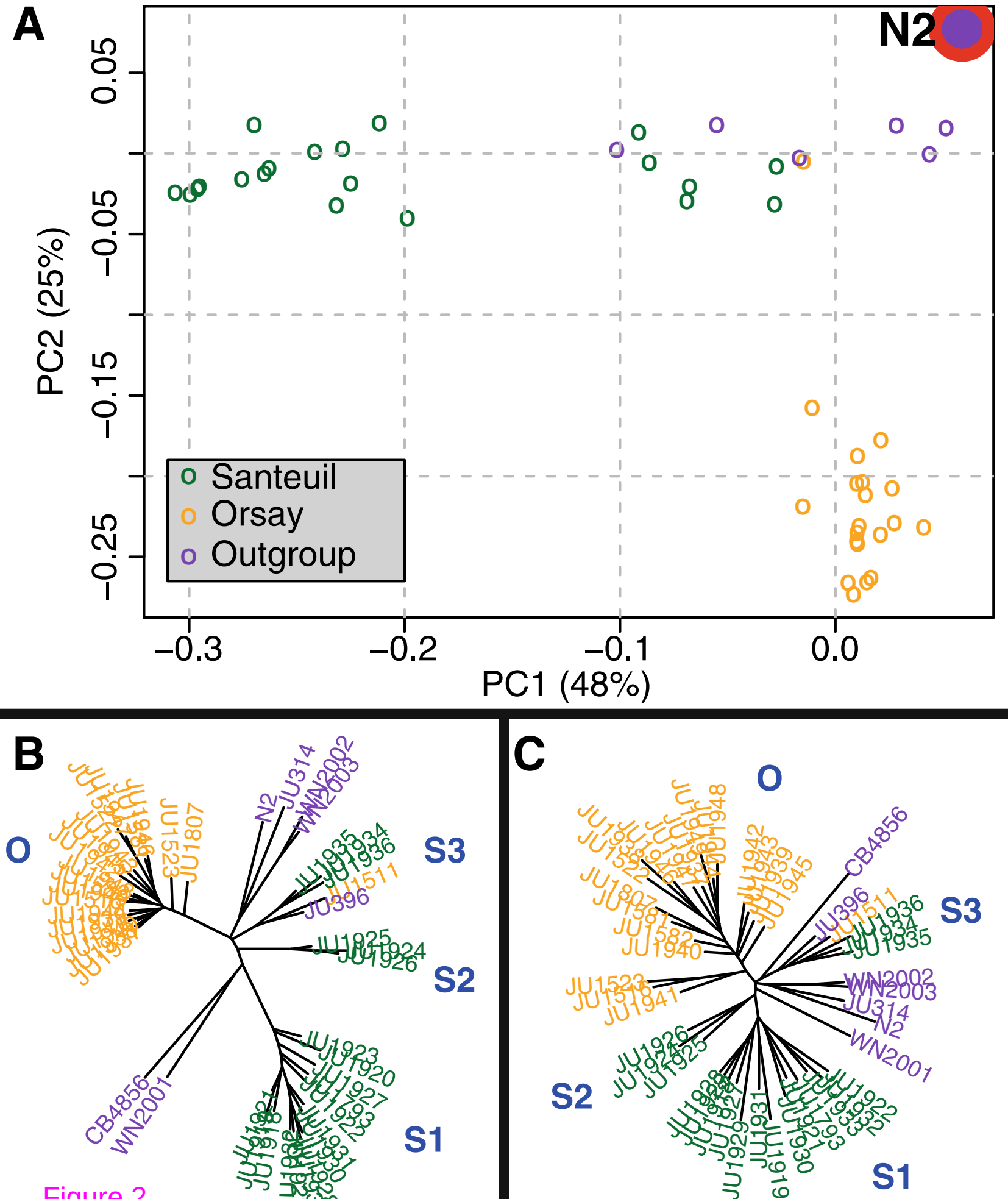

A
N2
0.05
−0.05
−0.15
−0.25
PC2 (25%)
Santeuil
Orsay
Outgroup
−0.3
−0.2
−0.1
0.0
PC1 (48%)
B
O
S3
S2
S1
N2
JU314
WN2002
WN2003
JU1807
JU1523
JU1935
JU1934
JU1936
JU1511
JU396
JU1925
JU1924
JU1926
JU1923
JU1920
CB4856
WN2001
C
O
S3
S2
S1
CB4856
JU396
JU1511
JU1936
JU1934
JU1935
WN2002
WN2003
JU314
N2
WN2001
JU1523
JU1941
JU1940
JU1926
JU1925
JU1931
JU1919

Figure 2

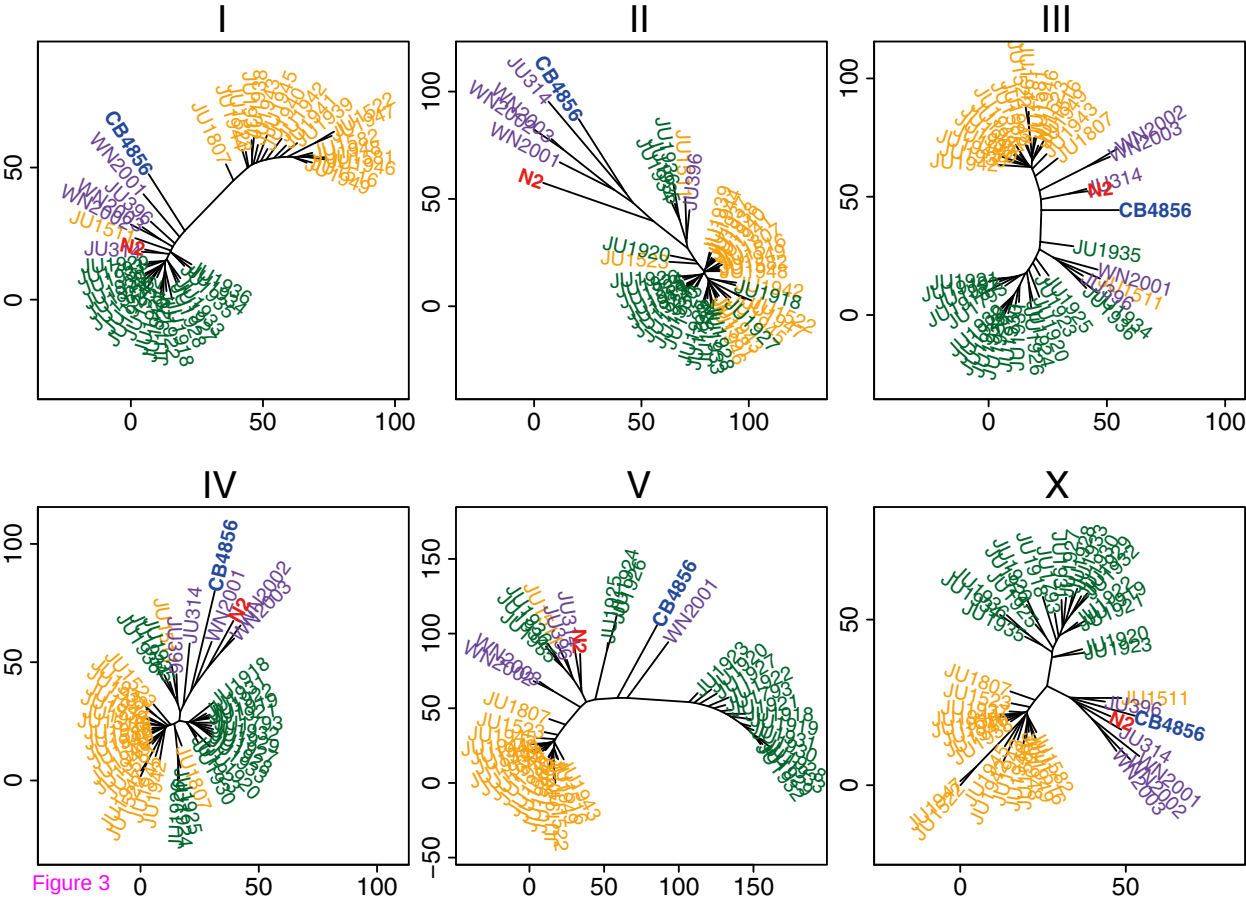

I
II
III
IV
V
X
CB4856
N2
JU1807
JU1511
JU396
JU314
JU1920
JU1935
JU1943
JU1918
WN2001
WN2002
WN2003

Figure 3

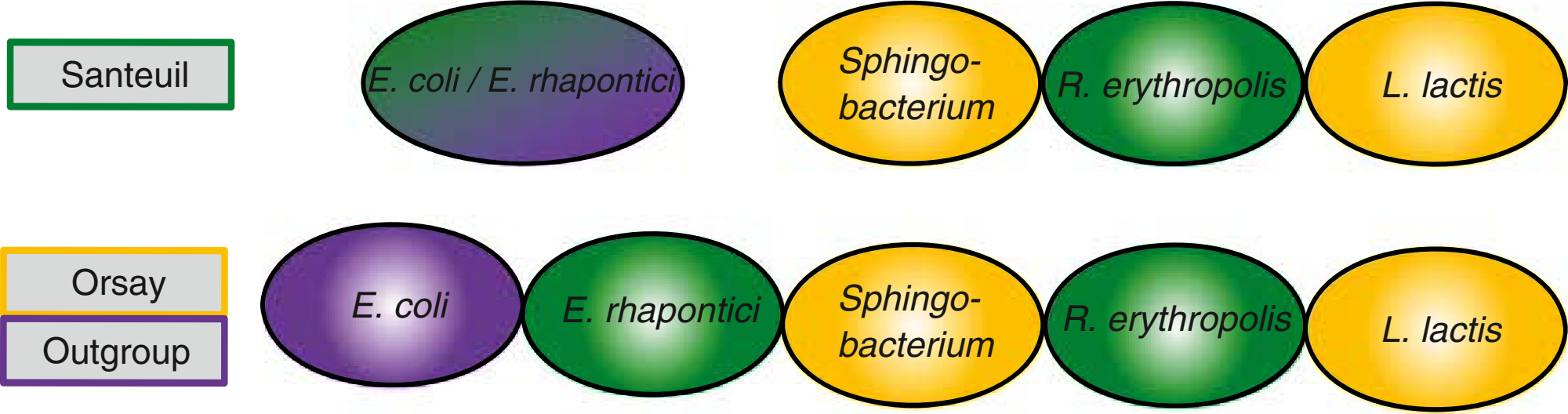

Santeuil
E. coli / E. rhapontici
Sphingo-bacterium
R. erythropolis
L. lactis
Orsay
Outgroup
E. coli
E. rhapontici
Sphingo-bacterium
R. erythropolis
L. lactis
Likes
Dislikes

Figure 4

Additional files provided with this submission:

Additional file 1: 1478686947102592_add1.xlsx, 1016K
http://www.biomedcentral.com/imedia/4277813021058438/supp1.xlsx
Additional file 2: 1478686947102592_add2.pdf, 9854K
http://www.biomedcentral.com/imedia/2024617280105843/supp2.pdf
Additional file 3: 1478686947102592_add3.pdf, 7683K
http://www.biomedcentral.com/imedia/1511465965105843/supp3.pdf
Additional file 4: 1478686947102592_add4.pdf, 5K
http://www.biomedcentral.com/imedia/1165723233105843/supp4.pdf
Additional file 5: 1478686947102592_add5.pdf, 2115K
http://www.biomedcentral.com/imedia/1707107924105843/supp5.pdf
Additional file 6: 1478686947102592_add6.pdf, 520K
http://www.biomedcentral.com/imedia/5709712631058438/supp6.pdf
Additional file 7: 1478686947102592_add7.pdf, 80K
http://www.biomedcentral.com/imedia/8319385561058438/supp7.pdf
Additional file 8: 1478686947102592_add8.pdf, 47K
http://www.biomedcentral.com/imedia/1566486138105843/supp8.pdf
Additional file 9: 1478686947102592_add9.pdf, 1574K
http://www.biomedcentral.com/imedia/2372462391058438/supp9.pdf
Additional file 10: 1478686947102592_add10.docx, 38K
http://www.biomedcentral.com/imedia/2172402810584385/supp10.docx