

2016-05-02

## The genetic basis of natural variation in *C. elegans* telomere length [preprint]

Daniel E. Cook  
*Northwestern University*

*Et al.*

Let us know how access to this document benefits you.

Follow this and additional works at: [https://escholarship.umassmed.edu/faculty\\_pubs](https://escholarship.umassmed.edu/faculty_pubs)



Part of the [Genetic Phenomena Commons](#), [Genetics Commons](#), and the [Population Biology Commons](#)

---

### Repository Citation

Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi D, Noble LM, Rockman MV, Alkema MJ, Braendle C, Kammenga J, Wang J, Kruglyak L, Felix M, Lee J, Andersen E. (2016). The genetic basis of natural variation in *C. elegans* telomere length [preprint]. University of Massachusetts Medical School Faculty Publications. <https://doi.org/10.1101/051276>. Retrieved from [https://escholarship.umassmed.edu/faculty\\_pubs/1573](https://escholarship.umassmed.edu/faculty_pubs/1573)

Creative Commons License



This work is licensed under a [Creative Commons Attribution 4.0 License](#).

This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in University of Massachusetts Medical School Faculty Publications by an authorized administrator of eScholarship@UMassChan. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

## The genetic basis of natural variation in *C. elegans* telomere length

Cook, D.C.<sup>1,2</sup>, Zdraljevic, S.<sup>1,2</sup>, Tanny, R.E.<sup>1</sup>, Seo, B.<sup>3</sup>, Riccardi D.D.<sup>4,5</sup> Noble L.M.,<sup>4,5</sup> Rockman M.V.<sup>4,5</sup>, Alkema, M.J.<sup>6</sup>, Braendle, C.<sup>7</sup>, Kammenga, J.E.<sup>8</sup>, Wang, J.<sup>9</sup>, Kruglyak L.<sup>10,11</sup>, Félix, MA.<sup>12</sup>, Lee, J.<sup>3,13</sup>, and Andersen, E.C.<sup>1,14,15,16,\*</sup>

1. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208
  2. Interdisciplinary Biological Science Program, Northwestern University, Evanston, IL 60208
  3. Department of Biological Sciences, Institute of Molecular Biology and Genetics, Seoul National University, Seoul 08826, Korea.
  4. Department of Biology, New York University, New York, NY 10003
  5. Center for Genomics and Systems Biology, New York University, New York, NY 10003
  6. Department of Neurobiology, University of Massachusetts Medical School, Worcester, MA 01605
  7. Univ. Nice Sophia Antipolis, CNRS, Inserm, IBV, Parc Valrose, 06100 Nice, France
  8. Laboratory of Nematology, Wageningen University, Wageningen, The Netherlands
  9. Biodiversity Research Center, Academia Sinica, Nankang Taipei, Taiwan
  10. Departments of Human Genetics and Biological Chemistry, University of California at Los Angeles, Los Angeles, CA 90095
  11. Howard Hughes Medical Institute, Chevy Chase, MD 20815
  12. Institut de Biologie de l'École Normale Supérieure, CNRS-Inserm, Paris, France
  13. Department of Biophysics and Chemical Biology, Seoul National University, Seoul 08826, Korea.
  14. Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Chicago, IL 60611
  15. Chemistry of Life Processes Institute, Northwestern University, Evanston, IL 60208
  16. Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208
- \* Corresponding Author

**Corresponding author:**

Erik C. Andersen

Dept. of Molecular Biosciences

Northwestern University

2205 Tech Dr.

Evanston, IL 60208

[erik.andersen@northwestern.edu](mailto:erik.andersen@northwestern.edu)

847-467-4382

## 1 **Abstract**

2 Telomeres are involved in the maintenance of chromosomes and the prevention of  
3 genome instability. Despite this central importance, significant variation in telomere  
4 length has been observed in a variety of organisms. The genetic determinants of  
5 telomere-length variation and their effects on organismal fitness are largely unexplored.  
6 Here, we describe natural variation in telomere length across the *Caenorhabditis*  
7 *elegans* species. We identify a large-effect variant that contributes to differences in  
8 telomere length. The variant alters the conserved oligosaccharide/oligonucleotide-  
9 binding fold of POT-2, a homolog of a human telomere-capping shelterin complex  
10 subunit. Mutations within this domain likely reduce the ability of POT-2 to bind telomeric  
11 DNA, thereby increasing telomere length. We find that telomere-length variation does  
12 not correlate with offspring production or longevity in *C. elegans* wild isolates,  
13 suggesting that naturally long telomeres play a limited role in modifying fitness  
14 phenotypes in *C. elegans*.

## 15 **Introduction**

16  
17 Genome-wide association (GWA) studies, in which phenotypic differences are  
18 correlated with genome-wide variation in populations, offer a powerful approach to  
19 understand the genetic basis of complex traits (McCarthy *et al.* 2008). GWA requires  
20 accurate and quantitative measurement of traits for a large number of individuals. Even  
21 in organisms that are studied easily in the laboratory, the measurement of quantitative  
22 traits is difficult and expensive. By contrast, the rapid decrease in sequencing costs has  
23 made the collection of genome-wide variation accessible. From *Drosophila* (Mackay *et al.*  
24 *al.* 2012; Lack *et al.* 2015) to *Arabidopsis* (Weigel and Mott 2009) to humans (Project *et*  
25 *al.* 2012), the whole genomes from large populations of individuals can be analyzed to  
26 identify natural variation that is correlated with quantitative traits. Because the genome  
27 itself can vary across populations, whole-genome sequence data sets can be mined for  
28 traits without measuring the physical organism. Specifically, large numbers of sequence  
29 reads generated from individuals in a species can be analyzed to determine attributes of  
30 genomes, including mitochondrial or ribosomal DNA copy numbers. Another such trait is  
31 the length of the highly repetitive structures at the ends of linear chromosomes called  
32 telomeres (Blackburn 1991).

33  
34 Telomeres are nucleoprotein complexes that serve as protective capping structures that  
35 prevent chromosomal degradation and fusion (O'Sullivan and Karlseder 2010). The  
36 DNA component of telomeres in most organisms consists of long stretches of nucleotide  
37 repeats that terminate in a single-stranded 3' overhang (McEachern *et al.* 2000). The  
38 addition of telomeric repeats is necessary because DNA polymerase is unable to  
39 completely replicate the lagging strand (Watson 1972; Levy *et al.* 1992). The length of  
40 telomeres can differ among cell populations (Samassekou *et al.* 2010), from organism to  
41 organism (Fulcher *et al.* 2014), and within proliferating cellular lineages (Frenck Jr. *et*  
42 *al.* 1998). Two antagonistic pathways regulate telomere length. In the first pathway, the  
43 reverse transcriptase telomerase adds *de novo* telomeric repeats to the 3' ends of  
44 chromosomes. In the second, telomere lengthening is inhibited by the shelterin

45 complex. Shelterin forms a protective cap at telomere ends, presumably through the  
46 formation of lariat structures known as t-loops (Griffith *et al.* 1999). The t-loops are  
47 hypothesized to inhibit telomerase activity by preventing access to the 3' tail.  
48 Additionally, because uncapped telomeres resemble double-stranded DNA breaks,  
49 shelterin association with telomeric DNA represses endogenous DNA damage repair  
50 pathways, preventing chromosomal fusion events and preserving genome integrity (De  
51 Lange 2010).

52  
53 Variation in telomere length has important biological implications. In cells lacking  
54 telomerase, chromosome ends become shorter with every cell division, which eventually  
55 triggers cell-cycle arrest (Harley *et al.* 1992). In this way, telomere length sets the  
56 replicative potential of cells and acts as an important tumor-suppressor mechanism  
57 (Harley *et al.* 1992; Deng *et al.* 2008). In populations of non-clonal human leukocytes,  
58 telomere lengths have been shown to be highly heritable (Broer *et al.* 2013).  
59 Quantitative trait loci (QTL) identified from human genome-wide association (GWA)  
60 studies of telomere length implicate telomere-associated genes, including telomerase  
61 (*TERT*), its RNA template (*TERC*), and *OBFC1* (Levy *et al.* 2010; Jones *et al.* 2012;  
62 Codd *et al.* 2013). QTL underlying variation in telomere length have been identified in  
63 *Arabidopsis thaliana*, *Saccharomyces paradoxus*, and *Saccharomyces cerevisiae* using  
64 both linkage and association approaches (Gatbonton *et al.* 2006; Liti *et al.* 2009; Kwan  
65 *et al.* 2011; Fulcher *et al.* 2014). In *S. paradoxus*, natural variation in telomere lengths is  
66 mediated by differences in telomerase complex components. In *S. cerevisiae*, natural  
67 telomere lengthening is caused by a loss of an amino acid permease gene. Thus far, no  
68 studies in multicellular animals or plants have been able to identify specific genes  
69 responsible for population telomere-length differences. Recent advances in wild strain  
70 genotypes and sequences (Andersen *et al.* 2012) in *Caenorhabditis elegans* make it a  
71 powerful model to address natural variation in telomere length and its fitness  
72 consequences.

73

74 Like in humans, telomerase and shelterin activities regulate *C. elegans* telomere length  
75 (Malik *et al.* 2000; Cheung *et al.* 2006; Meier *et al.* 2006; Shtessel *et al.* 2013). The  
76 TRT-1-containing telomerase complex is hypothesized to add TTAGGC repeats to the  
77 ends of chromosomes and prevents chromosome shortening (Meier *et al.* 2006), and  
78 the shelterin complex regulates access of the telomerase complex to chromosome ends  
79 (Raices *et al.* 2008; Cheng *et al.* 2012; Shtessel *et al.* 2013). The length of telomeres in  
80 the laboratory strain N2 is variable and ranges between 2-9 kb (Wicky *et al.* 1996;  
81 Raices *et al.* 2005). The telomere lengths in wild isolates of *C. elegans* are largely  
82 unexplored. Previous studies examined variation in telomere length using a small  
83 number of wild strains (Cheung *et al.* 2004; Raices *et al.* 2005). However, several of the  
84 supposed wild strains have since been determined to be mislabeled versions of the  
85 laboratory strain N2 (McGrath *et al.* 2009). Thus, it is not known if and how telomere  
86 lengths vary among *C. elegans* natural strains. Additionally, the fitness consequences of  
87 telomere length variation have not been defined.

88  
89 Here, we collected a new set of whole-genome sequences from 208 wild *C. elegans*  
90 strains and used these strains to investigate natural variation in telomere length across  
91 the species. Computational estimates of telomere lengths were confirmed using  
92 molecular measurements, indicating that the computational technique can be applied  
93 across this large number of wild strains. Using association mapping, we found that  
94 variation in the gene *pot-2* is correlated with differences in *C. elegans* telomere length.  
95 Natural variation in *pot-2* affects gene function and causes longer than average  
96 telomeres in some wild strains. Additionally, we examined whether population  
97 differences in telomere length connect to differences fitness traits, including brood size  
98 and longevity. Our results indicate that variation in *pot-2* does not correspond with  
99 variation in fitness as measured in the laboratory and does not show strong signatures  
100 of selection in nature. These data suggest that telomere length beyond a basal  
101 threshold is of limited consequence to *C. elegans*. Our results underscore how traits  
102 obtained from sequence data can be utilized to understand the dynamic nature of  
103 genomes within populations.

104

## 105 **Results**

106

### 107 **Whole-genome sequencing of a large number of wild *C. elegans* strains identifies** 108 **new isotypes and highly diverged strains**

109

110 Previous genome-wide analyses of *C. elegans* population diversity used single-  
111 nucleotide variants (SNVs) ascertained from only two strains (Rockman and Kruglyak  
112 2009), reduced representation sequencing that only studied a fraction of the genome  
113 (Andersen *et al.* 2012), or were limited to a small set of wild strains (Thompson *et al.*  
114 2013). To address these limitations, we sequenced the whole genomes of a collection of  
115 208 wild strains (Supplementary File 1). Because *C. elegans* reproduction occurs  
116 primarily through the self fertilization of hermaphrodites, highly related individuals  
117 proliferate in close proximity to one another (Barrière and Félix 2005; Félix and Braendle  
118 2010). As a result, strains isolated from similar locations in nature are frequently  
119 identical and share genome-wide haplotypes or isotypes. Sequencing data generated  
120 from strains belonging to the same isotype can be combined to increase depth of  
121 coverage and to improve downstream analyses. To identify which strains shared the  
122 same genome-wide haplotypes, we compared all of the variation identified in each of  
123 the 208 strains to each other in pairwise comparisons. The 208 strains reduce to 152  
124 unique genome-wide haplotypes or isotypes (Supplementary File 1). The combination of  
125 sequence data from all strains that make up an isotype led to a 70-fold median depth of  
126 coverage (Supplementary Figure 1), enabling the discovery of single nucleotide variants  
127 (SNVs) and other genomic features. The number of SNVs in comparisons of each  
128 isotype to the reference strain N2 ranged from strains highly similar to N2 to 402,436  
129 SNVs (Supplementary Figure 2), and the density of SNVs across the genome  
130 matched previous distributions with more variants on chromosome arms than  
131 centers (Supplementary Figure 3). A clustering analysis of these 152 isotypes  
132 recapitulated the general relationships previously identified among a set of 97 wild



133 isotypes (Andersen *et al.* 2012) (Supplementary Figure 4) with the addition of 55 new  
134 isotypes. Past studies identified one highly diverged strain isolated from San Francisco,  
135 QX1211, which had divergence almost three times the level of other wild *C. elegans*  
136 strains (Andersen *et al.* 2012). Among the 55 new isotypes, one additional strain,  
137 ECA36 from New Zealand, is equally diverged, suggesting that wider sampling will  
138 recover additional diversity for this species. Altogether, our considerably expanded  
139 collection of whole-genome sequence data serves as a powerful tool to interrogate how  
140 natural variation gives rise to differences among individuals in a natural population.

141

### 142 ***C. elegans* wild strains differ in telomere lengths**

143

144 Our collection of high-depth whole-genome sequence data samples a large number of  
145 strains in the *C. elegans* species. The recent development of TelSeq, a program  
146 designed to estimate telomere length using short-read sequence data (Ding *et al.* 2014),  
147 allowed us to examine natural variation in telomere lengths computationally across wild  
148 *C. elegans* strains. We detected considerable natural variation in the total length of  
149 telomeric DNA in a strain (Figure 1), ranging from 4.12 kb to a maximum of 83.7 kb and  
150 a median telomere length of 12.25 kb (Supplementary File 2). The Telseq telomere  
151 length estimate for N2 was 16.97 kb, which is higher than previous estimates (Wicky *et*  
152 *al.* 1996). The distribution of telomere lengths in the *C. elegans* population  
153 approximated a normal distribution with a right tail containing strains with longer than  
154 average telomeres. We found that our computational estimates of telomere length from  
155 Illumina sequence data were significantly influenced by library preparation, possibly  
156 driven by the method of DNA fragmentation (Supplementary Figure 5). However, we  
157 were able to control for these differences using a linear model. We also observed a  
158 weak correlation between depth of coverage and TelSeq length estimates, but  
159 adjustments for library preparation eliminated this relationship (Supplementary Figure  
160 6).

161

162 TelSeq length estimates have been shown to give similar results as molecular methods  
163 to measure human telomere length (Ding *et al.* 2014). As of now, no studies have used  
164 TelSeq to examine *C. elegans* telomeres, so we investigated how well TelSeq estimates  
165 correlated with molecular methods, including Terminal Restriction Fragment (TRF)  
166 Southern blot analyses, quantitative PCR (qPCR) of telomere hexamer sequences, and  
167 fluorescence *in situ* hybridization (FISH) analyses. Using twenty strains, we found that  
168 the results from these molecular assays correlated well ( $\rho = 0.445$  TRF, 0.815 FISH,  
169 0.699 qPCR, Spearman's rank correlation) with computational estimates of telomere  
170 lengths (Figure 2, Supplementary File 3). These molecular results validated our  
171 computational estimates of telomere lengths and indicate that we can use TelSeq  
172 estimates to investigate the genetic causes underlying telomere variation.

173

#### 174 **Species-wide telomere length differences correlate with genetic variation on** 175 **chromosome II**

176

177 To identify the genes that cause differences in telomere length across the *C. elegans*  
178 population, we used a GWA mapping approach as performed previously (Andersen *et*  
179 *al.* 2012) but taking advantage of the larger collection of wild strains. We treated our  
180 computational estimates of telomere length as a quantitative trait and identified one  
181 significant quantitative trait loci (QTL) on the right arm of chromosome II (Figure 3,  
182 Supplementary File 4). To identify the variant gene(s) that underlie this QTL, we  
183 investigated the SNVs within a large genomic region (12.9 to 15.3 Mb) surrounding the  
184 most significant marker on chromosome II. This region contains 557 protein-coding  
185 genes (Supplementary File 5), but only 332 of these genes contained variants that are  
186 predicted to alter the amino-acid sequences among the 152 strains. We examined  
187 genes with predicted protein coding variants that could alter telomere length by  
188 correlating their alleles with the telomere-length phenotype. Thirty-four genes  
189 possessed variation that was most highly correlated with telomere length ( $\rho \geq 0.4$ ,  
190 Supplementary File 5). The chromosome II QTL explains 28.4% of the phenotypic  
191 variation in telomere length. Three additional suggestive QTL on chromosomes I, II, and

192 III were detected close to but below the significance threshold. Taken together, the four  
193 QTL explain 56.7% of the phenotypic variation in telomere length.

194

### 195 **Variation in *pot-2* underlies differences in telomere length**

196

197 One of the 34 genes in the chromosome II large-effect QTL is *pot-2* (Protection Of  
198 Telomeres 2), a gene which was previously implicated in regulation of telomere length  
199 (Raices *et al.* 2008; Cheng *et al.* 2012; Shtessel *et al.* 2013). Given the large number of  
200 genes present within our confidence interval and challenges associated with examining  
201 telomere length using traditional genetic approaches, we sought alternative methods to  
202 confirm that variation in *pot-2* could cause long telomeres. A quantitative  
203 complementation test could be used to confirm that wild strains have the same  
204 functional effect as a *pot-2* deletion. However, differences in telomere length caused by  
205 mutations in genes that encode telomere-associated proteins often do not have  
206 observable telomere defects for a number of generations (Vulliamy *et al.* 2004;  
207 Armanios *et al.* 2005; Marrone *et al.* 2005). It is technically not feasible to keep the  
208 genome heterozygous during long-term propagation. Fortunately, the ability to  
209 computationally estimate telomere length allowed us to further validate our approach  
210 using data from the Million Mutation Project (MMP) (Thompson *et al.* 2013) and  
211 examine whether the equivalent of a mutant screen for telomere length would provide  
212 insight into our result examining wild isolate genomes.

213

214 The MMP generated over two thousand mutagenized strains using the laboratory N2  
215 background. After each strain was passaged by self-mating of hermaphrodites for ten  
216 generations, the strains were whole-genome sequenced to identify and to predict the  
217 effects of induced mutations. The MMP data set can be used to identify correlations of  
218 phenotype and mutant genes in the laboratory strain background. We obtained whole-  
219 genome sequence data from 1,936 mutagenized N2 strains, each of which has a unique  
220 collection of mutations. Importantly, ten generations of self-propagation of these  
221 mutagenized strains prior to sequencing likely allowed telomere lengths to stabilize in

222 response to mutations in genes that regulate telomere length, enabling us to observe  
223 differences. TelSeq returned telomere length estimates for this population, which had a  
224 right long-tailed distribution (Figure 4A). We classified 39 of 1936 strains within the  
225 population as long-telomere strains with telomere lengths greater than 6.41 kb (98th  
226 percentile). Reasoning that certain mutant genes would be overrepresented in these 39  
227 strains compared to the others, we performed a hypergeometric test to identify if  
228 enrichment for particular genes in long-telomere strains existed. After adjusting for  
229 multiple statistical tests, we identified *pot-2* as highly enriched for mutations in six of the  
230 39 long-telomere strains ( $p = 2.69e^{-11}$ , Bonferroni corrected, Figure 4B). No other genes  
231 within any of the QTL intervals or any other part of the genome were enriched for  
232 mutations among long-telomere strains. This approach was different from association  
233 mapping and identified the same locus regulating telomere length. Additionally, we  
234 computationally examined telomere length from whole-genome sequencing of a *pot-2*  
235 knockout strain. This strain possesses a large deletion that spans the first and second  
236 exons of *pot-2* likely rendering it nonfunctional. We propagated this mutant strain for ten  
237 generations prior to whole-genome sequencing and TelSeq analysis. The telomere  
238 length of *pot-2(tm1400)* mutants was calculated to be 30.62 kb. Given these data, we  
239 have three independent tests that indicate that variation in *pot-2* likely underlies natural  
240 differences in telomere lengths across the *C. elegans* species.

241  
242 Our results are consistent with the established role of *pot-2* as an inhibitor of telomere  
243 lengthening (Shtessel *et al.* 2013). However, no connection of *pot-2* to natural variation  
244 in telomere lengths has been described previously. POT-2 contains an OB-fold thought  
245 to interact with telomeric DNA. OB-folds are involved in nucleic acid recognition (Flynn  
246 and Zou 2010), and the OB-fold of the human POT-2 homolog (hPOT1) binds telomeric  
247 DNA (Lei *et al.* 2004). We investigated the variant sites altered in the *C. elegans*  
248 species along with the mutations found in the MMP mutagenized strains (Figure 5). We  
249 found that the natural variation in *pot-2* resulted in a putative phenylalanine-to-isoleucine  
250 (F68I) change in the OB-fold domain of 12 strains. Strains with the POT-2(68I) allele  
251 have long telomeres on average, whereas strains with POT-2(68F) allele have normal

252 length telomeres on average. Synonymous variants or variation outside of the OB-fold  
253 domain were rarely found in strains with long telomeres. Because loss of *pot-2* is known  
254 to cause long telomeres (Raices *et al.* 2008; Cheng *et al.* 2012; Shtessel *et al.* 2013),  
255 the F68I variant likely reduces or eliminates the function of *pot-2*. Additionally, six out of  
256 the 39 long-telomere MMP strains had mutations in *pot-2*, including five strains that had  
257 mutations within or directly adjacent to the OB-fold and an additional strain with a  
258 nonsense mutation outside the OB-fold domain that likely destabilizes the transcript.  
259 These data support the hypothesis that *pot-2* is the causal gene underlying variation in  
260 telomere lengths across the *C. elegans* species.

261

### 262 **Natural variants in *pot-2* do not have detectable fitness consequences**

263

264 We connected genetic variation in the gene *pot-2* with telomere-length differences  
265 across *C. elegans* wild strains. Specifically, an F68I variant in the putative telomere-  
266 binding OB-fold domain might cause reduction of function and long telomeres. A variety  
267 of studies have observed a relationship between telomere length and organismal  
268 fitness, including longevity or cellular senescence (Harley *et al.* 1992; Heidinger *et al.*  
269 2012; Soerensen *et al.* 2012). Our results with natural variation in telomere lengths  
270 provided a unique opportunity to connect differences in the length of telomeres with  
271 effects on organismal fitness. We measured offspring production for our collection of  
272 152 wild strains and found no correlation with telomere length ( $\rho=0.062$ , Figure 6A).  
273 Long telomeres allow for increased replicative potential of cells (Harley *et al.* 1992), but  
274 it is unclear how the replicative potential of individual cells contributes to organismal  
275 longevity phenotypes (Hornsby 2007). We chose nine strains covering the range of  
276 telomere-length differences and found no correlation with longevity ( $\rho=-0.008$ , Figure  
277 6B, Supplementary Figure 7). Taken together, these results suggest that the long  
278 telomeres found in some wild *C. elegans* strains do not have significant fitness  
279 consequences in these laboratory-based experiments.

280

281 Because we did not observe a strong effect on organismal fitness, we investigated the  
282 population genetics of *pot-2* to test whether that locus had any signature of selection.  
283 Examination of Tajima's D at the *pot-2* locus yielded no conspicuous signature, though  
284 the characteristic high linkage disequilibrium of *C. elegans* makes gene-focused tests  
285 challenging in this species (Supplementary Figure 8). Furthermore, the haplotypes that  
286 contain this variant are rare (Supplementary Figure 9) and not geographically restricted  
287 (Supplementary Figure 10) Like the measurements of organismal fitness and lack of  
288 correlation with telomere-length differences, the population genetic measures for  
289 selection indicate that the observed variation in *pot-2* is not under strong selective  
290 pressure. Together, these results suggest that natural variation in telomere length plays  
291 a limited role in modifying whole-organism phenotypes in *C. elegans*.

292

## 293 **Discussion**

294

295 In this study, we report the identification of a QTL on the right arm of chromosome II  
296 containing a variant within the gene *pot-2* that contributes to differences in telomere  
297 length of *C. elegans* wild isolates. Several lines of evidence support the F68I allele of  
298 *pot-2* as the variant modulating telomere lengths. First, others have shown previously  
299 that loss of *pot-2* results in progressive telomere lengthening in the laboratory strain  
300 background (Raices *et al.* 2008; Shtessel *et al.* 2013). Additionally, the F68I variant is  
301 the only SNV in *pot-2* that correlates with long telomeres. This variant falls within the  
302 OB-fold of POT-2, and our examination of strain telomere lengths within the MMP  
303 shows enrichment of mutations from long-telomere strains found within the OB-fold  
304 domain. Importantly, OB-folds are known to interact with single-stranded nucleic acids,  
305 and TelSeq telomere-length estimates of wild isolates and randomly mutagenized  
306 laboratory strains show that mutation or variation of the OB-fold domain reduces  
307 function and causes long telomeres, as we also observed in the *pot-2(tm1400)* deletion  
308 strain. Moreover, this amino acid change could plausibly alter the function of the OB-fold  
309 within POT-2. Nucleic acid recognition of OB-folds occurs through a variety of molecular

310 interactions, including aromatic stacking (Gatzeva-topalova *et al.* 2011). A change from  
311 phenylalanine to isoleucine would eliminate a potential aromatic stacking interaction and  
312 presumably reduce the binding affinity and function of POT-2. Somatic mutations in  
313 individuals with chronic lymphocytic leukemia were found to be concentrated within the  
314 OB-folds of hPOT1 (Ramsay *et al.* 2013).

315  
316 We wondered why additional genes involved in the regulation of telomeres were not  
317 identified from our study of telomere lengths across wild isolates and mutagenized  
318 laboratory strains. Homologs for both telomerase and shelterin complex components  
319 are found in *C. elegans* (Stein *et al.* 2001). We identified natural variation in *trt-1* but  
320 only in the highly diverged strains ECA36 and QX1211. These rare alleles are removed  
321 from the GWA mapping, because we require allele frequencies to be greater than 5%.  
322 Laboratory mutants in *trt-1* have short telomeres (Cheung *et al.* 2006; Meier *et al.* 2006),  
323 but we do not see enrichment of *trt-1* mutations in the MMP collection for short or long  
324 telomeres. *C. elegans* contains orthologous genes for two of the six shelterin complex  
325 members, hPOT1 and RAP1 (Harris *et al.* 2009). Four *C. elegans* genes with homology  
326 to hPOT1 have been identified (*mrt-1*, *pot-1*, *pot-2*, and *pot-3*) (Raices *et al.* 2008; Meier  
327 *et al.* 2009), and *C. elegans rap-1* is homologous to human RAP1 (Raices *et al.* 2008;  
328 Meier *et al.* 2009). The genes *rap-1* and *pot-3* had no variants or only rare variants,  
329 respectively. All of the other homologous genes contained variants in 5% or more of the  
330 wild isolates. None of these genes mapped by GWA besides *pot-2*, and none of the  
331 mutations in these genes were enriched in short- or long-telomere strains from the MMP  
332 collection. Perhaps shorter telomere strains are less fit and do not survive well in the  
333 wild or during the growth of mutant MMP strains. These results suggest that long  
334 telomeres are likely of limited consequence compared to short telomeres in natural  
335 settings. Additionally, because TelSeq provides an average estimate of telomere length,  
336 it is possible for the variance of telomere lengths to increase without affecting average  
337 length estimates. For this reason, we might not detect a QTL at *pot-1*, which has been  
338 previously reported to result in longer but more heterogeneous telomeres (Raices *et al.*  
339 2008).

340  
341 Our observation that considerable telomere-length variation in the wild isolate  
342 population exists allowed us to directly test whether variation in telomere length  
343 contributes to organismal fitness. We did not see any correlation between telomere  
344 length and offspring production, suggesting that fitness in wild strains is not related to  
345 telomere length. In contrast to findings in human studies, we did not identify a  
346 relationship between telomere length and longevity. Our results confirm past findings  
347 that telomere length is not associated with longevity in a small number of *C. elegans*  
348 wild isolates or laboratory mutants (Raices *et al.* 2005). Although the effects of telomere  
349 length on longevity have been observed in a well controlled study of the gene *hrp-1* on  
350 isogenic populations in the laboratory (Joeng *et al.* 2004), this study differs from our  
351 results in wild isolates. The background effects of wild isolate variation along with  
352 telomere-length variation could obfuscate a direct connection to longevity. Even though  
353 we did not identify a correlation between telomere length and either longevity or  
354 offspring production under laboratory conditions, our study suggests a limited role for  
355 telomeres in post-mitotic cells. Furthermore, the population genetic results do not  
356 strongly support evidence of selection on *pot-2* variants.

357  
358 In summary, this study demonstrates that a variant in *pot-2* likely contributes to  
359 phenotypic differences in telomere length among wild isolates of *C. elegans*. The  
360 absence of evidence for selection at the *pot-2* locus and the lack of strong effects on  
361 organismal fitness traits suggest that differences in telomere length do not substantially  
362 affect individuals at least under laboratory growth conditions. Additionally, our study  
363 demonstrates the ability to extract and to utilize phenotypic information from sequence  
364 data. A number of approaches can be employed to examine other dynamic components  
365 of the genome, including mitochondrial and ribosomal DNA copy numbers, the  
366 mutational spectrum, or codon biases. These traits present a unique opportunity to  
367 identify how genomes differ among individuals and the genetic variants underlying those  
368 differences.

369



## 370 **Materials and Methods**

371

### 372 **Strains**

373

374 *C. elegans* strains were cultured using bacterial strain OP50 on a modified nematode  
375 growth medium (NGMA, 1% agar, 0.7% agarose) to prevent burrowing of wild isolates  
376 (Andersen *et al.* 2014). Strain information is listed in Supplementary File 1. The  
377 following strains were scored for the molecular telomere assays described below: AB4  
378 (CB4858 isotype), CB4856, CX11285, CX11292, DL238, ECA248, ED3012, EG4349,  
379 JT11398, JU311, JU1400, JU2007, KR314, N2, NIC2, NIC3, NIC207, PB303, and  
380 QX1212.

381

### 382 **Library construction and sequence acquisition**

383

384 DNA was isolated from 100-300  $\mu$ l of packed animals using the Blood and Tissue DNA  
385 isolation kit (Qiagen). The provided protocol was followed with the addition of RNase (4  
386  $\mu$ l of 100 mg/ml) following the initial lysis for two minutes at room temperature. DNA  
387 concentration was determined using the Qubit dsDNA BR Assay Kit (Invitrogen).  
388 Libraries were generated using the Illumina Nextera Sample Prep Kit and indexed using  
389 the Nextera Index Kit. Twenty-four uniquely indexed samples were pooled by mixing  
390 100 ng of each sample. The pooled material was size-selected by electrophoresing the  
391 DNA on a 2% agarose gel and excising the fragments ranging from 300-500 bp. The  
392 sample was purified using the Qiagen MinElute Kit and eluted in 11  $\mu$ l of buffer EB. The  
393 concentration of the purified sample was determined using the Qubit dsDNA HS Assay  
394 Kit. Sequencing was performed on the Illumina HiSeq 2500 platform. To increase  
395 coverage of some strains, we incorporated data from two separate studies of wild  
396 strains (Thompson *et al.* 2013; Noble *et al.* 2015).

397

## 398 **Trimming and demultiplexing**

399

400 When necessary, demultiplexing and sequence trimming were performed using  
401 `fastx_barcode_splitter.pl` (version 0.0.14) (Gordon and Hannon 2010). Sequences were  
402 trimmed using `trimmomatic` (version 0.32) (Bolger *et al.* 2014). Nextera libraries were  
403 trimmed using the following parameters:

404

405 `NexteraPE-PE.fa:2:80:10 MINLEN:45`

406

407 TruSeq libraries were trimmed using:

408

409 `TruSeq2-PE.fa:2:80:10 TRAILING:30 SLIDINGWINDOW:4:30 MINLEN:30`

410

411 The full details of the preparation, source, and library are available in Supplementary  
412 File 6.

413

## 414 **Alignment, variant calling, and filtering**

415

416 FASTQ sequence data has been deposited under NCBI Bioproject accession  
417 PRJNA318647. Sequences were aligned to WS245 (<http://www.wormbase.org>) using  
418 BWA (version 0.7.8-r455) (Li and Durbin 2009). Optical/PCR duplicates were marked  
419 with PICARD (version 1.111). BAM and CRAM files are available at  
420 [www.elegansvariation.org/Data](http://www.elegansvariation.org/Data). To determine which type of SNV caller would perform  
421 best on our dataset and to set appropriate filters, we simulated variation in the N2  
422 background. We used `bamsurgeon` ([github.com/adamewing/bamsurgeon](https://github.com/adamewing/bamsurgeon)), which  
423 modifies base calls to simulate variants at specific positions within aligned reads and  
424 then realigns reads to the reference genome using BWA. We simulated 100,000 SNVs  
425 in 10 independent simulation sets. Of the 100,000 sites chosen in each simulation set,  
426 `bamsurgeon` successfully inserted an average of 95,172.6 SNVs. Using these 10  
427 simulated variant sets, we tested two different methods of grouping our strains for

428 variant calling: calling strains individually (comparing sequences from a single strain to  
429 the reference) or calling strains jointly (comparing all strains in a population to each  
430 other). After grouping, bcftools has two different calling methods: a consensus caller  
431 (specified using -c), and a more recently developed multiallelic caller (specified using -  
432 m) (Li 2011). We performed variant calling using all four combinations of individual/joint  
433 calling and the consensus/multiallelic parameters. Because of the hermaphroditic life  
434 cycle of *C. elegans*, heterozygosity rates are likely low. Occasionally, heterozygous  
435 variants will be called despite skewed read support for reference or alternative alleles.  
436 To account for these likely erroneous calls, we performed 'heterozygous polarization'  
437 using the log-likelihood ratios of reference to alternative genotype calls. When the log-  
438 likelihood ratio was less than -2 or greater than 2, heterozygous genotypes were  
439 polarized (or switched) to reference genotypes or alternative genotypes, respectively. All  
440 other SNVs with likelihood ratios between -2 and 2 were called NA. Following variant  
441 calling and heterozygous polarization on resulting calls, we observed increased rates of  
442 heterozygous calls using joint methods and decreased true positive rates using our  
443 simulation data set (Supplementary File 7, Supplementary File 8). Given *C. elegans*  
444 predominantly asexual mode of reproduction, we decided to focus on the individual-  
445 based calling method that performed better. Next, we determined the optimal filters to  
446 maximize true positive (TP) rates and minimize false positive (FP) and false negative  
447 (FN) results using our simulated data (Supplementary File 8). After implementing  
448 different combinations of filters, we found that depth (DP), mapping quality (MQ), variant  
449 quality (QUAL), and the ratio of high quality alternative base calls (DV) over DP filters  
450 worked well (Supplementary Figure 11). Variants with  $DP \leq 10$ ,  $MQ \leq 40$ ,  $QUAL < 30$ ,  
451 and  $DV/DP < 0.5$  were called NA. Using these filters, we called 1.3M SNVs across 152  
452 isotypes. This data set is available at [www.andersenlab.org/Research/Data/Cooketal](http://www.andersenlab.org/Research/Data/Cooketal).

453

#### 454 **Validation of SNV calling methods**

455

456 In addition to performing simulations to optimize SNV-calling filters, we compared our  
457 whole-genome sequence variant calls with SNVs identified previously in CB4856 (Wicks

458 *et al.* 2001). Out of 4,256 sites we were able to call in regions that were sequenced  
459 using Sanger sequencing, we correctly identified 4,223 variants (99.2% of all variants) in  
460 CB4856. One true positive was erroneously filtered and two false positives were  
461 removed using our filters, and we failed to call the non-reference allele for 30 variants  
462 (false negatives).

463 Additionally, we examined sequence variants with poor parameters in terms of depth,  
464 quality, heterozygosity, or modification by our heterozygous polarization filter. We used  
465 primer3 (Rozen and Skaletsky 1998) to generate a pair of primers for performing PCR  
466 and a single forward primer for Sanger sequencing. We successfully sequenced 73 of  
467 95 sites chosen from several strains. Comparison of variant calls after imputation and  
468 filtering yielded 46 true positives (TP) and 14 true negatives (TN). We successfully  
469 removed 3/11 false positives (FP) and erroneously filtered two sites that should have  
470 been called as non-reference (FN). We also validated the variant responsible for the  
471 F68I change in JT11398.

472

### 473 **Identification of clonal sets**

474

475 Some strains in our original collection were isolated from the same or nearly identical  
476 locations. Therefore, we determined if these strains share distinct genome-wide  
477 haplotypes or isotypes. To determine strain relatedness, we sequenced and called  
478 variants from sequencing runs independently (*e.g.* individual FASTQ pairs) to ensure  
479 that strains were properly labeled before and after sequencing. We then combined  
480 FASTQ files from sequencing runs for a given strain and examined the concordance  
481 among genotypes. Comparison of variants identified among sequencing strains were  
482 used to determine whether the strains carried identical haplotypes. We observed that  
483 some strains were highly related to each other as compared with the rest of the  
484 population. Strains that were greater than 99.93% identical across 1,589,559 sites were  
485 classified as isotypes (Supplementary Figure 12). Because LSJ1 and N2 share a  
486 genome-wide genotype but exhibit distinct phenotypes (Sterken *et al.* 2015), we treated  
487 each strain as a separate isotype. We found the following isotype differences from the

488 previous characterization of a large number of strains. JU360 and JU363 were  
489 previously thought to be separate, but highly related, isotypes. We found that, at the  
490 genome-wide level and at high depths of coverage, these strains are from the same  
491 isotype. Several wild strains isolated before 2000 had different genome-wide haplotypes  
492 compared to strains with the same names but stored at the *Caenorhabditis* Genetics  
493 Center (CGC). CB4851 from the CGC had a different genome-wide haplotype compared  
494 to a strain with the same name from Cambridge, UK. We renamed the CB4851 strain  
495 from the CGC as ECA243. By contrast, the version from Cambridge, UK was nearly  
496 identical to N2 and not studied further. CB4855 from the CGC has a genome-wide  
497 haplotype that matches CB4858, which has a different history and isolation location.  
498 Therefore, we cannot guarantee the fidelity of this strain, and it was not studied further.  
499 CB4855 from Cambridge, UK is different from the CGC version of CB4855. We gave  
500 this strain the name ECA248 to avoid confusion. CB4858 from CGC has a different  
501 genome-wide haplotype than CB4858 from Cambridge, UK. Therefore, we renamed  
502 CB4858 from Cambridge, UK to ECA252, and it is a separate isotype. The CB4858 from  
503 the CGC was renamed ECA251 and is the reference strain from the CB4858 isotype.

504

### 505 **Imputation and variant annotation**

506

507 Following SNV calling and filtering, some variant sites were filtered. Therefore, we  
508 generated an imputed SNV set using beagle (version r1399) (Browning and Browning  
509 2016). This imputed variant set is available at  
510 [www.andersenlab.org/Research/Data/Cooketal](http://www.andersenlab.org/Research/Data/Cooketal). We used SnpEff (version 4.1g)  
511 (Cingolani *et al.* 2012) on this SNV set to predict functional effects.

512

### 513 **Telomere-length estimation**

514

515 Telomere lengths were estimated using TelSeq (version 0.0.1) (Ding *et al.* 2014) on  
516 BAM files derived from wild isolates or MMP strain sequencing. To estimate telomere  
517 lengths, TelSeq determines the reads that contain greater than seven telomeric

518 hexamer repeats (TTAGGC for *C. elegans*). Compared to most hexamers, telomeric  
519 hexamers can be found tandemly repeated within sequenced reads. TelSeq calculates  
520 the relative proportion of reads that appear to be telomerically derived among all  
521 sequenced reads and transforms this value into a length estimate using the formula  
522  $l = t_k s c$  where  $l$  is the length estimate and  $t_k$  is the abundance of reads with a  
523 minimum of  $k$  telomeric repeats. The value of  $s$  is the fraction of all reads with a GC  
524 composition similar to the telomeric repeat (48-52% for *C. elegans*). The value of  $c$  is a  
525 constant representing the length of 100 bp windows within the reference genome where  
526 GC content is equal to the GC content of the telomeric repeat divided by the total  
527 number of telomere ends. By default, TelSeq provides length estimates applicable to  
528 humans. We found the number of 100 bp windows with a 50% GC content in the  
529 WS245 reference genome to be 58,087. We calculated  $c$  for *C. elegans* as  $58,087 \text{ kb} /$   
530  $12 \text{ telomeres} = 484$ . This value was used to transform human length estimates to  
531 length estimates appropriate to *C. elegans*.

532  
533 Notably, telomere length estimates are averaged across all chromosomes, as no  
534 specific data about any one particular telomere is determined. To assess how well the  
535 TTAGGC hexamer distinguishes telomeric reads from non-telomeric reads, we  
536 examined the frequencies of non-cyclical permutations of the *C. elegans* hexamer in the  
537 N2 laboratory strain using TelSeq (Supplementary Figure 13). We observe that the  
538 majority of hexamers examined were not present in more than six copies in a high  
539 frequency of reads. By contrast, the reads possessing the telomeric hexamer with seven  
540 copies or more were more abundant than any other hexamer. Tandem repeats of the  
541 telomeric hexamer are present within the reference genome at the ends of each  
542 chromosome and occasionally internally within chromosomes at between 2-71 copies  
543 (Supplementary File 9). After running TelSeq on our wild isolates, we removed eight  
544 sequencing runs (out of 868 total) that possessed zero reads with 15 or more copies of  
545 the telomeric hexamer. These sequencing runs provide additional support for SNV  
546 calling but had short read lengths that would underestimate telomere length. We used  
547 the weighted average of telomere length estimates for all runs of a given strain based

548 on total reads to calculate telomere-length estimates. Supplementary File 10 details  
549 telomere length estimates for every sequencing run.

550

### 551 **Quantitative PCR assays for telomere-length measurements**

552

553 Telomere lengths were measured by qPCR as described previously with some  
554 modifications (Cawthon 2009). Primer sequences were modified from the vertebrate  
555 telomere repeat (TTAGGG) to use the *C. elegans* telomere repeat (TTAGGC):

556

557 telG: 5'-ACACTAAGCTTTGGCTTTGGCTTTGGCTTTGGCTTAGTCT-3'

558 telC: 5'-TGTTAGGTATGCCTATGCCTATGCCTATGCCTATGCCTATGCCTAAGA-3'

559

560 The internal control, *act-1*, was amplified using the following primer pair:

561

562 forward: 5'-GTCGGTATGGGACAGAAGGA-3'

563 reverse: 5'-GCTTCAGTGAGGAGGACTGG-3'

564

565 Two primer pairs were amplified separately (singleplex qPCR). All the samples were run  
566 in triplicate. qPCR was performed using iQ SYBR green supermix (BIORAD) with  
567 iCycler iQ real-time PCR detection system (BIORAD). After thermal cycling,  $C_t$  (cycle  
568 thresholds) values were exported from BIORAD iQ5 software.

569

### 570 **Terminal restriction fragment (TRF) Southern blot assay**

571

572 Animals were grown on 100 mm Petri dishes with NGM seeded with OP50.  
573 Synchronized adult animals were harvested and washed four times with M9 buffer.  
574 Pelleted animals were lysed for four hours at 50°C in buffer containing 0.1 M Tris-Cl (pH  
575 8.5), 0.1 M NaCl, 50 mM EDTA (pH 8.0), 1% SDS, and 0.1 mg/mL proteinase K. DNA  
576 was isolated by phenol extraction and ethanol precipitation. DNA was eluted with buffer  
577 containing 10 mM Tris (pH 7.5) and 1 mM EDTA. DNA was then treated with 10  $\mu$ g/mL

578 boiled RNase A. DNA was again isolated with phenol extraction and ethanol  
579 precipitation. Five  $\mu\text{g}$  of DNA was digested with *Hinf*I at 37°C overnight. Telomere  
580 restriction fragment was blotted as described previously (Seo *et al.* 2015)  
581 (Supplementary Figure 14). Digoxigenin-labeled (TTAGGC)<sub>4</sub> oligonucleotides were used  
582 as probes. Digoxigenin probes were detected with DIG nucleic acid detection kit  
583 (Roche). Blots were imaged with ImageQuant LAS4000 (GE healthcare).

584

### 585 **Fluorescence *in situ* hybridization assays**

586

587 Fluorescence *in situ* hybridization (FISH) was performed as previously described (Seo  
588 *et al.* 2015). Embryos were isolated by bleaching synchronized adult animals using  
589 standard methods (Stiernagle 2006). Isolated embryos were fixed in 2%  
590 paraformaldehyde (PFA) for 15 minutes at room temperature (RT) on a polylysine  
591 treated glass slide. The slide was put on dry ice and freeze-cracked. The embryos were  
592 permeabilized in ice-cold methanol and acetone for 5 minutes each. The slides were  
593 washed with 1X phosphate buffered saline containing 0.1% Tween-20 (PBST) three  
594 times for 15 minutes each at room temperature. 10  $\mu\text{l}$  of hybridization buffer (50 nM  
595 Cy3-(TTAGGC)<sub>3</sub> peptide nucleic acids probe (PANAGENE), 50% formamide, 0.45 M  
596 sodium chloride, 45 mM sodium citrate, 10% dextran sulfate, 50  $\mu\text{g}/\text{mL}$  heparin,  
597 100  $\mu\text{g}/\text{mL}$  yeast tRNA, 100  $\mu\text{g}/\text{mL}$  salmon sperm DNA) was added on the slide. The  
598 samples were denatured on a heat block at 85°C for three minutes. After overnight  
599 incubation at 37°C, the samples were washed in the following order: 1X PBST once for  
600 five minutes at room temperature, 2X SSC (0.3 M sodium chloride, 30 mM sodium  
601 citrate) in 50% formamide once for 30 minutes at 37°C, 1X PBST three times for 10  
602 minutes each at room temperature. The samples were incubated in DAPI and mounted  
603 in anti-bleaching solution (Vectashield). The samples were imaged with a confocal  
604 microscope (LSM700, Zeiss). Telomere spots were quantified with TFL-TELO software  
605 (Dr. Peter Lansdorp, Terry Fox Laboratory, Vancouver) (Poon SS1, Martens UM, Ward  
606 RK 1999).



607

## 608 **Genome-wide association (GWA) mapping**

609

610 GWA mapping was performed on marker genotype data and telomere-length estimates  
611 using the rrBLUP package (version 4.3) (Endelman 2011) and GWAS function. rrBLUP  
612 requires a kinship matrix and a SNV set to perform GWA. We generated a kinship  
613 matrix using our imputed SNV set with the A.mat function within rrBLUP. Genomic  
614 regions of interest were determined empirically from simulating a QTL that explained  
615 20% of the phenotypic variance at each marker in our mapping data set. All simulated  
616 QTL were mapped within 100 markers (50 markers to the left and 50 markers to the  
617 right) of the simulated marker position. To generate a SNV set for mapping, we again  
618 used our imputed SNV set. However, we filtered the number of SNVs to a set of 38,688  
619 markers. This set was generated by lifting over (from WS210 to WS245) a set of 41,888  
620 SNVs previously used for GWA mapping (Andersen *et al.* 2012) and filtering our  
621 imputed SNVs to those sites.

622

## 623 **Million Mutation Project analysis**

624

625 Whole-genome sequence data from mutagenized strains within the Million Mutation  
626 Project (MMP) was obtained from the sequence read archive (SRA, project accession  
627 number SRP018046). We removed 59 strains that were contaminated with other strains.  
628 We also were unable to locate the sequence data for 12 MMP strains on SRA, leaving  
629 us with a total of 1,936 mutagenized strains. Within the MMP project, read lengths  
630 varied among sequencing runs, being either 75 bp or 100 bp. We ran TelSeq on all  
631 sequencing runs assuming 100 bp reads. To utilize 75 bp sequencing runs, we took the  
632 448 strains that were sequenced at both 75 and 100 bp lengths and used those  
633 estimates to develop a linear model. Then, this model was used to transform 75 bp  
634 length estimates to 100 bp estimates (Supplementary Figure 15). We then used the  
635 weighted average of telomere length estimates for all runs of a given strain based on  
636 total reads to calculate telomere-length estimates. Because telomeric reads resemble

637 PCR duplicates, TelSeq utilizes them in calculating telomere length. However, we  
638 observed very low PCR and optical duplicate rates among MMP sequence data likely  
639 due to differences in library preparation in contrast to wild isolate sequence data. These  
640 differences likely account for shorter telomere estimates from the MMP sequence data.

641  
642 Long-telomere strains from the MMP were classified as strains with telomere lengths  
643 greater than the 98<sup>th</sup> quantile of all MMP strains (6.41 kb). Mutation data was obtained  
644 from the MMP website ([http://genome.sfu.ca/mmp/mmp\\_mut\\_strains\\_data\\_Mar14.txt](http://genome.sfu.ca/mmp/mmp_mut_strains_data_Mar14.txt)).  
645 A hypergeometric test was performed to identify which genes were enriched for  
646 mutations from long-telomere strains (Supplementary File 11) using the *phyper* function  
647 in R (R Core Team 2013). FX1400 was propagated for ten generations prior to whole-  
648 genome sequencing. Telomere length was estimated using TelSeq.

649

## 650 **Statistical analyses**

651

652 Statistical analyses were performed using R (version 3.2.3). Plots were produced using  
653 ggplot2 (version 2.0.0).

654

## 655 **Longevity assays**

656

657 At least 80 fourth larval stage animals were plated onto each of three separate 6 cm  
658 NGMA plates in two independent assays and viability assessed each day until all  
659 animals were scored as dead or censored from the analysis as a result of bagging or  
660 missing animals. Animals were scored as dead in the absence of touch response and  
661 pharyngeal pumping. Animals were transferred to fresh plates every day from the  
662 initiation of the assay until day seven of adulthood to remove progeny and transferred  
663 every other day until the completion of the assay. The following short telomere strains  
664 were scored: EG4349, JU2007, NIC1, and NIC3. The following long-telomere strains  
665 were scored: KR314, NIC207, QX1212, and RC301. Additionally, N2 and CB4856 were  
666 scored.

667

## 668 **High-throughput fecundity assays**

669

670 Assays were performed similar to previously reported (Andersen *et al.* 2015) with the  
671 following differences. Animals were bleached, synchronized, and grown to L4 larvae in  
672 96-well plates. From the L1 to L4 stage, animals were fed 5 mg/mL of a large-scale  
673 production HB101 lysate in K medium (Boyd *et al.* 2010) to provide a stereotyped and  
674 constant food source. Then, three L4 larvae from each of the 152 genotypes were  
675 dispensed using a COPAS BIOSORT instrument to wells containing 10 mg/mL HB101  
676 lysate in K medium and progeny were counted 96 hours later. Fecundity data were  
677 calculated using 12 samples – triplicate technical replicates from four biological  
678 replicates. The data were processed using COPASutils (Shimko and Andersen 2014)  
679 and statistically analyzed using custom R scripts.

680

## 681 **Clustering of relatedness**

682

683 Variant data for dendrogram comparisons were assembled by constructing a FASTA file  
684 with the genome-wide variant positions across all strains and subsetting by regions as  
685 described. MUSCLE (version v3.8.31) (Edgar 2004) was used to generate neighbor-  
686 joining trees. The R packages ape (version 3.4) (Paradis *et al.* 2004) and phyloseq  
687 (version 1.12.2) (McMurdie and Holmes 2013) were used for data processing and  
688 plotting.

689

## 690 **Acknowledgements**

691

692 We would like to thank Joshua Bloom and members of the Andersen laboratory for  
693 critical comments on this manuscript. We also thank M. Barkoulas, T. BÉlicard, D.  
694 Bourc'his, N. Callemeyn-Torre, S. Carvalho, J. Dumont, L. Frézal, C.-Y. Kao, L.  
695 Lokmane, I. Ly, K. Ly, A. Paaby, J. Riksen, G. Wang for isolating new wild *C. elegans*

696 strains. The National Bioresource Project provided the FX1400 strain, and Wormbase  
697 data made a variety of analyses possible. This work was supported by an NIH R01  
698 subcontract to E.C.A. (GM107227), the Chicago Biomedical Consortium with support  
699 from the Searle Funds at the Chicago Community Trust, and an American Cancer  
700 Society Research Scholar Grant to E.C.A. (127313-RSG-15-135-01-DD), along with  
701 support from the National Science Foundation Graduate Research Fellowship to D.E.C.  
702 (DGE-1324585).

## 703 **References**

- 704
- 705 Andersen E. C., Gerke J. P., Shapiro J. A., Crissman J. R., Ghosh R., Bloom J. S., Félix  
706 M.-A., Kruglyak L., 2012 Chromosome-scale selective sweeps shape  
707 *Caenorhabditis elegans* genomic diversity. *Nat. Genet.* **44**: 285–290.
- 708 Andersen E. C., Bloom J. S., Gerke J. P., Kruglyak L., 2014 A Variant in the  
709 Neuropeptide Receptor *npr-1* is a Major Determinant of *Caenorhabditis elegans*  
710 Growth and Physiology. *PLoS Genet.* **10**.
- 711 Andersen E. C., Shimko T. C., Crissman J. R., Ghosh R., Bloom J. S., Seidel H. S.,  
712 Gerke J. P., Kruglyak L., 2015 A Powerful New Quantitative Genetics Platform,  
713 Combining *Caenorhabditis elegans* High-Throughput Fitness Assays with a Large  
714 Collection of Recombinant Strains. *G3 (Bethesda)*. **5**: 911–20.
- 715 Armanios M., Chen J.-L., Chang Y.-P. C., Brodsky R. a, Hawkins A., Griffin C. a,  
716 Eshleman J. R., Cohen A. R., Chakravarti A., Hamosh A., Greider C. W., 2005  
717 Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in  
718 autosomal dominant dyskeratosis congenita. *Proc. Natl. Acad. Sci. U. S. A.* **102**:  
719 15960–15964.
- 720 Barrière A., Félix M.-A., 2005 *Natural variation and population genetics of*  
721 *Caenorhabditis elegans*. *WormBook (2005)*. doi:10.1895/wormbook.1.43.1.
- 722 Blackburn E. H., 1991 Structure and function of telomeres. *Nature* **350**: 569–573.
- 723 Bolger A. M., Lohse M., Usadel B., 2014 Trimmomatic: A flexible trimmer for Illumina  
724 sequence data. *Bioinformatics* **30**: 2114–2120.
- 725 Boyd W. A., McBride S. J., Rice J. R., Snyder D. W., Freedman J. H., 2010 A high-  
726 throughput method for assessing chemical toxicity using a *Caenorhabditis elegans*  
727 reproduction assay. *Toxicol. Appl. Pharmacol.* **245**: 153–159.
- 728 Broer L., Codd V., Nyholt D. R., Deelen J., Mangino M., Willemsen G., Albrecht E., Amin  
729 N., Beekman M., Geus E. J. C. de, Henders A., Nelson C. P., Steves C. J., Wright  
730 M. J., Craen A. J. M. de, Isaacs A., Matthews M., Moayyeri A., Montgomery G. W.,  
731 Oostra B. A., Vink J. M., Spector T. D., Slagboom P. E., Martin N. G., Samani N. J.,  
732 Duijn C. M. van, Boomsma D. I., 2013 Meta-analysis of telomere length in 19,713

- 733 subjects reveals high heritability, stronger maternal inheritance and a paternal age  
734 effect. *Eur. J. Hum. Genet.* **21**: 1163–8.
- 735 Browning B. L., Browning S. R., 2016 Genotype Imputation with Millions of Reference  
736 Samples. *Am. J. Hum. Genet.* **98**: 116–126.
- 737 Cawthon R. M., 2009 Telomere length measurement by a novel monochrome multiplex  
738 quantitative PCR method. *Nucleic Acids Res.* **37**: 1–7.
- 739 Cheng C., Shtessel L., Brady M. M., Ahmed S., 2012 *Caenorhabditis elegans* POT-2  
740 telomere protein represses a mode of alternative lengthening of telomeres with  
741 normal telomere lengths. *Proc. Natl. Acad. Sci.* **109**: 7805–7810.
- 742 Cheung I., Schertzer M., Baross A., Rose A. M., Lansdorp P. M., Baird D. M., 2004  
743 Strain-specific telomere length revealed by single telomere length analysis in  
744 *Caenorhabditis elegans*. *Nucleic Acids Res.* **32**: 3383–3391.
- 745 Cheung I., Schertzer M., Rose A., Lansdorp P. M., 2006 High incidence of rapid  
746 telomere loss in telomerase-deficient *Caenorhabditis elegans*. *Nucleic Acids Res.*  
747 **34**: 96–103.
- 748 Cingolani P., Platts A., Wang L. L. L., Coon M., Nguyen T., Wang L. L. L., Land S. J., Lu  
749 X., Ruden D. M., 2012 A program for annotating and predicting the effects of single  
750 nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila*  
751 *melanogaster* strain w 1118; iso-2; iso-3. *Fly (Austin)*. **6**: 80–92.
- 752 Codd V., Nelson C. C. P. C., Albrecht E., Mangino M., Deelen J., Buxton J. L., Hottenga  
753 J. J., Fischer K., Esko T., Surakka I., Broer L., Nyholt D. R., Mateo Leach I., Salo  
754 P., Hägg S., Matthews M. K., Palmen J., Norata G. D., O'Reilly P. F., Saleheen D.,  
755 Amin N., Balmforth A. J., Beekman M., Boer R. a de, Böhringer S., Braund P. S.,  
756 Burton P. R., Craen A. J. M. de, Denniff M., Dong Y., Douroudis K., Dubinina E.,  
757 Eriksson J. G., Garlaschelli K., Guo D., Hartikainen A.-L., Henders A. K., Houwing-  
758 Duistermaat J. J., Kananen L., Karssen L. C., Kettunen J., Klopp N., Lagou V.,  
759 Leeuwen E. M. van, Madden P. a, Mägi R., Magnusson P. K. E., Männistö S.,  
760 McCarthy M. I., Medland S. E., Mihailov E., Montgomery G. W., Oostra B. a, Palotie  
761 A., Peters A., Pollard H., Pouta A., Prokopenko I., Ripatti S., Salomaa V., Suchiman  
762 H. E. D., Valdes A. M., Verweij N., Viñuela A., Wang X., Wichmann H.-E., Widen

- 763 E., Willemsen G., Wright M. J., Xia K., Xiao X., Veldhuisen D. J. van, Catapano A.  
764 L., Tobin M. D., Hall A. S., Blakemore A. I. F., Gilst W. H. van, Zhu H., Consortium  
765 C., Erdmann J., Reilly M. P., Kathiresan S., Schunkert H., Talmud P. J., Pedersen  
766 N. L., Perola M., Ouwehand W., Kaprio J., Martin N. G., Duijn C. M. van, Hovatta I.,  
767 Gieger C., Metspalu A., Boomsma D. I., Jarvelin M.-R., Slagboom P. E., Thompson  
768 J. R., Spector T. D., Harst P. van der, Samani N. J., 2013 Identification of seven loci  
769 affecting mean telomere length and their association with disease. *Nat. ...* **45**:  
770 422–7, 427e1–2.
- 771 Deng Y., Chan S. S., Chang S., 2008 Telomere dysfunction and tumour suppression:  
772 the senescence connection. *Nat. Rev. Cancer* **8**: 450–458.
- 773 Ding Z., Mangino M., Aviv A., Spector T., Durbin R., 2014 Estimating telomere length  
774 from whole genome sequence data. *Nucleic Acids Res.* **42**: 1–4.
- 775 Edgar R. C., 2004 MUSCLE: Multiple sequence alignment with high accuracy and high  
776 throughput. *Nucleic Acids Res.* **32**: 1792–1797.
- 777 Endelman J. B., 2011 Ridge Regression and Other Kernels for Genomic Selection with  
778 R Package rrBLUP. *Plant Genome J.* **4**: 250–255.
- 779 Félix M.-A., Braendle C., 2010 The natural history of *Caenorhabditis elegans*. *Curr. Biol.*  
780 **20**: R965–R969.
- 781 Flynn R. L., Zou L., 2010 Oligonucleotide/oligosaccharide-binding fold proteins: a  
782 growing family of genome guardians. *Crit. Rev. Biochem. Mol. Biol.* **45**: 266–275.
- 783 Frenck Jr. R. W., Blackburn E. H., Shannon K. M., Frenck R. W., Blackburn E. H.,  
784 Shannon K. M., 1998 The rate of telomere sequence loss in human leukocytes  
785 varies with age. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 5607–10.
- 786 Fulcher N., Teubenbacher A., Kerdaffrec E., Farlow A., Nordborg M., Riha K., 2014  
787 Genetic Architecture of Natural Variation of Telomere Length in *Arabidopsis*  
788 *thaliana*. *Genetics* **199**: 625–635.
- 789 Gatbonton T., Imbesi M., Nelson M., Akey J. M., Ruderfer D. M., Kruglyak L., Simon J.  
790 A., Bedalov A., 2006 Telomere length as a quantitative trait: Genome-wide survey  
791 and genetic mapping of telomere length-control genes in yeast. *PLoS Genet.* **2**:  
792 0304–0315.

- 793 Gatzeva-topalova P. Z., Warner L. R., Pardi A., Carlos M., 2011 Nucleic Acid  
794 Recognition by Ob-Fold Proteins. **18**: 1492–1501.
- 795 Gordon A., Hannon G. J., 2010 Fastx-toolkit. FASTQ/A short-reads pre-processing  
796 tools. Unpubl. [http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit).
- 797 Griffith J. D., Comeau L., Rosenfield S., Stansel R. M., Bianchi A., Moss H., Lange T.  
798 De, 1999 Mammalian telomeres end in a large duplex loop. *Cell* **97**: 503–514.
- 799 Harley C. B., Vaziri H., Counter C. M., Allsopp R. C., 1992 The telomere hypothesis of  
800 cellular aging. *Exp. Gerontol.* **27**: 375–382.
- 801 Harris T. W., Antoshechkin I., Bieri T., Blasiar D., Chan J., Chen W. J., Cruz N. D. La,  
802 Davis P., Duesbury M., Fang R., Fernandes J., Han M., Kishore R., Lee R., Mu H.,  
803 Nakamura C., Ozersky P., Petcherski A., Rangarajan A., Rogers A., Schindelman  
804 G., Schwarz E. M., Tuli M. A., Auken K. Van, Wang D., Wang X., Williams G., Yook  
805 K., Durbin R., Stein L. D., Spieth J., Sternberg P. W., la Cruz N. De, Davis P.,  
806 Duesbury M., Fang R., Fernandes J., Han M., Kishore R., Lee R., M??ller H. M.,  
807 Nakamura C., Ozersky P., Petcherski A., Rangarajan A., Rogers A., Schindelman  
808 G., Schwarz E. M., Tuli M. A., Auken K. Van, Wang D., Wang X., Williams G., Yook  
809 K., Durbin R., Stein L. D., Spieth J., Sternberg P. W., 2009 Wormbase: A  
810 comprehensive resource for nematode research. *Nucleic Acids Res.* **38**: 463–467.
- 811 Heidinger B. J., Blount J. D., Boner W., Griffiths K., Metcalfe N. B., Monaghan P., 2012  
812 Telomere length in early life predicts lifespan. : 1–6.
- 813 Hornsby P. J., 2007 Telomerase and the aging process. *Exp. Gerontol.* **42**: 575–81.
- 814 Joeng K. S., Song E. J., Lee K.-J., Lee J., 2004 Long lifespan in worms with long  
815 telomeric DNA. *Nat. Genet.* **36**: 607–611.
- 816 Jones a. M., Beggs a. D., Carvajal-Carmona L., Farrington S., Tenesa a., Walker M.,  
817 Howarth K., Ballereau S., Hodgson S. V., Zuber a., Bertagnolli M., Midgley R.,  
818 Campbell H., Kerr D., Dunlop M. G., Tomlinson I. P. M., 2012 TERC  
819 polymorphisms are associated both with susceptibility to colorectal cancer and with  
820 longer telomeres. *Gut* **61**: 248–254.
- 821 Kwan E. X., Foss E., Kruglyak L., Bedalov A., 2011 Natural polymorphism in *bul2* links  
822 cellular amino acid availability with chronological aging and telomere maintenance



- 823 in yeast. *PLoS Genet.* **7**.
- 824 Lack J. B., Cardeno C. M., Crepeau M. W., Taylor W., Corbett-Detig R. B., Stevens K.  
825 a., Langley C. H., Pool J. E., 2015 The *Drosophila* Genome Nexus: A Population  
826 Genomic Resource of 623 *Drosophila melanogaster* Genomes, Including 197 from  
827 a Single Ancestral Range Population. *Genetics* **199**: 1229–1241.
- 828 Lange T. De, 2010 How shelterin solves the telomere end-protection problem. *Cold*  
829 *Spring Harb. Symp. Quant. Biol.* **75**: 167–177.
- 830 Lei M., Podell E. R., Cech T. R., 2004 Structure of human POT1 bound to telomeric  
831 single-stranded DNA provides a model for chromosome end-protection. *Nat. Struct.*  
832 *Mol. Biol.* **11**: 1223–1229.
- 833 Levy M. Z., Allsopp R. C., Futcher A. B., Greider C. W., Harley C. B., 1992 Telomere  
834 end-replication problem and cell aging. *J. Mol. Biol.* **225**: 951–960.
- 835 Levy D., Neuhausen S. L., Hunt S. C., Kimura M., Hwang S.-J., Chen W., Bis J. C.,  
836 Fitzpatrick A. L., Smith E., Johnson A. D., Gardner J. P., Srinivasan S. R., Schork  
837 N., Rotter J. I., Herbig U., Psaty B. M., Sastrasinh M., Murray S. S., Vasan R. S.,  
838 Province M. A., Glazer N. L., Lu X., Cao X., Kronmal R., Mangino M., Soranzo N.,  
839 Spector T. D., Berenson G. S., Aviv A., 2010 Genome-wide association identifies  
840 OBFC1 as a locus involved in human leukocyte telomere biology. *Proc. Natl. Acad.*  
841 *Sci. U. S. A.* **107**: 9293–8.
- 842 Li H., Durbin R., 2009 Fast and accurate short read alignment with Burrows-Wheeler  
843 transform. *Bioinformatics* **25**: 1754–1760.
- 844 Li H., 2011 A statistical framework for SNP calling, mutation discovery, association  
845 mapping and population genetical parameter estimation from sequencing data.  
846 *Bioinformatics* **27**: 2987–2993.
- 847 Liti G., Haricharan S., Cubillos F. a., Tierney A. L., Sharp S., Bertuch A. a., Parts L.,  
848 Bailes E., Louis E. J., 2009 Segregating YKU80 and TLC1 alleles underlying  
849 natural variation in telomere properties in wild yeast. *PLoS Genet.* **5**.
- 850 Mackay T. F. C., Richards S., Stone E. a., Barbadilla A., Ayroles J. F., Zhu D., Casillas  
851 S., Han Y., Magwire M. M., Cridland J. M., Richardson M. F., Anholt R. R. H.,  
852 Barrón M., Bess C., Blankenburg K. P., Carbone M. A., Castellano D., Chaboub L.,

- 853 Duncan L., Harris Z., Javid M., Jayaseelan J. C., Jhangiani S. N., Jordan K. W.,  
854 Lara F., Lawrence F., Lee S. L., Librado P., Linheiro R. S., Lyman R. F., Mackey A.  
855 J., Munidasa M., Muzny D. M., Nazareth L., Newsham I., Perales L., Pu L.-L., Qu  
856 C., Ràmia M., Reid J. G., Rollmann S. M., Rozas J., Saada N., Turlapati L., Worley  
857 K. C., Wu Y.-Q., Yamamoto A., Zhu Y., Bergman C. M., Thornton K. R., Mittelman  
858 D., Gibbs R. a., 2012 The Drosophila melanogaster Genetic Reference Panel.  
859 Nature **482**: 173–178.
- 860 Malik H. S., Burke W. D., Eickbush T. H., 2000 Putative telomerase catalytic subunits  
861 from Giardia lamblia and Caenorhabditis elegans. Gene **251**: 101–108.
- 862 Marrone A., Walne A., Dokal I., 2005 Dyskeratosis congenita: Telomerase, telomeres  
863 and anticipation. Curr. Opin. Genet. Dev. **15**: 249–257.
- 864 McCarthy M. I., Abecasis G. R., Cardon L. R., Goldstein D. B., Little J., Ioannidis J. P. a,  
865 Hirschhorn J. N., 2008 Genome-wide association studies for complex traits:  
866 consensus, uncertainty and challenges. Nat. Rev. Genet. **9**: 356–369.
- 867 McEachern M. J., Krauskopf A., Blackburn E. H., 2000 Telomeres and their control.  
868 Annu. Rev. Genet. **34**: 331–358.
- 869 McGrath P. T., Rockman M. V., Zimmer M., Jang H., Macosko E. Z., Kruglyak L.,  
870 Bargmann C. I., 2009 Quantitative Mapping of a Digenic Behavioral Trait Implicates  
871 Globin Variation in C. elegans Sensory Behaviors. Neuron **61**: 692–699.
- 872 McMurdie P. J., Holmes S., 2013 Phyloseq: An R Package for Reproducible Interactive  
873 Analysis and Graphics of Microbiome Census Data. PLoS One **8**.
- 874 Meier B., Clejan I., Liu Y., Lowden M., Gartner A., Hodgkin J., Ahmed S., 2006 trt-1 is  
875 the Caenorhabditis elegans catalytic subunit of telomerase. PLoS Genet. **2**: 187–  
876 197.
- 877 Meier B., Barber L. J., Shtessel L., Boulton S. J., Gartner A., Ahmed S., 2009 The MRT-  
878 1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in  
879 Caenorhabditis elegans. EMBO J. **28**: 3549–3563.
- 880 Noble L. M., Chang A. S., McNelis D., Kramer M., Yen M., Nicodemus J. P., Riccardi D.  
881 D., Ammerman P., Phillips M., Islam T., Rockman M. V., 2015 Natural Variation in  
882 plep-1 Causes Male-Male Copulatory Behavior in C. Elegans. Curr. Biol. **25**: 2730–

- 883 2737.
- 884 O'Sullivan R. J., Karlseder J., 2010 Telomeres: protecting chromosomes against  
885 genome instability. *Nat. Rev. Mol. Cell Biol.* **11**: 171–181.
- 886 Paradis E., Claude J., Strimmer K., 2004 APE: Analyses of phylogenetics and evolution  
887 in R language. *Bioinformatics* **20**: 289–290.
- 888 Poon SS1, Martens UM, Ward RK L. P., 1999 Telomere length measurements using  
889 digital fluorescence microscopy. *Cytometry*. **278**: 267–278.
- 890 Project G., Project G., Asia E., Africa S., Figs S., Tables S., 2012 An integrated map of  
891 genetic variation from 1,092 human genomes. *Nature* **135**: 0–9.
- 892 R Core Team, 2013 R Core Team. *R A Lang. Environ. Stat. Comput. R Found. Stat.*  
893 *Comput. Vienna, Austria.*: ISBN 3–900051–07–0, URL <http://www.R-project.org/>.
- 894 Raices M., Maruyama H., Dillin A., Karlseder J., 2005 Uncoupling of longevity and  
895 telomere length in *C. elegans*. *PLoS Genet.* **1**: 295–301.
- 896 Raices M., Verdun R. E., Compton S. A., Haggblom C. I., Griffith J. D., Dillin A.,  
897 Karlseder J., 2008 *C. elegans* Telomeres Contain G-Strand and C-Strand  
898 Overhangs that Are Bound by Distinct Proteins. *Cell* **132**: 745–757.
- 899 Ramsay A. J., Quesada V., Foronda M., Conde L., Martínez-Trillos A., Villamor N.,  
900 Rodríguez D., Kwarciak A., Garabaya C., Gallardo M., López-Guerra M., López-  
901 Guillermo A., Puente X. S., Blasco M. a, Campo E., López-Otín C., 2013 POT1  
902 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nat. Genet.*  
903 **45**: 526–30.
- 904 Rockman M. V., Kruglyak L., 2009 Recombinational landscape and population  
905 genomics of *caenorhabditis elegans*. *PLoS Genet.* **5**.
- 906 Rozen S., Skaletsky H. J., 1998 Primer3. *Bioinforma. Methods Protoc. Methods Mol.*  
907 *Biol.* **3**: 1–41.
- 908 Samassekou O., Gadji M., Drouin R., Yan J., 2010 Sizing the ends: Normal length of  
909 human telomeres. *Ann. Anat.* **192**: 284–291.
- 910 Seo B., Kim C., Hills M., Sung S., Kim H., Kim E., Lim D. S., Oh H.-S., Choi R. M. J.,  
911 Chun J., Shim J., Lee J., 2015 Telomere maintenance through recruitment of  
912 internal genomic regions. *Nat. Commun.* **6**: 8189.

- 913 Shimko T. C., Andersen E. C., 2014 COPASutils: An R Package for Reading,  
914 Processing, and Visualizing Data from COPAS Large-Particle Flow Cytometers.  
915 PLoS One **9**: e111090.
- 916 Shtessel L., Lowden M. R., Cheng C., Simon M., Wang K., Ahmed S., 2013  
917 *Caenorhabditis elegans* POT-1 and POT-2 repress telomere maintenance  
918 pathways. *G3 (Bethesda)*. **3**: 305–13.
- 919 Soerensen M., Thinggaard M., Nygaard M., Dato S., Tan Q., Hjelmberg J., Andersen-  
920 Ranberg K., Stevnsner T., Bohr V. A., Kimura M., Aviv A., Christensen K.,  
921 Christiansen L., 2012 Genetic variation in TERT and TERC and human leukocyte  
922 telomere length and longevity: A cross-sectional and longitudinal analysis. *Aging*  
923 *Cell* **11**: 223–227.
- 924 Stein L., Sternberg P., Durbin R., Thierry-Mieg J., Spieth J., 2001 WormBase: network  
925 access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.*  
926 **29**: 82–86.
- 927 Sterken M. G., Snoek L. B., Kammenga J. E., Andersen E. C., 2015 The laboratory  
928 domestication of *Caenorhabditis elegans*. *Trends Genet.*: 1–8.
- 929 Stiernagle T., 2006 Maintenance of *C. elegans*. *WormBook*: 1–11.
- 930 Thompson O., Edgley M., Strasbourger P., Flibotte S., Ewing B., Adair R., Au V.,  
931 Chaudhry I., Fernando L., Hutter H., Kieffer A., Lau J., Lee N., Miller A., Raymant  
932 G., Shen B., Shendure J., Taylor J., Turner E. H., Hillier L. W., Moerman D. G.,  
933 Waterston R. H., 2013 The million mutation project: A new approach to genetics in  
934 *Caenorhabditis elegans*. *Genome Res.* **23**: 1749–1762.
- 935 Vulliamy T., Marrone A., Szydlo R., Walne A., Mason P. J., Dokal I., 2004 Disease  
936 anticipation is associated with progressive telomere shortening in families with  
937 dyskeratosis congenita due to mutations in TERC. *Nat Genet* **36**: 447–449.
- 938 Watson J. D., 1972 Origin of concatemeric T7 DNA. *Nat. New Biol.* **239**: 197–201.
- 939 Weigel D., Mott R., 2009 The 1001 genomes project for *Arabidopsis thaliana*. *Genome*  
940 *Biol.* **10**: 107.
- 941 Wicks S. R., Yeh R. T., Gish W. R., Waterston R. H., Plasterk R. H., 2001 Rapid gene  
942 mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat.*

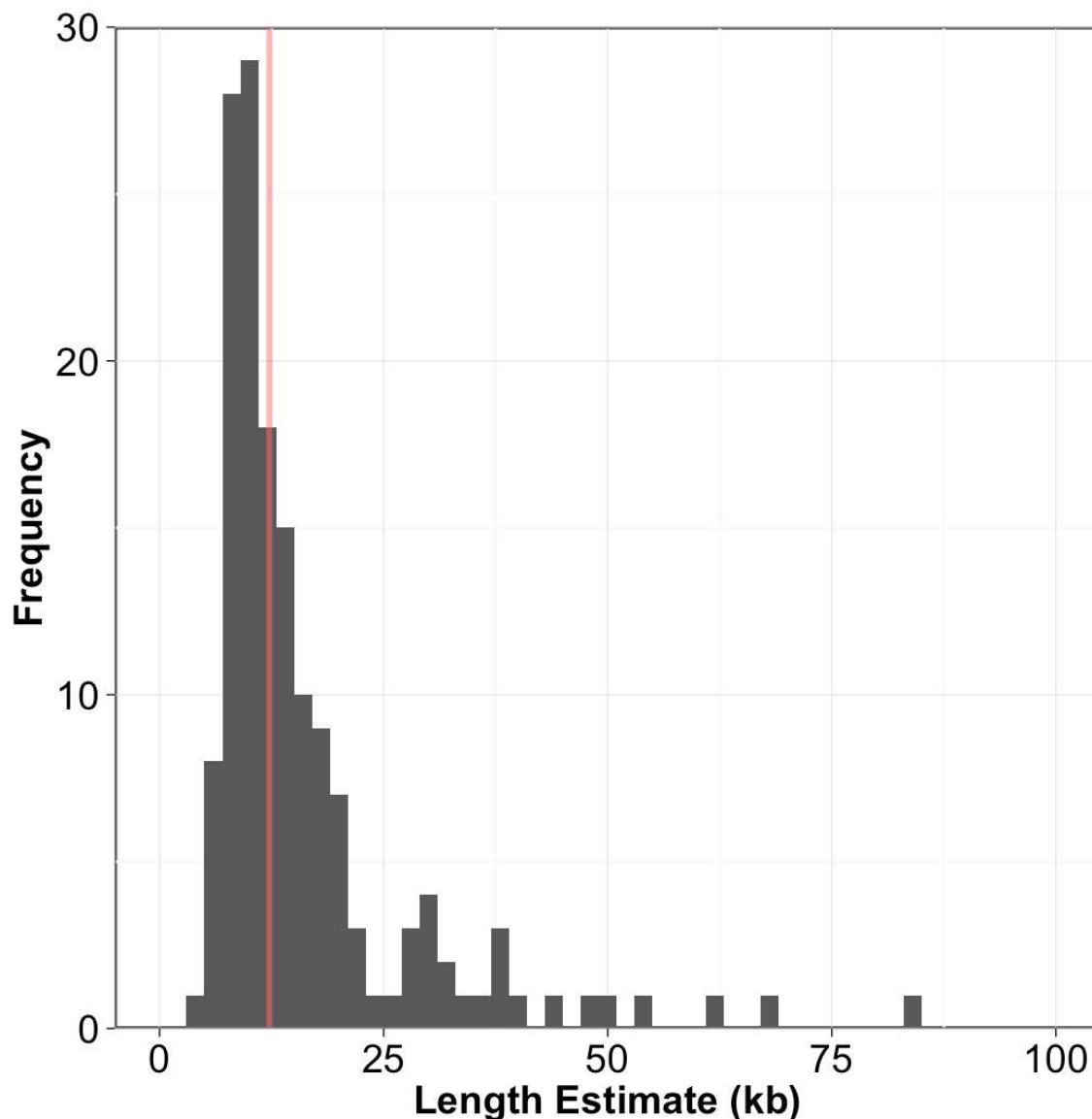
943 Genet. **28**: 160–164.

944 Wicky C., Villeneuve a M., Lauper N., Codourey L., Tobler H., Müller F., 1996

945 Telomeric repeats (TTAGGC)<sub>n</sub> are sufficient for chromosome capping function in

946 *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U. S. A. **93**: 8983–8988.

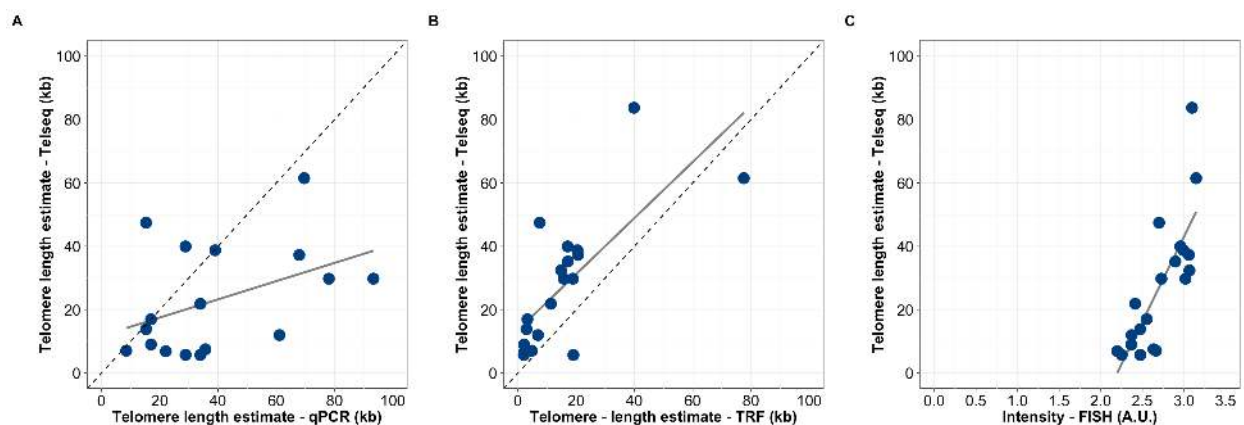
947 Figures



948 **Fig 1. Distribution of telomere-length estimates** A histogram of telomere-length  
949 estimates weighted by the number of reads sequenced per run is shown. Bin width is 2.  
950 The red line represents the median telomere-length estimate of 12.2 kb.  
951

952

953



954

955 **Fig 2. Telomere-length estimates correlate with alternative molecular**

956 **measurement methods** Scatterplot of Telseq Telomere-length estimates (y-axis)

957 plotted against alternative methods of telomere length measurement on the x-axis.

958 Alternative methods plotted on the x-axis and their associated Spearman's rank

959 correlation are (A) qPCR measurements normalized by N2 qPCR value and scaled

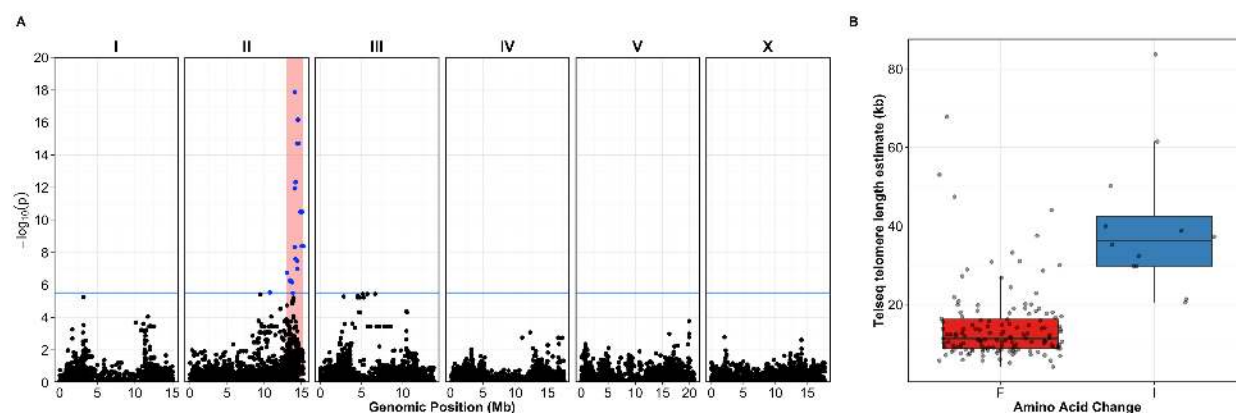
960 relative to the TelSeq N2 telomere length estimate ( $\rho = 0.445$ ,  $p = 0.049$ ), (B) TRF

961 ( $\rho = 0.699$ ,  $p = 8.5e^{-4}$ ) and (C) FISH ( $\rho = 0.815$ ,  $p = 1.03e^{-5}$ ). Grey lines represent

962 the regression lines between Telseq and each method. Dashed diagonal lines represent

963 identity lines.

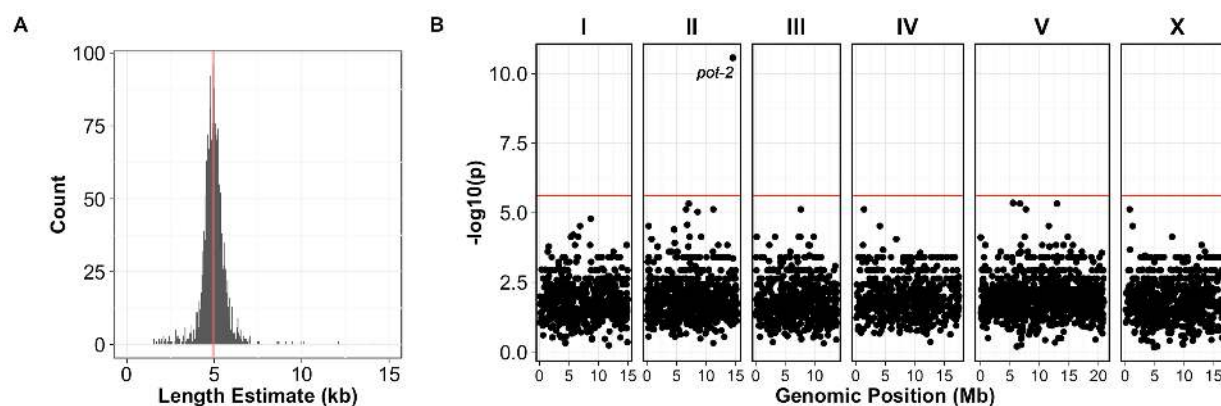
964



965  
966 **Fig 3. Genome-wide association of telomere length** (A) Genome-wide association of  
967 telomere length residuals (conditioned on DNA library) is visualized using a Manhattan  
968 plot. Genomic coordinates are plotted on the x-axis against the negative of the log-  
969 transformed  $p$ -value of a test of association on the y-axis. The blue bar indicates the  
970 Bonferroni-corrected significance threshold ( $\alpha = 0.05$ ). Blue points represent SNVs  
971 above the significance threshold whereas black points represent SNVs below the  
972 significance threshold. Light-red regions represent the confidence intervals surrounding  
973 significantly associated peaks. (B) Shown is the split between TelSeq estimated  
974 telomere lengths (y-axis) by genotype of *pot-2* at the presumptive causative allele as  
975 boxplots (x-axis). The variant at position 14,524,396 on chromosome II results in a  
976 putative F68I coding change. Horizontal lines within each box represent the median,  
977 and the box represents the interquartile range (IQR) from the 25th-75th percentile.  
978 Whiskers extend to 1.5x the IQR above and below the box. The plotted points represent  
979 estimates beyond 1.5x the IQR.  
980



981

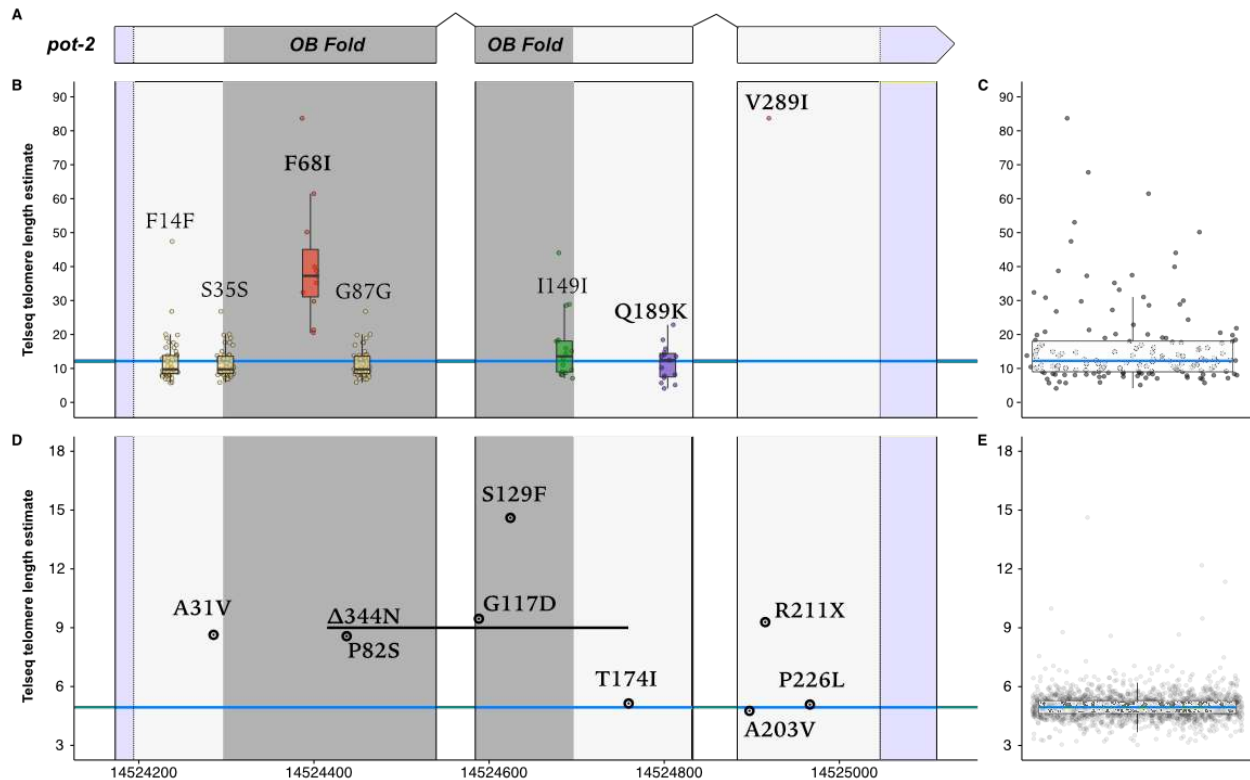


982

983 **Fig 4. Mutations in *pot-2* are more often found in strains with long telomeres than**  
984 **in strains with short telomeres** (A) A histogram of telomere-length estimates among  
985 the 1,936 mutagenized strains from the Million Mutation Project. Median telomere length  
986 is 4.94 kb. (B) Plot of significance from a hypergeometric test for every *C. elegans*  
987 protein-coding gene. The red line represents the Bonferroni ( $\alpha = 0.05$ ) threshold set  
988 using the number of protein coding genes (20,447). Each point represents a gene  
989 plotted at its genomic position on the x-axis, and the log transformed  $p$ -value testing for  
990 enrichment of mutations in long-telomere strains.

991

992

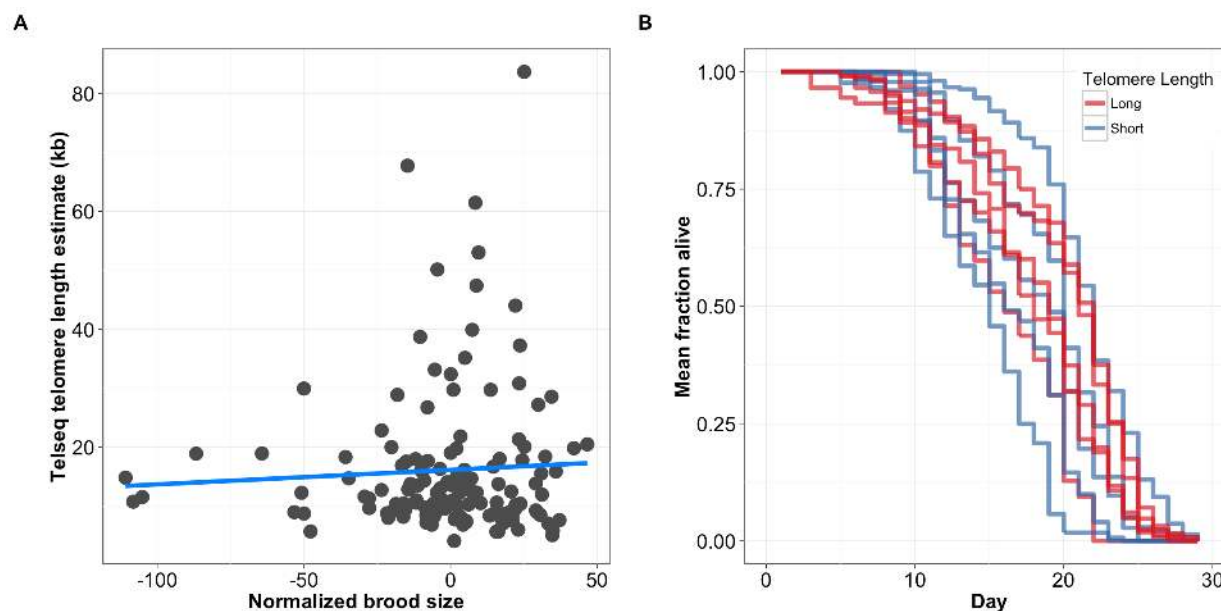


993

994 **Fig 5. Variation within *pot-2* in wild isolate and Million Mutation Project strains**

995 Natural variation and induced mutations that alter codons across *pot-2* are shown along  
 996 with the telomere-length estimates for all strains. In panel (A), a schematic illustrating  
 997 the *pot-2* genomic region is shown. The dark gray region represents the part of the  
 998 genome encoding the OB-fold domain. Purple regions represent untranslated regions.  
 999 (B) Strains that harbor the alternative (non-reference) allele are plotted by telomere  
 1000 length on the y-axis and genomic position on the x-axis. Both synonymous and  
 1001 nonsynonymous variants are labeled. Variants resulting in a nonsynonymous coding  
 1002 change are bolded. The blue line indicates the median telomere length value for wild  
 1003 isolates. The color of boxplots and markers indicates variants from the same  
 1004 haplotypes. (C) Boxplot of natural isolate distribution of telomere lengths. Blue lines  
 1005 within the center of each box represent the median while the box represents the  
 1006 interquartile range (IQR) from the 25th – 75th percentile. Whiskers extend to 1.5x the  
 1007 IQR above and below the box. All plotted points represent estimates beyond 1.5x the  
 1008 IQR. (D) Telomere length is plotted on the y-axis as in (B), but strains do not share

1009 mutations because strains harbor unique collections of induced alleles. The blue line  
1010 indicates median telomere length for the MMP population. (E) Boxplot of the distribution  
1011 of telomere lengths in the MMP is shown. Boxplot follows same conventions as in (C).  
1012 N2 telomere length in our wild isolate population was estimated to be 16.9 kb whereas  
1013 median telomere length in MMP was 4.94 kb. Differences are likely caused by library  
1014 preparation method and/or sequencing method.  
1015



1016  
1017 **Fig 6. Fitness traits are not associated with telomere length** (A) Normalized brood  
1018 sizes (x-axis) of 152 wild isolates are plotted against the telomere-length estimates from  
1019 those same strains (y-axis). The blue line indicates a linear fit of the data. However, the  
1020 correlation is not significant ( $\rho = -0.062$ ,  $p = 0.463$ ). (B) Survival curves of nine wild  
1021 isolates with long and short telomeres. Lines represent aggregate survival curves of  
1022 three replicates. Survival among long and short telomere length strains is not  
1023 significantly different ( $p = 0.358$ ; Mantel-Cox analysis).  
1024