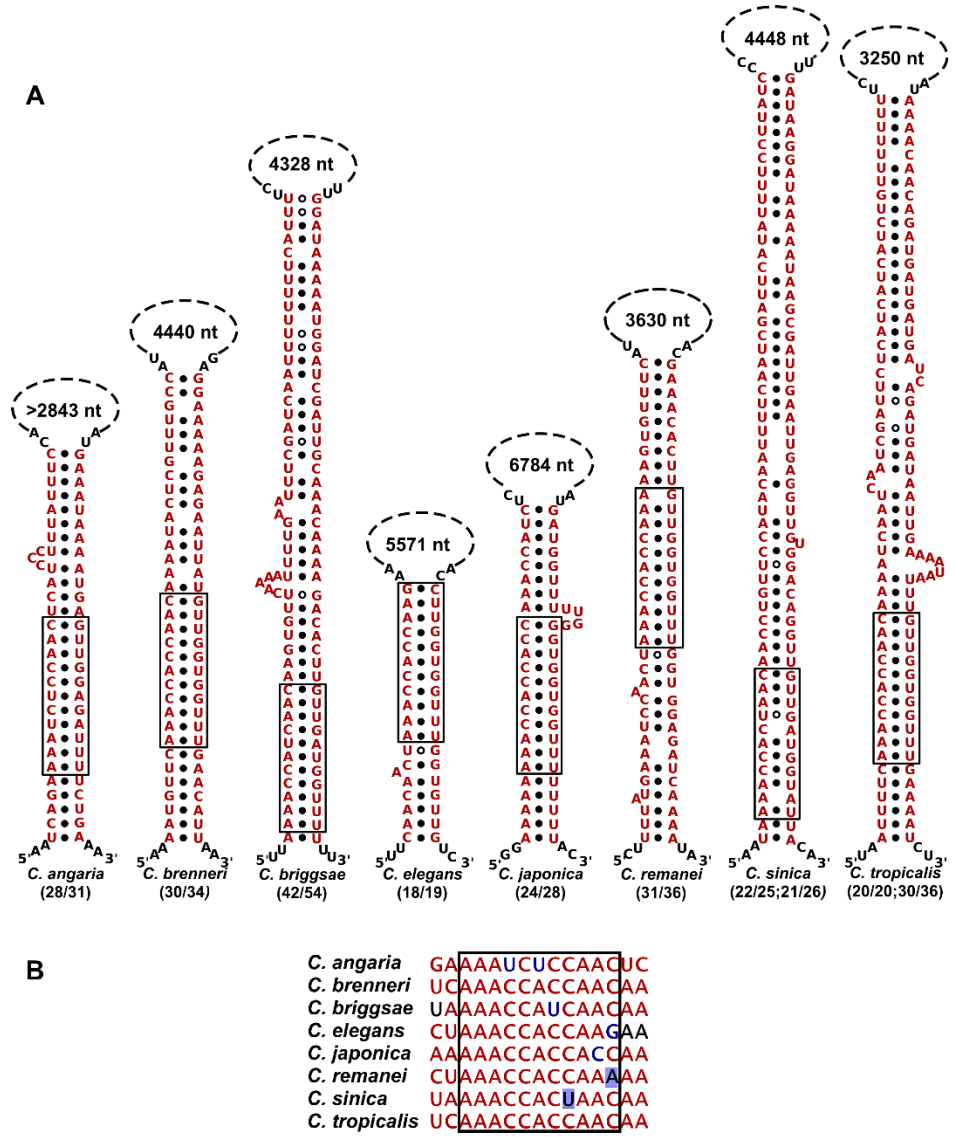

Figure S1 Complementary R1 elements from the cholinergic loci of eight Caenorhabditis species. (A) R1 stem-loop structures. The fractional pairing of the stem is given below each species name; the length of the putative loop is shown above each stem. The consensus sequence (see below) is boxed. The solid black circles represent standard (A-U and G-C) pairing; circles with white centers represent G-U pairing. (B) Identification of an 11-nucleotide R1 consensus sequence (AAACCACCAAC) in the upstream stem sequences shown in panel A. The consensus (containing nucleotides present in the stems of at least six of the eight species) is boxed. Nucleotides in **Red** are paired (in the stem); nucleotides in **Black** are unpaired; the nucleotides in **Blue** deviate from the consensus, but are paired with complementary downstream nucleotides; the nucleotide with the **Blue background** deviates from the consensus and is unpaired. The species previously known as *C. sp5* and *C. sp11* have recently been renamed *C. sinica* and *C. tropicalis*, respectively (Félix *et al.* 2014; Huang *et al.* 2014).

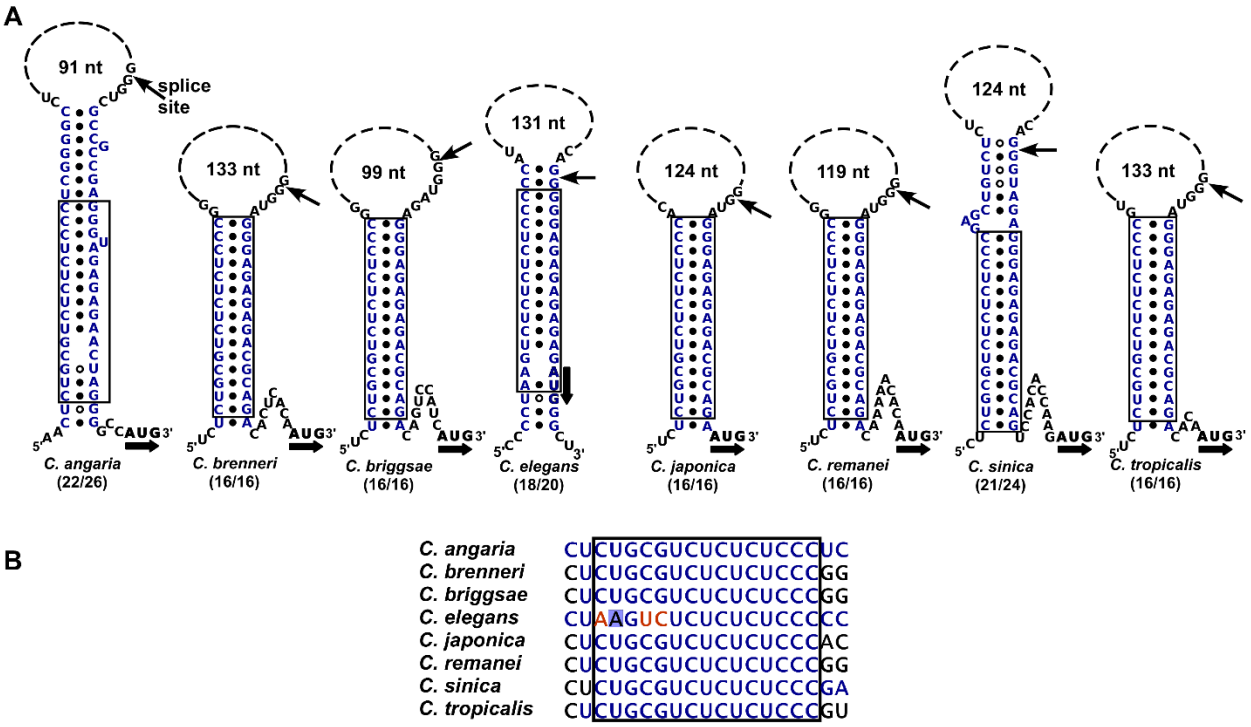

Figure S2 Complementary R2 elements from the cholinergic loci of eight Caenorhabditis species. (A) R2 stem-loop structures. The fractional pairing of the stem, the length of the putative loop, the consensus sequence, and the black circles are the same as in Figure S1. For each R2 structure, the splice acceptor site at the 5'-end of exon u2 (Figure 1) is indicated with a thin arrow, and the UNC-17 initiation (AUG) codon is indicated with a thick arrow. (B) Identification of a 16-nucleotide R2 consensus sequence (UCUGCGUCUCUCUCCC when transcribed) in the upstream stem sequences shown in panel A. The consensus (containing nucleotides present in the stems of at least seven of the eight species) is boxed. Nucleotides in **Blue** are paired (in the stem); nucleotides in **Black** are unpaired; the three nucleotides in **Red** deviate from the consensus, but are paired with complementary downstream nucleotides; the nucleotide with the **Blue background** deviates from the consensus and is unpaired. The species previously known as *C. sp5* and *C. sp11* have recently been renamed *C. sinica* and *C. tropicalis*, respectively (Félix *et al.* 2014; Huang *et al.* 2014).

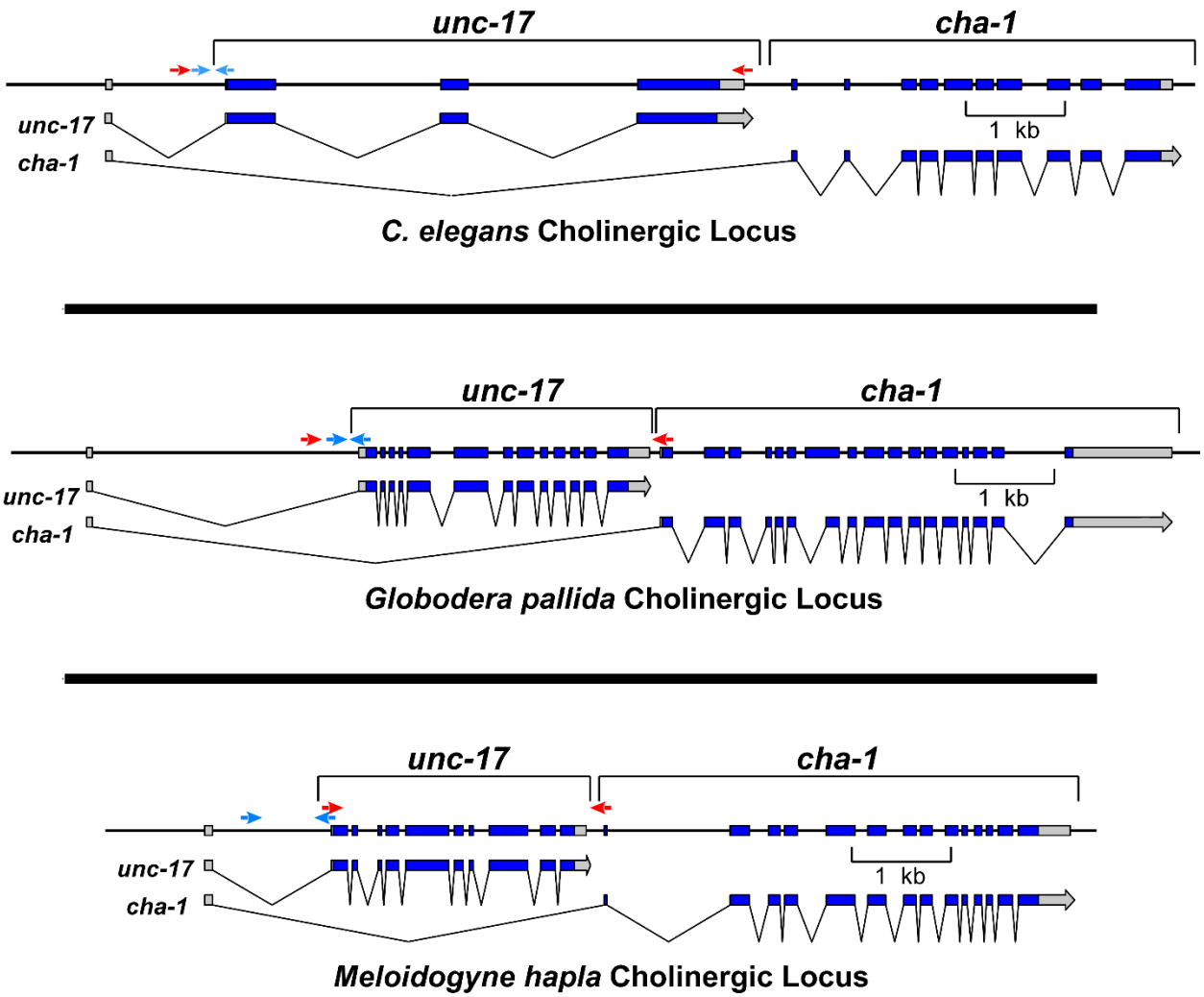

Figure S3 Genomic structures of the cholinergic loci in *Globodera pallida* and *Meloidogyne hapla*, two nematode species with an abundance of UNC-17 exons. The structure of the *C. elegans* locus (at the same scale) is shown for comparison. Because the common exon is non-coding, it could not be identified easily by homology. Instead, the criteria used for provisional identification of the common exon in *G. pallida* and *M. hapla* were: 1) good matches to canonical splice acceptor and donor sequences; 2) size was 50 - 90 bp; and 3) the sequence contained no ATG triplets. The predicted extent of the 3'-UTRs for *G. pallida* and *M. hapla* were based on location of AATAAA sequences. Red and blue arrows indicate the locations of the sequences corresponding to the R1 and R2 complementary elements, respectively.

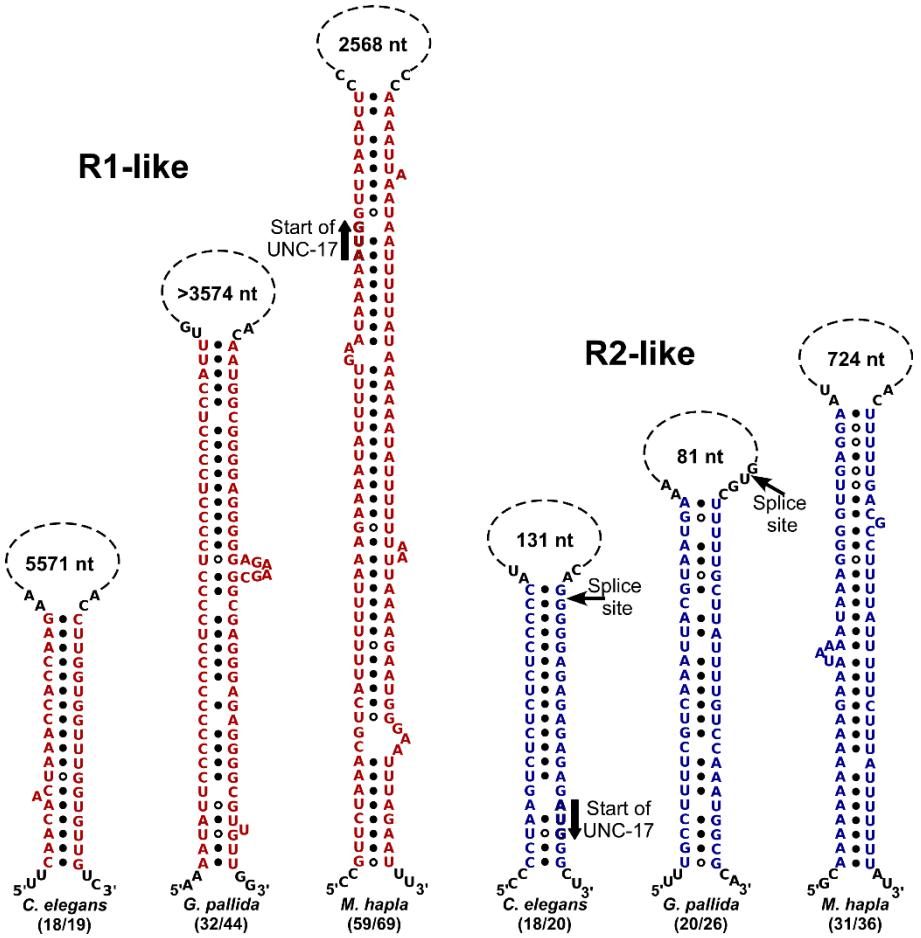

Figure S4 Putative stem-loop alignments of the R1-like and R2-like sequences from the cholinergic loci of *Globodera pallida* and *Meloidogyne hapla* (see Figure S3). The corresponding *C. elegans* alignments are shown for comparison. Splice sites and UNC-17 initiation codons are indicated (when present). The exact size of the *G. pallida* R1-like loop is uncertain because of a gap in the genomic sequence.

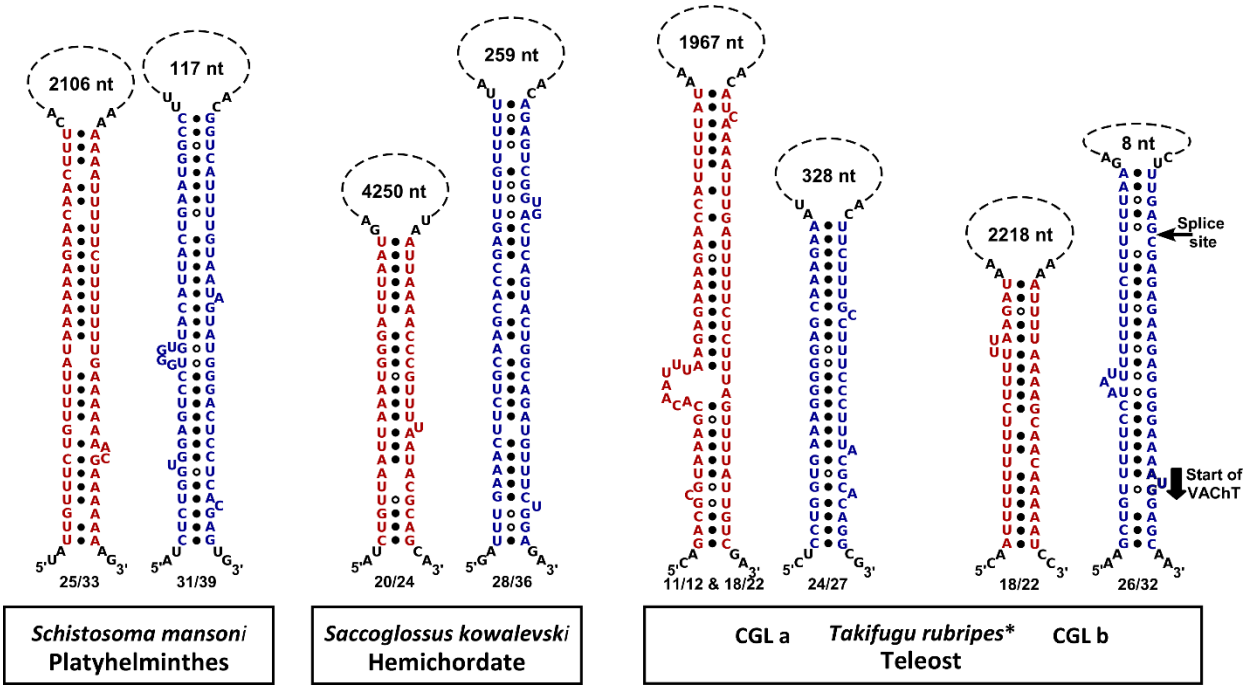

Figure S5 Putative stem-loop alignments of the R1-like (red) and R2-like (blue) sequences from the cholinergic loci of representative species from three different animal phyla (these are in addition to the species presented in Figure 6). The fractional pairing of the stem is given below each stem; the length of the putative loop is shown above each stem. Splice sites at the 5'-end of the first VAChT coding exon are indicated with a thin arrow, and the VAChT initiation codon is indicated with a thick arrow. *The *T. rubripes* genome reflects a whole genome duplication, and therefore has 2 copies of the CGL. Official *T. rubripes* gene names for the CGLs are still unassigned or provisional; for the present investigation, we refer to the CGL on Chromosome 1 as CGL a and the locus on Chromosome 4 as CGL b.

File S1

Supporting Materials and Methods

Strain List

N2	Wild type	Figure 4A, B
RM2256	*pha-1(e2123)III*	Recipient for transgenic injections
RM523	*unc-17(cn355)IV*	Figure 4C, D
RM3288	*pha-1(e2123)III; mdEx833[Punc-17::dual-reporter(wt) pBX pBS]*	Figure 3; 4E, F; 5A
RM3302	*pha-1(e2123)III; mdEx847[Punc-17::dual-reporter(cn355) pBX pBS]*	Figure 4G, H
RM3376	*pha-1(e2123)III; mdEx909[dual-reporter(wt) pBX pBS]*	No promoter (Control)
RM3317	*pha-1(e2123)III; mdEx857[Punc-17::dual-reporter(R1u:scr)* pBX pBS]*	Figure 5B
RM3323	*pha-1(e2123)III; mdEx863[Punc-17::dual-reporter(R1ud:scr) pBX pBS]*	Figure 5C
RM3319	*pha-1(e2123)III; mdEx859[Punc-17::dual-reporter(R2d:scr) pBX pBS]*	Figure 5D
RM3321	*pha-1(e2123)III; mdEx861[Punc-17::dual-reporter(R2ud:scr) pBX pBS]*	Figure 5E
RM3347	*pha-1(e2123)III; mdEx885[Punc-17::dual-reporter(R1uR2d:scr) pBX pBS]*	Figure 5F

*Note: R1u:scr = R1 upstream element scrambled (see Table S1);

R1ud:scr = R1 upstream and downstream elements scrambled; etc.

Details for the *unc-17(cn355)* allele (see Figure 4)

Nature of mutation: A>G transition

Flanking Sequences: AAATTTAGAAAAAATAAAATATTCC/ **A>G** /GGGGGAGAGAGAGAGATGGGCTTCA

(in direction of transcription)

Sources and accession numbers of genomic CGL sequences

Genomic sequences for all Caenorhabditis species were downloaded from WormBase, Release WS240 (www.wormbase.org).

Downloaded from the Sanger Center (www.sanger.ac.uk/resources/downloads/helminths/globodera-pallida.html):

Globodera pallida pathogens_Gpal_scaffold_214.1

pathogens_Gpal_scaffold_214.2

pathogens_Gpal_scaffold_146

Downloaded from NCBI (www.ncbi.nlm.nih.gov):

Ciona intestinalis	NW_001955240 REGION: 1..30000
Drosophila melanogaster	NT_033777 REGION: 14525001..14560000
Meloidogyne hapla	ABLG01001183 **and** ABLG01002582
Mus musculus	AC167565
Rattus norvegicus	NW_047469 REGION: complement(7900000..8006000)
Saccoglossus kowalevskii	NW_003151267 REGION: 100001..150000
Schistosoma mansoni	NS_000200 REGION: 148400001..148540000
Takifugu rubripes *	CGL a (chr 1): NC_018890 REGION: complement(3185000..3200000)
	CGL b (chr 4): NC_018893 REGION: 10170000..10190000

* Note: The genomes of *T. rubripes* and other teleosts reflect a whole genome duplication, and therefore have 2 (somewhat diverged) copies of many genes, including the CGL. Official *T. rubripes* gene names for the CGLs are still unassigned, but we provisionally refer to the CGL on Chromosome 1 as CGL a, and the locus on Chromosome 4 as CGL b.

DNA/RNA sequence analysis

Sequence analysis utilized Vector NTI® software (Life Technologies Corporation, Carlsbad, CA) or the Lasergene® Suite (DNASTAR, Inc., Madison, WI). For several species, genomic annotation of the CGL was incomplete or incorrect; in such cases, the genomic organization of the *unc-17* and *cha-1* homologs was deduced through analysis with TBlastN, identification of splice-site consensus sequences, and direct determination of homology. RNA structures were analyzed with Mfold (Zuker 2003) or Sfold (Ding and Lawrence 2003; Ding *et al.* 2005). The criteria for R1-like elements (in addition to complementarity) were that they flank the complete VAChT coding sequence but do not include any part of the ChAT coding sequence; the criterion for R2-like elements was that they flank or overlap the splice site of the first VAChT coding exon.

Construction of dual-reporter plasmid RM#942p from 7 fragments cloned into pCRII-TOPO

Starting Plasmids:

pBS: pBlueScript (Stratagene)

TOPO: pCRII-TOPO (Invitrogen)

pAA64: mCherry plasmid with 3 artificial introns (McNally *et al.* 2006; Green *et al.* 2008).

RM#651p: Fire-type modular plasmid with an empty cloning site 1 and GFP in cloning site 2.

RM#691p: Fire-type plasmid with CFP driven by 4.4 kb of genomic sequence upstream of the *unc-17* start codon.

Clone A: RM#691p was digested with *Xba*I and recircularized to remove the 3.2 kb *unc-17* promoter, the common exon, and the first 339 bp of the *unc-17* 1st intron.

Clone D: mCherry fragment (minus the initiation Met) amplified from pAA64 with p2288 and p2206. **Note**: A lower-case "p" followed by a number refers to a primer listed below.

Clone E: *unc-17* 3'-UTR amplified from N2 genomic DNA with p13 and p2301.

Clone F: *cha-1* 3'-UTR amplified from N2 genomic DNA with p2290 and p45.

Clone H: GFP fragment from RM#651p amplified with p2291 and p2292.

Clone I: *unc-17* 5'-UTR amplified from N2 genomic DNA with p62 and p2293.

Clone J: *unc-17* 3'-UTR amplified from N2 genomic DNA with p2294 and p49.

Cloning Steps:

Clone D was transferred from TOPO to pBS with *Xho*I+*Spe*I → Clone D'

Clone F/*Nhe*I+*Spe*I was subcloned into D'/*Spe*I → Clone D'F

Clone E/*Sac*I(blunt) was subcloned into D'F/*Sna*BI → Clone ED'F

Clone ED'F was subcloned into Clone A with *Avr*II+*Spe*I → Clone B

Clone H was transferred from TOPO to pBS with *Apa*I+*Sac*I → Clone H'

Clone I was subcloned into Clone H' with *Bam*HI → Clone IH'(B)

Clone IH'(B) was digested with *Xba*I and recircularized → Clone IH'

Clone J was subcloned into Clone IH' with *Sac*I → Clone IH'J

Clone IH'J was subcloned into Clone B with *Avr*II → RM#942p.

Primers cited:

p13	CCTTCTCTGTTACCTACAA
p45	GTCTGGTGTTTCTGGGATGA
p49	CATTTGGTGGAAGTTCGTCAAC
p62	TTCCGCATCTCTTGTTCAA
p2206	GAGCTCTTAGGATCCACTAGTCTTATAC (final amino acids of mCherry with *Sac*I site at 5'-end)

p2288 TACGTATCAAAGGGTGAAGAAGATAACA (adds *Sna*BI site; introduces a silent mutation in Val2 of mCherry)

p2290 GCTAGCGGATCCTGAATTTTATTATTATTTTTTGAG (adds the last four amino acids of mCherry with two mutations to alter the *Spe*I site to *Nhe*I - amino acid sequence from Thr-Ser to Ala-Ser - uses the *cha-1* stop codon TGA)

p2291 GGATCCAAAGGAGAAGAACTTTTCACTG (introduces a *Bam*HI site that changes GFP N-terminal amino acid sequence from Met-Ser-Lys to Gly-Ser-Lys)

p2292 GAGCTCATCCATGCCATGTGTAATC (makes two silent mutations; inserts *Sac*I site in Glu-Leu near end of GFP)

p2293 GGATCCCATCTCTCTCTCTCC (creates *Bam*HI site - changes GFP amino acid sequence from Met-Ser to Met-Gly)

p2294 GAGCTCTACAAATAGTCGTAGATTTGGATCTCTG (5'-end corresponds to the last four amino acids of GFP with two silent mutations to generate *Sac*I site)

p2301 GAGCTCATCTGGAACAAAATTTACTTCT (*Sac*I site added at 5'-end)

LITERATURE CITED

Ding, Y., C. Y. Chan and C. E. Lawrence, 2005 RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. RNA 11**:** 1157-1166.

Ding, Y., and C. E. Lawrence, 2003 A statistical sampling algorithm for RNA secondary structure prediction. Nucleic Acids Res. 31**:** 7280-7301.

Félix, M. A., C. Braendle, and A. D. Cutter, 2014 A streamlined system for species diagnosis in Caenorhabditis (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. PLoS One 9: e94723.

Green, R. A., A. Audhya, A. Pozniakovsky, A. Dammermann, H. Pemble *et al.*, 2008 Expression and Imaging of Fluorescent Proteins in the *C. elegans* Gonad and Early Embryo, pp. 179-218 in *Methods in Cell Biology: Fluorescent Proteins*, edited by F. S. Kevin. Academic Press.

Huang, R. E., X. Ren, Y. Qiu and Z. Zhao, 2014 Description of *Caenorhabditis sinica* sp. n. (Nematoda: Rhabditidae), a Nematode Species Used in Comparative Biology for *C. elegans*. PLoS One 9: e110957.

McNally, K., A. Audhya, K. Oegema and F. J. McNally, 2006 Katanin controls mitotic and meiotic spindle length. J. Cell Biol. 175**:** 881-891.

Zuker, M., 2003 Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31**:** 3406-3415.

Table S1 Properties and evaluation of *C. elegans* R1 and R2 sequences

(A) Scrambled R1 and R2 sequences

Element	WT sequence	Scrambled sequence
R1 upstream	CAACAACTAAACCACCAAG	ACACACACCACAAGCCATA (p2524)
R1 downstream	CTTGGTGGTTTGGTGTTG	TATGGCTTGTGGTGTGTGT (p2525)
R2 upstream	TCTCTCTCTCCCCC	CTCCTCCTTCCTCC (p2527)
R2 downstream	GGGGGAGAGAGAGA	GGAGGAAGGAGGAG (p2526)

(B) Genes with coordinated exon skipping

Transcript*	# of exons skipped	Complementary sequence elements?
cha-1	**3**	**Yes**
unc-49B	8	No
unc-49C	16	No
unc-60B	4	No
avr-14B (gbr-2B)	7	No

*Transcript terminology and structure for *cha-1* from Alfonso *et al.* 1994; *unc-49* from Bamber *et al.* 1999; for *unc-60* from McKim *et al.* 1994; for *avr-14* (aka *gbr*-2) from Laughton *et al.* 1997 and Dent *et al.* 2000. Genomic sequences for all genes were downloaded from WormBase, Release WS240 (www.wormbase.org).

(C) Long introns evaluated for complementary sequence elements

Gene/transcript*	Intron examined	Intron size	Complementary sequence elements?
cha-1	**From exon 1 to exon c2**	**6.9 kb**	**Yes**
ric-4A(snap-25A)	From exon 1A to exon 2	7.8 kb	No
gar-3B	From exon 2 to exon 3	9.0 kb	No
gcy-28D	From exon 5 to exon 6	13.6 kb	No

*Transcript terminology and structure for *cha-1* from Alfonso *et al.* 1994; for *ric-4* (aka *snap-25)*, *gar-3*, and *gcy-2* from WormBase. Genomic sequences for all genes were downloaded from WormBase, Release WS240 (www.wormbase.org).

LITERATURE CITED

Alfonso, A., K. Grundahl, J. R. McManus, J. M. Asbury and J. B. Rand, 1994 Alternative splicing leads to two cholinergic proteins in *Caenorhabditis elegans*. J. Mol. Biol. 241**:** 627-630.

Bamber, B. A., A. A. Beg, R. E. Twyman, and E. M. Jorgensen, 1999 The *Caenorhabditis elegans unc-49* locus encodes multiple subunits of a heteromultimeric GABA receptor. J. Neurosci. 19: 5348–5359.

Dent, J. A., M. M. Smith, D. K. Vassilatis, and L. Avery, 2000 The genetics of ivermectin resistance in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA 97: 2674–2679.

Laughton, D. L., G. G. Lunt, and A. J. Wolstenholme, 1997 Alternative splicing of a *Caenorhabditis elegans* gene produces two novel inhibitory amino acid receptor subunits with identical ligand binding domains but different ion channels. Gene 201: 119–125.

McKim, K. S., C. Matheson, M. A. Marra, M. F. Wakarchuk, and D. L. Baillie, 1994 The *Caenorhabditis elegans unc-60* gene encodes proteins homologous to a family of actin-binding proteins. Mol. Gen. Genet. 242: 346–357.