Natural genetic variation is the raw material of evolution and influences disease development and progression. An important question is how this genetic variation translates into variation in protein abundance. To analyze the effects of the genetic background on gene and protein expression in the nematode *Caenorhabditis elegans*, we quantitatively compared the two genetically highly divergent wild-type strains N2 and CB4856. Gene expression was analyzed by microarray assays, and proteins were quantified using stable isotope labeling by amino acids in cell culture. Among all transcribed genes, we found 1,532 genes to be differentially transcribed between the two wild types. Of the total 3,238 quantified proteins, 129 proteins were significantly differentially expressed between N2 and CB4856. The differentially expressed proteins were enriched for genes that function in insulin-signaling and stress-response pathways, underlining strong divergence of these pathways in nematodes. The protein abundance of the two wild-type strains correlates more strongly than protein abundance *versus* transcript abundance within each wild type. Our findings indicate that in *C. elegans* only a fraction of the changes in protein abundance can be explained by the changes in mRNA abundance. These findings corroborate with the observations made across species. *Molecular & Cellular Proteomics* 15: 10.1074/mcp.M115.052548, 1670–1680, 2016.

Natural genetic variation in gene expression shapes the diversity in phenotypic traits and is the raw material for evolutionary processes (1). Variation in gene expression can be very extensive across individuals with different genotypes. The additive effects (narrow-sense heritability) of independent loci on gene expression variation can reach 35% in humans (2). The broad-sense heritable variation in gene expression has been estimated to be up to 70% in the nematode *Caenorhabditis elegans* (3, 4) and up to 80% in yeast (5). This high heritability and the ability to construct genetically segregating populations facilitate mapping of gene expression regulation and subsequent detection of expression quantitative trait loci (eQTL) (5–11). eQTLs are genomic regions containing a polymorphism associated with variation in transcript abundance between genotypes (12). eQTL analysis provides insight into the underlying genetic architecture of complex traits and is valuable for the identification of pathways and gene networks (10, 13–15). A key question is whether gene expression variation is translated into variation at the proteome level and whether it affects functionally relevant proteins. Genetic model species provide an ideal platform to explore the relationship between gene expression variation and variation at the proteome level due to their tractability. Although it is well established that there is a correlation between transcript and protein abundances, the relationship between natural variation in gene expression and variation in protein abundance is less well understood. The proteome provides information required to understand the functioning of cells (16), and combined with natural genetic variation, it allows for mapping of protein expression regulators (17). In yeast, gene expression variation in 354 genes between genetically different strains was found to lead to variation in protein abundance for a limited set of proteins reflecting mainly transcription-independent mechanisms (18). In mice, the natural variation of transcript abundance and protein levels among inbred strains correlates slightly (0.27) (19). Over half of the identified eQTLs in yeast contributed to changes in protein levels of regulated genes, but several protein-QTLs did not correlate with their cognate transcript levels (20).

The abbreviations used are: eQTL, expression quantitative trait loci; FDR, false discovery rate; GO, gene ontology; iBAQ, intensity-based absolute quantification; L4, larval stage 4; SILAC, stable isotope labeling by amino acids in cell culture; ACN, acetonitrile.
C. elegans is an ideal model organism to analyze the effect of the genetic background on its transcriptome (8, 9, 21–23) and proteome because comprehensive proteome catalogues have been generated (24, 25). The quantitative proteome comparison in C. elegans and Drosophila melanogaster showed that the interspecies’ protein abundance correlation was higher than the intraspecies’ correlation between protein and mRNA abundance. This suggests that protein levels are under evolutionary selection acting post-transcriptionally (26).

In a recent interspecies comparison of the two nematodes C. elegans and Caenorhabditis briggsae, the protein and transcript changes were found to be conserved throughout development (27).

To complete these comparative interspecies expression analyses with an intraspecies analysis and to investigate the effect of natural genetic variation from the transcriptome to the proteome, the genetically highly divergent C. elegans wild-type strains, N2 (isolated in Bristol, UK) and CB4856 (from Hawaii), were compared quantitatively. Transcriptome data were acquired by microarray analysis, and for quantitative proteome analysis, the SILAC labeling method was used, which was recently established for C. elegans (28, 29). We analyzed animals at the developmental larval stage 4 (L4), because at this stage the largest differences in gene expression levels were detected between N2 and CB4856 (3, 4, 30–32). The genomes of these two wild-type strains, as well as polymorphisms between them, have been characterized extensively (33–40). These genetic polymorphisms often lead to gene expression differences, which are playing a role in a whole range of phenotypic differences between N2 and CB4856, such as life history traits (4, 41–44), behavior (40, 45–49), and life span (41, 50). Here, we asked whether natural genetic variation causes variation in protein levels between N2 and CB4856 and whether it is related to variation in transcript abundance.

Using SILAC on three independent biological replicates, we quantified a total of 3,238 proteins, 2,485 proteins were identified in at least two biological replicates. We found 129 proteins to be significantly differentially expressed between N2 and CB4856 with at least a 1.3-fold change in abundance and a p value < 0.000457. They were enriched for genes that function in insulin-signaling and stress-response pathways, underlying strong divergence of these pathways in nematodes. The protein abundance of the two wild-type strains correlated more strongly than the protein abundance versus transcript abundance within each wild type.

**EXPERIMENTAL PROCEDURES**

**C. elegans Strains and Culture Conditions**—The two C. elegans wild-type strains N2 (Bristol) and CB4856 (Hawaii) were grown at 20 °C on 9-cm NGM agar plates without peptone (3 g/liter NaCl, 20 g/liter bacto-agar, 5 mg/liter cholesterol, 25 mM K2PO4, 1 mM MgSO4, 1 mM CaCl2) and with a lawn of *Escherichia coli* OP50 strain bacteria.

**Transcriptome Comparison**—For RNA isolation we used a Maxwell® 16 AS2000 instrument with a Maxwell® 16 LEV simplyRNA tissue kit (both from Promega Corp., Madison, WI). The mRNA isolation was preceded by a modified lysis step. In short, 200 µl of homogenization buffer, 200 µl of lysis buffer, and 10 µl of a 20 mg/ml stock solution of proteinase K were added to each sample. The samples were then incubated for 10 min at 65 °C, 1000 rpm in a Thermomixer (Eppendorf, Hamburg, Germany). After cooling on ice for 1 min, the samples were pipetted into the cartridges, and the protocol as recommended by Promega was continued. After mRNA isolation, the “Two-color Microarray-based Gene Expression Analysis, Low Input Quick Amp Labeling” protocol, version 6.0, was followed, starting from step 5 (4, 30, 31).

**Protein Isolation, SDS-PAGE, Protein Digestion**—To extract proteins, worm samples were homogenized with glass beads (G1277 acid-washed beads, diameter of 212–300 µm, Sigma-Aldrich) in freshly prepared cell lysis buffer (8 µM urea, 5% 1 µM Tris-HCl, pH 8.3) in a ratio of 1:1.2 (worms/beads/buffer) at 4 °C for 30 s at 5 m/s four times (FastPrep®-24, MP Biomedicals). The lysates were centrifuged three times at 20,000 × g for 10 min at room temperature to remove debris. Protein concentration in the supernatant was determined using the Bradford reagent (Sigma-Aldrich), and 1 µg of each protein sample was checked for equimolarity by SDS-PAGE.
The L4 animal lysates of N2 and CB4856 were combined in a 1:1 ratio. Protein disulfide bridges were reduced with 5 mM dithiothreitol (DTT) at 60 °C for 30 min and alkylated with 15 mM iodoacetamide in the dark at 37 °C for 1 h. 600 μg of proteins were digested with trypsin (modified sequencing grade porcine, Promega) in a ratio of 1:50 w/w overnight at 37 °C. Samples were kept frozen at −20 °C until further processing.

**HPLC Fractionation of Peptides and ZipTip C18 Sample Clean Up**—Peptide samples were dried in a SpeedVac concentrator and resuspended in solvent A (5% ACN, 20 mM K3HPO4, pH 11) to a final urea concentration below 2 M. The pH was adjusted to 11 with 20% KOH, and the peptides were loaded on a 150 × 4.6-mm YMC Triart C18 column filled with 5 μm of silica beads (YMCE Europe GmbH) at a flow rate of 1 ml/min using an Agilent 1100 liquid chromatography system (Agilent Technologies Inc.). Peptides were separated by a linear gradient from 2 to 50% of solvent B (50% ACN, 20 mM K3HPO4, pH 11) within 50 min. In total, 47 fractions were collected and pooled into 10 fractions based on the intensity measured by a UV-light detector at 214 nm. The reproducibility of the separation was controlled by running twice a self-mixed standard peptide mixture (adenocorticotropin(1–16), angiotensinogen(1–14), bradykinin(2–9), leu-enkephalin, substance P, and vasopressin). Pooled fractions were vacuum-centrifuged until complete dryness and resuspended in 50 μl of 5% ACN, 0.1% TFA. Salts were removed using ZipTip C18 pipette tips (Millipore Corp.), and the peptides were eluted with 15 μl of 60% ACN, 0.1% TFA. Samples were dried and re-dissolved in 12 μl of 3% ACN, 0.1% formic acid to a final peptide concentration of 0.4 μg/μl.

**LC-MS/MS Analysis**—LC-MS/MS was performed on a reversed-phase nano-LC system (Eksigent) at pH 3. Peptides were separated on a self-packed reverse-phase column (75 μm × 10 cm) packed with C18 beads (Magic C18, AC, 3 μm, 200 Å, Bischoff GmbH, Leonberg, Germany) at a flow rate of 200 nl/min. The column was equilibrated with 95% solvent A (0.1% formic acid in water) and 5% solvent B (0.1% formic acid in ACN). Peptides were eluted using the following gradient: 0–1 min; 5–9% B, 1–56 min; 9–40% B, 56–60 min; 40–50% B and 60–64 min; 50–95% B. Peptides were analyzed on an LTQ Orbitrap XL mass spectrometer in the data-dependent acquisition mode. High accuracy mass spectra were acquired in the mass range of 300–1,800 m/z. Ion mass tolerance was set to 20 ppm, and the fragment ion mass tolerance was set to ±0.6 Da. The peptide FDR was set to 1%; protein FDR was set to 5%. In total, 40 raw data files were generated for every biological replicate and combined for database searching with the match between runs set to 2 min. Proteins with all shared peptides were combined into one protein group. The “Majority Protein IDs” column was used for protein assignments. The annotated MS/MS spectra with the lowest posterior error probability for single peptide identifications were extracted from 2,485 protein dataset using R package protViz (S3) and are shown in supplemental Figs. S3–S8. The lists of corresponding single peptides are included in supplemental Tables S6–S8 for all three replicates.

**Correction of Arginine-to-Proline Conversion**—To estimate the effect of the arginine-to-proline conversion, all raw files were analyzed with the Progenesis QI for Proteomics software (Nonlinear Dynamics). The MS1 m/z feature maps filtered for the charge states from 2 to 7 were generated using the automatic method for peak picking with a sensitivity value of 5. The number of fragment ion counts was limited to 200, and deisotoping and charge deconvolution were applied. The generated mgf files were searched against the C. elegans database with Mascot for the identification of heavy proline peaks. The precursor mass tolerance was set to 10 ppm, and the fragment ion mass tolerance was set to 0.6 Da. Only fully tryptic termini with two missed cleavages were considered. The isotopic labeling of arginine (Arg-10) and lysine (Lys-8), heavy proline (Pro-6), heavy glutamate (Glu-6) labels and methionine oxidation were set as variable modifications. Cysteine carbamidomethylation was set as fixed modification. The analysis yielded 301 peptides with heavy proline and heavy arginine and/or heavy lysine. Based on this, the contribution of the heavy proline signal to the heavy arginine-lysine signal was estimated to be 20% on average for peptides carrying one proline (supplemental Fig. 1A). This contribution to the signal intensity was proportional to the number of prolines per peptide (supplemental Fig. 1, A and B). Using a Perl script kindly provided by Jacob D. Jaffe (Broad Institute, Cambridge, MA), new protein group log2 heavy/light (H/L) ratios were calculated. In a first step, the intensity values of heavy peptides with prolines were corrected by adding to the measured signal intensity a value, which corresponds to 20% × n of heavy peptides abundance, where n is the number of prolines in the peptide. In a second step, adjusted peptide log2 H/L ratios were built and combined into the protein groups based on the MaxQuant protein group identifications. For each protein group median log2 H/L ratios were calculated. The protein group log2 H/L median ratios were calculated separately for each biological replicate; median was normalized to 1 and used for further analysis (supplemental Table 2). The effect of the arginine-to-glutamate conversion was negligible, because only 15 peptides identified had a heavy glutamate with heavy arginine and/or heavy lysine. Corrected peptide intensity values were used to calculate the corrected iBAQ values (intensity-based absolute quantification). The C. elegans database was digested in silico with trypsin using the Skyline software (version 1303, MacCoss Lab (54)), and the number of tryptic peptides per protein was extracted. The length of the peptides ranged from 6 to 30 amino acids; the N-terminal amino acids and potential ragged ends were not excluded. Using an in-house R function, corrected intensities of peptides per protein were summed up and divided by the corresponding number of tryptic peptides. Heavy and light iBAQ values were calculated separately and normalized using an intercept of 0 and slopes of 1.01896, 1.00456, and 1.00487 for the first, second, and third replicates, respectively. The proline corrected log2 iBAQ and protein abundance ratios correlated well with the uncorrected ratios (0.66 and 0.71, respectively), and the correlations for the differentially expressed proteins were even stronger (~0.9, supplemental Fig. 1, C and D). Proline corrected heavy and light peptide intensities correlated well between the replicates (r ≥0.74, see supplemental Fig. 2).

**Experimental Design and Statistical Rationale**—mRNAs were quantified in three biological replicates. In total, six worm samples were analyzed, three for each strain (N2 and CB4856). Gene expression differences between the genotypes were determined by a linear model using “log2 hybridization intensities ~ genotype + error.” Correction for multiple testing was done using a permutation determined threshold (3, 4, 31, 55). For each probe the values were

N2 and CB4856 Proteome and Transcriptome Comparison

Molecular & Cellular Proteomics 15.5
permutated once and analyzed by the same model. The FDR rate was set to a ratio of false to true positives of 0.013, which occurred at \( p \leq 0.0032 \) (\(-\log_{10}(p) = 2.5\)). False positive was the number of genes with a significant \( p \) value from the permuted set, and true positives were the number of genes with significant values from the original set. Proteins were quantified in three biological replicates, including label switching. For every biological replicate, two biochemical and two technical replicates were analyzed. In total, six combined protein samples were measured. To identify proteins that are differentially expressed between N2 and CB4856, SILAC data were filtered for proteins that were quantified at least twice with at least a 1.3-fold difference in abundance (the biological significance of the 1.3-fold change cutoff has been shown by Ref. 56, and a \( p \) value \( < 0.000457 \) (\( z \)-test, permutation determined threshold with the FDR cutoff of 0.05)).

**Enrichment Analysis**—The enrichment analyses were undertaken using the following databases with annotations. Gene ontologies were downloaded from Ensembl (version 83) using the BioMart package in R. The enrichment of GO terms was analyzed using the topGO package from the Bioconductor suite in R (58). GO terms with less than two annotated genes were excluded from the analysis. The enrichments were computed using the Fisher test. The elim algorithm was applied to account for the underlying GO graph topology (57). The anatomy terms, protein domains, and gene classes were obtained via WormMart of the WS220 WormBase release. Genes from WormBook chapters were obtained from the 2012 version of WormBook. eQTLs (8, 9) were obtained from WormQTL (21–23). KEGG pathways were obtained from release 65.0 of the Kyoto Encyclopedia of Genes and Genomes. These were tested for the enrichments by a hypergeometric test using R. Specific gene sets were obtained from supplementary data of the following papers: the aging set (59); DAF-16 datasets (60, 61); and the DAF-16/PQM-1 set (62). The enrichments for specific gene sets were performed using Benjamini-Hochberg multiple testing corrected two-sided Fisher exact (aging and DAF-16 datasets) and hypergeometric tests (DAF-16/PQM-1 dataset). Note that the hypergeometric test and one-sided Fisher exact test are basically the same tests (63).

**Data Storage**—The transcript profiles and protein levels were stored in WormQTL (21–23) and can be accessed on line. The MS proteomics data have also been deposited to the ProteomeXchange Consortium (64) via the PRIDE partner repository with the dataset identifier PXD002010.

**RESULTS**

**Transcriptome Comparison**—As the strongest mRNA expression differences between N2 and CB4856 were observed at stage L4 (3, 4, 30–32), we measured the genome-wide transcription levels of three biological replicates of L4 stage synchronized larvae using microarrays. We found 1,532 genes to be differentially expressed (7.4%, \(-\log_{10}(p) > 2.5;\) FDR = 0.013) of which 712 showed higher expression in CB4856 and 820 higher expression levels in N2 (Fig. 1). Differentially expressed genes were enriched for groups of genes involved in binding of proteins, sugars, and DNA, like f-box, math/bath/btb, clec, and nhr genes (supplemental Table 4) as observed previously (3, 4, 32). These groups of genes are highly morphologic between wild isolates (4, 33, 34). No major differences in genes linked to L4 development (30) could be detected between CB4856 and N2. We identified specific enrichments between the genes expressed higher in N2 compared with CB4856 (supplemental Table 4). For example, the GO-term “monooxygenase activity” was specifically enriched for genes expressed higher in CB4856. Because many of the gene expression differences between CB4856 and N2 are caused by their genetic differences, most of these will also be present in recombinant offspring. The loci causal for gene expression differences can be observed as eQTLs, which yield information about the genetic and regulatory architecture of gene expression. As expected, the group of 1,532 genes differentially expressed between CB4856 and N2 was highly enriched for genes with eQTLs that were previously identified in these recombinant inbred line populations (6, 8, 9) (928/1,532; \( p \ll 10^{-100} \) (8), 431/1,532; \( p \ll 10^{-100} \) (9)). To investigate whether these mRNA expression level differences are also found on the protein level, we measured protein abundance in both strains.

**Protein Abundances Are Very Similar in N2 and CB4856**—We quantitatively compared protein abundances between the two C. elegans wild-type strains N2 and CB4856 in the L4 stage using SILAC and mass spectrometry. Protein extracts were analyzed in three biological replicates, including label switching. Almost complete labeling (99.6%) of the second generation L4 worms was observed. In total, we quantified 3,238 distinct proteins, ranging from 2,536 to 2,622 proteins.
per SILAC experiment. 2,002 proteins (61.8%) were quantified in all three biological replicates and 2,485 proteins in at least two (protein FDR 5%, Fig. 2A, and supplemental Table 1). We compared our SILAC data to the integrated PaxDB dataset (65), which includes protein abundance data generated using various fractionation methods. We quantified 24.5% of the *C. elegans* proteins presented in PaxDB (13,224) in N2 and CB4856, distributed over 6 orders of magnitude, with a clear bias for high abundance proteins (Fig. 2B). Protein abundance between the two strains correlated very strongly.

Fig. 2. Proteins quantified in three CB4856/N2 SILAC experiments. A, Venn diagram of the quantified proteins in three biological replicates. In total 3,238 proteins were quantified, and 2,002 (61.8%) proteins were quantified in all three biological replicates. Proteins quantified in at least two biological replicates (2,485; 76.7%) in the pink sections were used for further analysis. B, log2 MS intensity distribution of the proteins quantified at least twice (pink section) compared with the entire *C. elegans* proteome. C, volcano plot of protein abundance differences between N2 and CB4856. Differentially expressed proteins (129; 5.2%) are shown in violet (threshold: \( p < 0.000457 \), fold-change \( >1.3 \)), and non-significant changes (2,356; 94.8%) are shown in gray. Inset represents the enlarged area from a \( p \) value of 0.95 to \( e^{-40} \). Proteins with a \( p \) value of 0 were set to \( e^{-290} \) for plotting.
To identify proteins that are differentially expressed between N2 and CB4856, the combined SILAC data (3,238 proteins) were filtered for proteins that were quantified at least twice (2,485 proteins) with at least a 1.3-fold difference in abundance and a \( p \) value \(< 0.000457 \) (\( z \)-test, FDR = 0.05, Fig. 2C). Using these criteria, 129 proteins were differentially expressed between N2 and CB4856. CB4856 in yellow and N2 in green, orange, and dark green dots represent the 129 differentially expressed proteins in CB4856 and N2. A, mRNA versus protein abundance. CB4856 and N2 are shown in yellow and green, respectively. Orange and dark green dots indicate other unique mRNAs overlapping with the protein dataset. For different protein isoforms, the same mRNA value was used for plotting. \( r \) is the Pearson correlation coefficient for all genes quantified at mRNA and/or protein levels. (a.u., arbitrary units).
proteins (5.2%) were found to be differentially expressed between the strains distributed over 5 orders of magnitude in expression. 70 proteins were up-regulated, and 59 proteins were down-regulated (Fig. 5).

Local and Distant Regulation of the Conserved mRNA and Protein Expression—In total, 2,481 quantified proteins (14 proteins with two isoforms each included) overlapped with 2,467 corresponding unique mRNAs in both strains (supplemental Table 3). The mRNA and protein expression patterns of N2 and CB4856 were very similar, whereas the correlation between mRNA and protein abundances within the strains was quite variable ($r = 0.54$ for both, Fig. 3A). By contrast, mRNA abundance, protein abundance, as well as protein-to-mRNA ratios correlated strongly between the strains ($r > 0.97$, Fig. 3, B–D). This suggests that for the detected genes both transcriptional and translational regulation of gene expression are well conserved between N2 and CB4856. The conservation of protein expression was also observed between distantly related species (26).

We also compared the protein abundance differences versus mRNA abundance differences between the two strains (Fig. 4). Of the 2,481 quantified genes, only 24 genes were found to be changed significantly at both the mRNA and protein levels between N2 and CB4856, and 105 proteins were different only at the protein level and 89 only at the mRNA level. To investigate the genetic loci possibly responsible for the mRNA and protein differences, we compared our data to eQTLs found in Vinuela et al. (9) and Rockman et al. (8). The vast majority of genes (16/23 and 23/25) found to be differentially expressed at both mRNA and protein levels had a QTL (cis- or trans) associated with them, with cis-eQTLs outnumbering trans-eQTLs with a ratio of about 2 to 1. By contrast, for the genes with no expression difference, eQTLs were relatively infrequent (~7–21%), with a slight bias toward trans-eQTLs. Genes varying at the protein or mRNA level only had intermediate eQTL frequencies, with the number of trans-eQTLs being higher for the genes differentially expressed only on a protein level (Table I).

Aging and Stress Response-related Proteins Are Enriched among the 129 Differentially Expressed Proteins—To characterize the function of the 129 differentially expressed proteins, we first analyzed their GO annotations. In our dataset, the categories carbohydrate metabolisms, innate immune response, defense response to bacterium, and cysteine biosynthesis from serine were over-represented. We have also found the NPL4 domain and domains involved in sugar metabolism over-represented among other enriched multiple protein domains (supplemental Table 4). Interestingly, np14 genes were shown to be involved in the life span of C. elegans (8, 66). As altered sugar metabolism, oxidative stress, and pathogen resistance pathways were shown to be associated with aging (67–70), and because the average life span of CB4856 is shorter than that of N2 (41, 50, 71), we compared our list with recently published aging and stress response transcriptome datasets. Transcripts differentially regulated during aging ($p = 0.007$ (59) and transcripts responsive to the DAF-16 transcription factor ($p = 0.0002$ (60, 61)) were over-represented (Fig. 5 and supplemental Table 5). 2,485 proteins were used as the background list, and Benjamini-Hochberg multiple testing correction was applied. We further compared our data to a recent study that showed that the DAF-16-associated element-binding transcription factor PQM-1 complements DAF-16 aging-related functions (62). Genes transcriptionally regulated by PQM-1 alone were enriched in the 2,485 quantified proteins ($p < 2.15e^{-13}$, data not shown), whereas DAF-16 and PQM-1 targets were not ($p = 0.76$, data not shown). By contrast, among the 129 differentially expressed proteins, enrichment was found only for the DAF-16 and PQM-1 targets ($p = 0.0008$), whereas the PQM-1-specific targets were not enriched ($p = 0.24$, Fig. 5 and supplemental Table 5). This shows that highly abundant proteins have a higher chance to be PQM-1 targets, and the variation in protein abundance acts downstream of DAF-16 and PQM-1. In total, 51 proteins were enriched in at least one of the mentioned stud-
ies, and 18 proteins were identified in at least two of them (Fig. 5 and supplemental Table 5). Together, these observations suggest that the 129 differentially expressed proteins might contribute to the differences in life span of CB4856 and N2.

The 129 differentially expressed proteins were also enriched for eQTLs identified in aging worms (9) (47 proteins, $p < 10^{-16}$) and for stage-specific eQTLs (8) (62 proteins, $p < 10^{-16}$) (Table I). Eleven proteins were found to have an eQTL under mildly stressful conditions of 24 °C (Y37A1B.5, Y48G1A.4, SSP-16, GRSP-3, MCE-1, MPPB-1, TATN-1, T06F4.1a, NDX-4, SPDS-1, and LEC-8) (6). Combined, these eQTL enrichments provide further evidence suggesting that at least part of the protein expression level differences can be explained by mRNA expression level differences between CB4856 and N2, and they might play a role in aging.

---

### TABLE 1

Summary of the enriched eQTLs for the genes differentially expressed at mRNA and/or protein levels

| Study/eQTL                  | No difference | mRNA levels changed | Protein levels changed | Both levels changed |
|-----------------------------|---------------|---------------------|------------------------|---------------------|
| Vinuela et al. 2010 all present in dataset | 2,270         | 91                  | 103                    | 23                  |
| Vinuela et al. 2010 cis-eQTLs | 23 (1%)       | 11 (12.1%)          | 8 (7.8%)               | 11 (47.8%)          |
| Vinuela et al. 2010 trans-eQTLs | 130 (5.7%)   | 9 (9.9%)            | 20 (19.4%)             | 5 (21.7%)           |
| Rockman et al. 2010 all present in dataset | 2,288         | 86                  | 109                    | 25                  |
| Rockman et al. 2010 cis-eQTLs | 209 (9.1%)   | 40 (46.5%)          | 20 (18.3%)             | 18 (72%)            |
| Rockman et al. 2010 trans-eQTLs | 266 (11.6%) | 13 (15.1%)         | 28 (25.7%)             | 5 (20%)             |

Fig. 5. Functional categorization of the 129 differentially expressed proteins. Differentially expressed proteins were significantly enriched in genes that are transcriptionally regulated during aging (yellow (59)), by the DAF-16 transcription factor (green (60, 61)), or by the PQM-1 and DAF-16 transcription factors (violet (62)) based on Benjamini-Hochberg-corrected two-sided Fisher exact and hypergeometric tests. Values shown at average ± S.D. of at least two biological replicates.
Genes with eQTLs on the npr-1 Locus Are Enriched among the 129 Proteins—CB4856 favors lower oxygen concentrations, mainly caused by an allelic difference in NPR-1 (39, 40), and this is reflected in its clumping behavior (72, 73). We found an enrichment for genes with an eQTL on the npr-1 locus in the 129 candidates, suggesting that at least part of the gene expression differences lead to differences in protein levels. Furthermore, two cysteine synthases, CYSL-2 and CYSL-3, proposed sensors of cellular H2S levels upon hypoxia (74), showed up-regulated protein levels in CB4856 (Fig. 5). The oxygen-binding protein GLB-1, the transcript of which is up-regulated under hypoxic conditions and in hif-1 mutants (75, 76), was also up-regulated in CB4856 (Fig. 5), as well as the oxygen-sensing globin protein GLB-5 (72, 73). These differences on protein level might contribute to the different oxygen responses of N2 and CB4856.

DISCUSSION

In this paper, we analyzed the influence of genetic background and natural variation on the proteome and the transcriptome in two genetically highly divergent C. elegans wild-type strains, N2 (Bristol) and CB4856 (Hawaii). Using SILAC, we quantified in total 3,238 distinct proteins in three biological replicates; 2,485 proteins were quantified in at least two replicates. These numbers are comparable with other SILAC studies in C. elegans. More than 1,400 proteins were reliably quantified in worm samples using size exclusion chromatography (28), and 3,470 proteins were quantified using LysC digestion followed by hydrophilic interaction liquid chromatography separation of the peptides (29). Between 1,461 and 4,072 proteins were identified in different developmental stages of C. elegans and C. briggsae after separation on SDS gels (27). To quantify less abundant proteins, additional separation methods would be required, as was done for the quantitative comparison of distantly related species (26). In our study, primarily highly expressed proteins were quantified. These show relatively little variation between N2 and CB4856, also at the transcript level.

Among the 2,485 quantified proteins, 129 (5.2%) proteins distributed over 5 orders of magnitude in expression were differentially expressed between N2 and CB4856. By contrast, of all the genes, 1,532 genes (7.4%) were differentially expressed at the mRNA level. Of the 2,481 genes for which proteins were detected and that overlapped with the mRNA dataset, only 113 (4.6%) were different at the mRNA level. The overall correlations for both protein and mRNA abundances were close to one between the strains. This is consistent with the quantitative comparison of C. elegans and C. briggsae (27), but it is different for C. elegans and D. melanogaster, where the conservation in expression was found only on protein level (26). Thus, in this intraspecies comparison, we observed a strong selective pressure to maintain protein levels, an effect that was also shown for closely (27) and even distantly related species (26).

The 129 differentially expressed proteins were strongly enriched for downstream targets of the insulin-signaling pathway (DAF-16 and PQM-1 targets) (60, 61, 62). An increased expression divergence of DAF-16 targets was also observed in a quantitative comparison of gene expression between the two nematode species C. briggsae and C. elegans (27). These results imply that differential regulation of the insulin-signaling downstream targets can be observed already between strains of the same species, underlining the strong divergence of this signaling pathway in nematodes. Importantly, the genes differentially expressed between N2 and CB4856 were also strongly enriched in eQTLs associated with aging. These results suggest that at least a subset of our differentially expressed genes might contribute to the longevity differences between N2 and CB4856.

Acknowledgments—We thank Xiaoxue Li and Xue Zheng for carrying out pre-experiments in proteomics and Peter Hunziker for valuable advice. We thank Nikolaus Rajewsky and Nadine Thierfelder for kindly providing the SILAC labeling protocol and Jacob D. Jaffe for help with the correction of arginine-to-proline conversion. We also thank Morris Swertz and Joeri van der Velde for help in making the data accessible through WormQTL. We thank Christian Panse for help with MS/MS spectra annotation of the single peptide identifications.

* This work was supported by European Community Health Seventh Framework Programme FP7/2007-2013 under Grant PANACEA 222936, the Swiss National Science Foundation, and the Kanton of Zurich. L.B.S. was supported by Netherlands Organization for Scientific Research Project 823.01.001. R.J.M.V. was supported by the NWO-ALW Project 855.01.151. M.G.S. was funded by the Graduate School Production Ecology and Resource Conservation. The authors declare that they have no conflicts of interest with the contents of this article.

[5] This article contains supplemental material.

†† To whom correspondence should be addressed: Laboratory of Nematology, Wageningen Universiteit, Droevendaalsesteeg 1, NL-6708 PB, Wageningen, The Netherlands. Tel.: 31-317-482998/482197; Fax: 31-317-484254; E-mail: jan.kammenga@wur.nl or Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland. Tel.: 41-44-6353112; E-mail: sabine.schrimpf@imls.uzh.ch.

¶ Both authors contributed equally to this work.

REFERENCES

1. Oleksiak, M. F., Churchill, G. A., and Crawford, D. L. (2002) Variation in gene expression within and among natural populations. Nat. Genet. 32, 261–266
2. Gaffney, D. J. (2013) Global properties and functional complexity of human gene regulatory variation. PLOS Genet. 9, e1003501
3. Viruela, A., Snoek, L. B., Riksen, J. A., and Kammenga, J. E. (2012) Aging uncouples heritability and expression-QTL in Caenorhabditis elegans. G3 2, 597–605
4. Volkers, R. J., Snoek, L. B., Hubar, C. J., Coopman, R., Chen, W., Yang, W., Sterken, M. G., Schulenburg, H., Braeckman, B. P., and Kammenga, J. E. (2013) Gene-environment and protein-degradation signatures characterize genomic and phenotypic diversity in wild Caenorhabditis elegans populations. BMC Biol. 11, 93
5. Brem, R. B., Yvert, G., Clinton, R., and Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296, 752–755
6. Li, Y., Alvarez, O. A., Guttelting, E. W., Tijsterman, M., Fu, J., Riksen, J. A., Hazendonk, E., Prins, P., Plasterk, R. H., Jansen, R. C., Breitling, R., and
Kamminga, J. E. (2006) Mapping determinants of gene expression plasticity by genetic linkages in C. elegans. PLoS Genet. 2, e222

7. Li, Y., Breitling, R., Snoek, L. B., van der Velde, K. J., Swertz, M. A., Riksen, J., Jansen, R. C., and Kamminga, J. E. (2010) Global genetic robustness of the alternative splicing machinery in Caenorhabditis elegans. Genetics 186, 405–410

8. Rockman, M. V., Skrovanek, S. S., and Kruglyak, L. (2010) Selection at linked sites shapes hortepitic phenotypic variation in C. elegans. Science 330, 372–376

9. Vrueuia, A., Snoek, L. B., Riksen, J. A., and Kamminga, J. E. (2010) Genome-wide gene expression regulation as a function of genotype and age in C. elegans. Genome Res. 20, 929–937

10. Snoek, L. B., Terpstra, I. R., Dekker, R., Van den Ackerveken, G., and Peeters, A. J. (2012) Genetic genomics reveals large scale genotype-phenotype interactions in Arabidopsis thaliana. Front. Genet. 3, 317

11. King, E. G., Sanderson, B. J., McNeil, C. L., Long, A. D., and Macdonald, S. J. (2014) Genetic dissection of the Drosophila melanogaster female head transcriptome reveals widespread allelic heterogeneity. PLoS Genet. 10, e1004322

12. Jansen, R. C., and Nap, J. P. (2001) Genetical genomics: the added value from segregation. Trends Genet. 388–391

13. Birney, E. and Hoeschele, I. (2005) Genetical genomics analysis of a yeast segregant population for transcription network inference. Genetics 170, 533–542

14. Keurentjes, J. J., Fu, J., Terpstra, I. R., Garcia, J. M., van den Ackerveken, G., Snoek, L. B., Peeters, A. J., Vreugdenhil, D., Koornneef, M., and Jansen, R. C. (2007) Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proc. Natl. Acad. Sci. U.S.A. 104, 1708–1713

15. Terpstra, I. R., Snoek, L. B., Keurentjes, J. J., Peeters, A. J., and van den Ackerveken, G. (2010) Regulatory network identification by genetical genomics: signaling downstream of the Arabidopsis receptor-like kinase ERECTA. Plant Physiol. 154, 1067–1078

16. Cox, J., and Mann, M. (2007) Is proteomics the new genomics? Cell 130, 395–398

17. Albert, F. W., Treusch, S., Shockley, A. H., Bloom, J. S., and Kruglyak, L. (2011) Genetics of single-cell protein abundance variation in large yeast populations. Nature 506, 494–497

18. Foss, E. J., Radulovic, D., Shaffer, S. A., Goodlett, D. R., Kruglyak, L., and Rajewsky, N. (2014) Conservation of mRNA and protein expression during development of C. elegans. Cell Rep. 6, 565–577

19. Laranca, M., Baille, A. P., Pourkarim, E., Hay, R. T., Buchman, G., Coulthurst, S., Xirodimas, D. P., Gartner, A., and Lamond, A. I. (2011) Stable-isotope labeling with amino acids in netomedes. Nat. Methods 8, 849–851

20. Watanabe, K., Eng, H., Aoki, S., Kato, Y., Norimatsu, H., and Aoki, S. (2010) Comparative proteomics of Caenorhabditis elegans. Proteomics 10, 4072–4080

21. Capra, E. J., Skrovanek, S. S., and Kruglyak, L. (2008) Comparative developmental expression profiling of two C. elegans isolates. PLoS One 3, e4055

22. Thompson, O., Edgley, M., Strasbourger, P., Filibotte, S., Ewing, B., Adair, R., Au, V., Chaudhry, I., Lombardo, S., Hutter, H., Kieffer, A., Lau, J., Lee, N., Miller, A., Raymant, G., et al. (2013) The million mutation project; a new approach to genetics in Caenorhabditis elegans. Genome Res. 23, 1749–1762

23. Mandre, A. and Félix, M. A. (2005) High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations. Curr. Biol. 15, 1176–1184

24. Andersen, E. C., Gerke, J. P., Shapiro, J. A., Crissman, J. R., Ghosh, R., Bloom, J. S., Félix, M. A., and Kruglyak, L. (2012) Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. Nat. Genet. 44, 285–290

25. Barriere, A., and Félix, M. A. (2005) Natural variation and population genet-ics of Caenorhabditis elegans. WormBook 2005, 1–19

26. Maydan, J. S., Filibotte, S., Edgley, M., Lau, J., Selzer, R. R., Richmond, T. A., Pofahl, N. J., Thomas, J. H., and Moerman, D. G. (2007) Efficient high-resolution deletion discovery in Caenorhabditis elegans by array comparative genomic hybridization. Genome Res. 17, 337–347

27. Schmid, T., Snoek, L. B., Nijveen, H., Schram, P., van der Bent, M. L., Kammenga, J. E., and Harvey, S. C. (2013) The laboratory domestication of Caenorhabditis elegans. Trends Genet. 29, 224–231

28. Andersen, E. C., Bloom, J. S., Gerke, J. P., and Kruglyak, L. (2014) A variant in the neuropeptide receptor npr-1 is a major determinant of Caenor- habditis elegans growth and physiology. PLoS Genet. 10, e1004156

29. Statina, J. S., Snoek, L. B., Kamminga, J. E., and Harvey, S. C. (2015) Genotype-dependent lifespan effects in peptone deprived Caenorhabditis elegans. Sci. Rep. 5, 16259

30. Schmid, T., Snoek, L. B., Fröhlich, E., van der Bent, M. L., Kamminga, J., and Najwal, A. (2015) Systemic regulation of RAS/MAPK signaling by the serotonin metabolite 5-HIAA. PLoS Genet. 11, e1005236

31. Green, J. W., Snoek, L. B., Kamminga, J. E., and Harvey, S. C. (2013) Genetic mapping of variation in dauer larvae development in growing populations of Caenorhabditis elegans. Heredity 111, 306–313

32. Duveaux, F., and Félix, M. A. (2012) Role of pleiotropy in the evolution of a cryptic developmental variation in Caenorhabditis elegans. PLoS Biol. 10, e1001230

33. Reedy, K. C., Andersen, E. C., Kruglyak, L., and Kim, D. H. (2009) A polymorphism in npr-1 is a behavioral determinant of pathogen suscept-ibility in C. elegans. Science 323, 382–384

34. Glauser, D. A., Chen, W. C., Agin, R., Macinnis, B. L., Hellman, A. B., Garrity, P. A., Tan, M. W., and Goodman, M. B. (2011) Heat avoidance is
regulated by transient receptor potential (TRP) channels and a neuro-peptide signaling pathway in Caenorhabditis elegans. Genetics 188, 91–103.

47. Gaetner, B. E., Parmarter, M. D., Rockman, M. V., Kruglyak, L., and Phillips, P. C. (2012) More than the sum of its parts: a complex epistatic network underlies natural variation in thermal preference behavior in Caenorhabditis elegans. Genetics 192, 1533–1542.

48. Bendesky, A., Pitts, J., Rockman, M. V., Chen, W. C., Tan, M. W., Kruglyak, L., and Bargmann, C. I. (2012) Long-range regulatory polymorphisms affecting a GABA receptor constitute a quantitative trait locus (QTL) for social behavior in Caenorhabditis elegans. PLoS Genet. 8, e1003157.

49. Bendesky, A., Tsuzuki, M., Rockman, M. V., Kruglyak, L., and Bargmann, C. I. (2011) Catecholamine receptor polymorphisms affect decision-making in C. elegans. Nature 472, 313–318.

50. Rodríguez, M., Snoek, L. B., Riksen, J. A., Bevers, R. P., and Kammenga, J. E. (2012) Genetic variation for stress-response hoxmerism in C. elegans lifespan. Exp. Gerontol. 47, 581–587.

51. Zahurak, M., Parmigiani, G., Yu, W., Scharpf, R. B., Berman, D., Schaeffer, E., Shabbeer, S., and Cope, L. (2007) Pre-processing agilent microarray data. BMC Bioinformatics 8, 142.

52. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372.

53. Panse, C. G., Grossman, J., and Oesterreicher, S. B. (2013) protViz: Visualizing and analyzing mass spectrometry related data in proteomics. R Package, version 3.0.2, Functional Genomics Center Zurich, Zurich.

54. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966–968.

55. Viruelas, A., Snoek, L. B., Riksen, J. A., and Kammenga, J. E. (2010) Genome-wide gene expression analysis in response to organophosphorus pesticide chlorpyrifos and diazinon in C. elegans. PLoS One 5, e12145.

56. Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Groth, M., Platzer, M., Ristow, M., and Guthke, R. (2013) Extension of life span by impaired glucose metabolism in Caenorhabditis elegans is accompanied by structural rearrangements of the transcriptomic network. PLoS One 8, e77776.

57. 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. elegansdaf-2 mutants are resistant to bacterial pathogens. Science 300, 1921.

58. Papp, D., Csernely, P., and Sőti, C. (2012) A role for SKN-1/Nrf in pathogen resistance and immunosenescence in Caenorhabditis elegans. PLoS Pathog. 8, e1002673.

59. Doroszuk, A., Snoek, L. B., Fradin, E., Riksen, J., and Kammenga, J. (2009) A genome-wide library of CB4856/N2 introgression lines of Caenorhabditis elegans. Nucleic Acids Res. 37, e110.

60. McGrath, P. T., Rockman, M. V., Zimmer, M., Jang, H., Macoesko, E. Z., Kruglyak, L., and Bargmann, C. I. (2009) Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors. Neuron 61, 692–699.

61. Persson, A., Gross, E., Laurent, P., Busch, K. E., Bretes, H., and de Bono, M. (2009) Natural variation in a neural globin tunes oxygen sensing in wild Caenorhabditis elegans. Nature 458, 1030–1033.

62. Gabszvard, B., Ahmed, S., Li, L., Arii, V. M., Moore, P. K., and Stürzenbaum, S. R. (2013) C. elegans aging is modulated by hydrogen sulfide and the sulfhydrylase/cysteine synthase cysl-2. PLoS One 8, e60135.

63. Geuens, E., Hoogewijs, D., Nardini, M., Vinck, E., Pesce, A., Kiger, L., Fago, A., Tilleman, L., de Henau, S., Marden, M. C., Weber, R. E., Van Doorslaer, J., Vanfleteren, J., Moens, L., Bolognesi, M., and Dewilde, S. (2010) Globin-like proteins in Caenorhabditis elegans: in vivo localization, ligand binding and structural properties. BMC Biochem. 11, 17.

64. Hoogewijs, D., Geuens, E., Dewilde, S., Vlerstraete, A., Moens, L., Vinogradov, S., and Vanfleteren, J. R. (2007) Wide diversity in structure and expression profiles among members of the Caenorhabditis elegans globin protein family. BMC Genomics 8, 356.