

Additional file 1. Supplementary Results and Discussion

Bulk RNA-seq dataset QC

After mapping the RNA-seq datasets to the *H. bakeri* genome using STAR, reads overlapping each transcript were counted (See main Methods). The read counts per transcript were used to perform principal component analysis and hierarchical clustering (Additional file 2: Fig S1 and S2). Biplots of the first two principal components are shown for either the raw counts (Additional file 2: Fig S1A, accounting for 86.3% of the variation) or the VST-transformed counts from DESeq2 (Additional file 2: Fig S1B, accounting for 83.7% of the variation). In both cases, all replicates of each sample condition cluster together and sample conditions cluster relative to each other in an expected way (ie. Male and female samples cluster separately, lumen-dwelling and tissue-dwelling samples cluster separately). These patterns indicate our samples are suitable for further comparison.

Hierarchical clustering using either Euclidean (Additional file 2: Fig S2A), Manhattan (Additional file 2: Fig S2B), or Canberra (Additional file 2: Fig S2C) distance shows sample replicates clustering together with sample groups tending to cluster by sex rather than by age (ex. lumen females cluster together rather than D21 worms clustering together). This pattern matches that seen in the principal component biplots and further indicates our samples are what we intended and are suitable for further comparison.

Agreement with other Genome Sequences

Ten million reads were randomly selected from all libraries using seqtk. They were mapped to genome assemblies for *H. bakeri* (GCA_947359475.1_nxHelBake1.1, Lewis Stevens personal communication) and *H. polygyrus* (GCA_947396885.1_ngHelPoly1.1, Lewis Stevens personal communication) using STAR v2.7.10b with default parameters. The STAR alignment score is recorded in the 'AS' field for each read, which was used to determine to which assembly each read mapped better using commands in Additional file 4. We found 93% of the reads mapped to *H. bakeri* with 0.5% mismatch rate per base and 89% of the reads mapped to *H. polygyrus* with 2% mismatch rate per base. When considering the alignment scores, 78% of reads mapped to better to *H. bakeri*, 8% mapped better to *H. polygyrus*, and 14% mapped equally well to both. Together, these mapping rates indicate that, unsurprisingly, the lab isolates (*H. bakeri*) are more similar to each other than to the wild isolate (*H. polygyrus*).

H. bakeri PRJEB15396 Annotations

To evaluate the completeness and accuracy of the genome annotations, the RNA-seq reads were aligned to the transcript sequences predicted from the genome annotation that accompanies the PRJEB15396 assembly using the non-splice-aware aligner bwa v0.7.17-r1188 (Additional file 3: Table S1). Among all the datasets, 83.70 – 89.41% of the reads mapped to the predicted transcript sequences, suggesting that the annotations have captured the bulk of the polyadenylated transcript content, but there is still room for improvement. In comparison to the mapping rates to the genome, the lower mapping rates to the predicted transcripts point to the annotations missing regions that are actually

transcribed. Moreover, the rates of mapped singletons, mismatched pairs, and misoriented pairs point to occasional missing exons and/or untranslated regions among the predicted transcripts.

Samples included in pairwise comparisons

Of the original 32 samples sequenced (four for each timepoint/sex combination examined in the study), two stood out in preliminary quality control analyses as problematic. D10 F6 and D10 F7 both clustered with D10 M samples in PCA biplots and hierarchical clustering dendrograms (data not shown). Moreover, the most abundant transcript in these samples is a major sperm protein (HPOL_0001609601), as is the case for the adult male samples. Furthermore, expression of orthologs of transcriptionally regulated male patterning genes in *C. elegans* *mab-3* (HPOL_0001902701) and *mab-23* (HPOL_0000964001) in these two samples match the levels seen in the other male samples. We therefore sequenced an additional five D10 F samples and sequenced two technical replicates (D10 M3 and D10 M6) to ensure no batch effects were seen between sequencing runs. No batch effects were observed in PCA analyses (data not shown) so the seven D10 F samples that clustered together (and near the D21 F samples) were used in all analyses. The D10 F6 and D10 F7 samples were excluded from all analyses except the assessment of the genome assembly and annotation.

Intestinal oxygen in mice

While air at sea level has a partial pressure of oxygen (pO_2) of ~145 mmHg (~21% O_2), pO_2 varies considerably throughout the different tissues in mammalian bodies [1]. Similarly to humans, the mouse gastrointestinal (GI) tract has a pO_2 gradient along the proximal-distal axis and along the radial axis [1]. Much research focuses on the distal GI tract and the anaerobic colon, however, *H. bakeri* larvae encyst in the wall of the duodenum in the small intestine and stay as adults in the lumen of the duodenum. Initial measurements in mice using spatial and spectral electron paramagnetic resonance (EPR) imaging placed the pO_2 of the lumen of the duodenum at 32 ± 8 mmHg [2]. A second set of measurements in mice using phosphorescence quenching placed the pO_2 of the lumen of the stomach at ~22 mmHg, the stomach tissue at ~27 mmHg, the lumen of the duodenum at ~61 mmHg, and the tissue of the duodenum at ~45 mmHg [3]. In this study it was noted that the measurement for the lumen of the duodenum required further investigation due to the possibility of small molecule quenchers in the pancreatic/bile secretions that could have given an artificially high measurement.

Sources of oxygen in the lumen of the duodenum include: 1) oxygen still in the material entering from the stomach, 2) oxygen diffusing from the duodenum tissue, and 3) oxygen in the pancreatic and bile secretions entering the duodenum [3]. The measures of the first two sources were well below the 61 mmHg measurement in question. In the periportal area of the human liver (where fresh oxygenated blood enters), the pO_2 is 65 mmHg [4]. Assuming a similar level in the mouse liver, bile in secretions would not be sufficient to boost oxygen concentration to higher than the surrounding tissue unless those secretions made up the vast majority of the content in the lumen. The alternative measurement of 32 mmHg is therefore a more reasonable estimate of the pO_2 of the duodenum lumen. This level is generally considered 'physiologically hypoxic' because in addition to being a low pO_2 , the epithelial cells

lining the GI tract lumen display typical hypoxia responses that are regulated by the transcription factor hypoxia-inducible factor [1].

Aging-related gene expression

To identify transcripts in *H. bakeri* with potential roles in aging we sought to find all the transcripts with expression patterns matching the orthologs of genes in *C. elegans* that have already been implicated in aging. We identified an initial 18 candidate orthologs (See table below), whose expression is plotted in Additional file 2: Fig S4. In addition to many sex-specific expression patterns, the 18 candidate orthologs show a varied assortment of expression patterns including: 1) increasing with age (ex. *dcap-2* males), 2) decreasing with age (ex. *eat-2* males), 3) changing upon final molt (ex. *aak-2* females), 4) spiking at a certain time point (ex. *daf-7* males), and 5) not changing (ex. *skn-1* males). Identifying transcripts whose expression mirrors a candidate may identify potential interacting partners of the candidate gene product. However, in the case of the process of aging, all potential expression patterns were found among genes already implicated in the process so new candidates could not be identified based on expression pattern alone. Additional time points may refine expression patterns of interest, but without the ability to follow up on new candidate genes with the use of mutants or RNAi it is unlikely true involvement in aging can be confirmed in this worm.

<i>C. elegans</i> Gene Implicated in Aging	Reference
<i>daf-2</i>	[5]
<i>age-1</i>	[5]
<i>daf-16</i>	[5]
<i>skn-1</i>	[5]
<i>aak-2</i>	[5]
<i>hsf-1</i>	[5]
<i>eat-2</i>	[5]
<i>dcr-1</i>	[6]
<i>alg-1</i>	[6]
<i>dcap-1</i>	[6]
<i>dcap-2</i>	[6]
<i>xrn-1</i>	[6]
<i>jnk-1</i>	[5]
<i>cst-1</i>	[5]
<i>daf-7</i>	[5]
<i>pqm-1</i>	[5]
<i>sir-2.1</i>	[5]
<i>sod-2</i>	[5]

Immunomodulatory gene expression

Expression of immunomodulatory genes that could be identified is described in the main text. HbTGM1-3 were excluded from this analysis because the locus in the PRJEB15396 that was previously

associated with these proteins was incorrectly assigned. The locus does not match the length of the reported protein(s), nor does the sequence from that locus match the sequences associated with these proteins. Analysis of additional unidentified immunomodulatory proteins was attempted based on gene annotations but cannot be linked to any of the peptides detected in analyses of secreted products because the information needed to do so was never made available.

References

1. Zheng L, Kelly CJ, Colgan SP. Physiologic hypoxia and oxygen homeostasis in the healthy intestine. A review in the theme: Cellular responses to hypoxia. *Am J Physiol Cell Physiol*. 2015;309:C350–60.
2. He G, Shankar RA, Chzhan M, Samouilov A, Kuppusamy P, Zweier JL. Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. *PNAS*. 1999;96:4586–91.
3. Friedman ES, Bittinger K, Esipova T V., Hou L, Chau L, Jiang J, et al. Microbes vs. chemistry in the origin of the anaerobic gut lumen. *PNAS*. 2018;115:4170–5.
4. Singhal R, Shah YM. Oxygen battle in the gut: Hypoxia and hypoxia-inducible factors in metabolic and inflammatory responses in the intestine. *J Biol Chem*. 2020;295:10493–505.
5. Uno M, Nishida E. Lifespan-regulating genes in *C. elegans*. *npj Aging Mech Dis*. 2016;2:16010.
6. Son HG, Altintas O, Kim EJE, Kwon S, Lee SJ V. Age-dependent changes and biomarkers of aging in *Caenorhabditis elegans*. *Aging Cell*. 2019;18:e12853.