Supplementary methods

3' UTR sequences, alignments, and PicTar runs for *C. elegans* and *C. briggsae*

We used WormBase (http://www.wormbase.org), version WS133 of the annotated *C. elegan*s genome. All available annotated 3' UTRs were used. For genes with no annotated 3'UTR, a 3' UTR was defined as the region immediately downstream from the stop codon, extending either to a length of 1,500 bp or up to the next adjacent coding region, whichever was shorter. We thus assembled a set of 22,217 sequences. *C. briggsae* 3'UTRs (WormBase CB25) were obtained using orthologous genes from *C. elegans*[1]. The length of the putative 3'UTR was defined either as the length of the corresponding *C. elegans* 3'UTR (if annotated), or as extending from the *C. briggsae* stop codon to a length of 1,500 bp or up to the next adjacent *C. briggsae* coding region (whichever was shorter). We used DIALIGN[2] 2.2.1 with default parameters to align orthologous 3' UTRs. If the sequence from one species was more than five times the length of the other sequence, we adjusted the length to that of the shorter sequence. Pairs of sequences were discarded if one member was less than 30 bp. Our final data set contained 10,607 orthologous 3' UTR pairs. The total length of all 3' UTR sequences was 5,794,092 bp for *C. elegans* and 6,078,515 bp for *C. briggsae*. All PicTar runs were performed with the number of required anchors set to 1.

Promoter sequences and alignments in *C. elegans* and *C. briggsae*

For all *C. elegans* genes in Wormbase WS133 we extracted 500 bp upstream from the start of the transcript or of the coding sequence if the transcription start was not annotated. Sequences overlapping another gene were truncated accordingly. Orthologous sequences in *C. briggsae* were defined analogously by extracting 500 bp upstream of the orthologous gene or up to the next adjacent coding region. A total of 11,121 orthologous pairs of promoters were constructed and aligned as described above.

Correlation with evolutionary conservation

For the 300 top scoring UTRs, we counted how many of the predicted binding sites fell into regions that are conserved between *C. elegans* and *C. briggsae*. We then randomly shuffled the positions of all predicted binding sites within each UTR and recorded the number of shuffled sites that resided in conserved sequence. The randomization was repeated 10,000 times for each UTR. 146 (that is, 14%) of a total of 1,049 PicTar predicted sites were conserved, but on average only 4% of the shuffled sites were conserved. The distribution of the number of conserved shuffled sites was very close to a Gaussian distribution with a standard deviation of 6.3 (data not shown). The significance S of observing 14% of the predicted sites to be evolutionarily conserved was thus very high (>10 standard deviations).

Testing the quality of the 3' UTR multiple alignments

To test the degree of true orthology between the sequences in our dataset we compared the human/mouse sequence pairs from the genome-wide multiple alignments to a set of independently defined curated 14,869 human/mouse RefSeq orthologs from the Jackson lab (ftp://ftp.informatics.jax.org/pub/reports/HMD_Human1.rpt). We extracted a test set

of ~1000 randomly selected unique pairs of putative human-mouse orthologous sequences from our multiple alignments and used the UCSC mappings of the mouse RefSeq dataset to the mouse genome to look up the mouse transcripts that corresponded to our mouse sequences. 64% of transcripts from our test set are present in the ortholog table. From these, 98% coincide with our human/mouse sequences as defined by the multiple alignments. The remaining genes (2%) in the UCSC data appeared to be mapped to paralogs. Virtually all of the remaining 425 genes from our test set mapped to known mouse RefSeq transcripts, and we found that the corresponding functional annotation of human/mouse RefSeq transcripts strongly supported orthology for additional 212 genes (22% from our test set). 11% from the test set had no functional annotation, but strikingly for only 3% of our test set genes was the RefSeq functional annotation different in both organisms.

References

(1) Stein, L.D. et al. The Genome Sequence of Caenorhabditis briggsae: A Platform for Comparative Genomics. *PLoS Biol.* **1(2)**, E45 (2003).

(2) International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716 (2004).