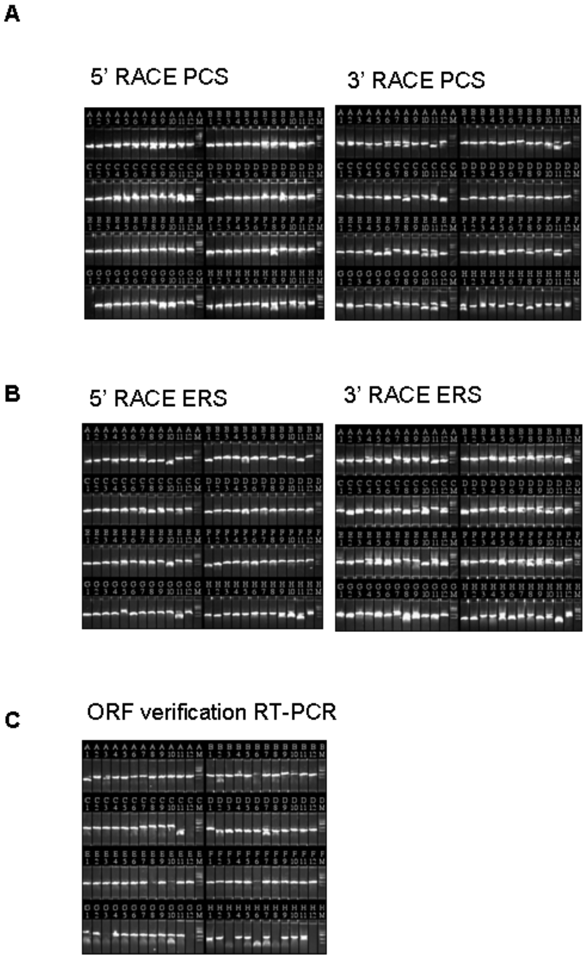
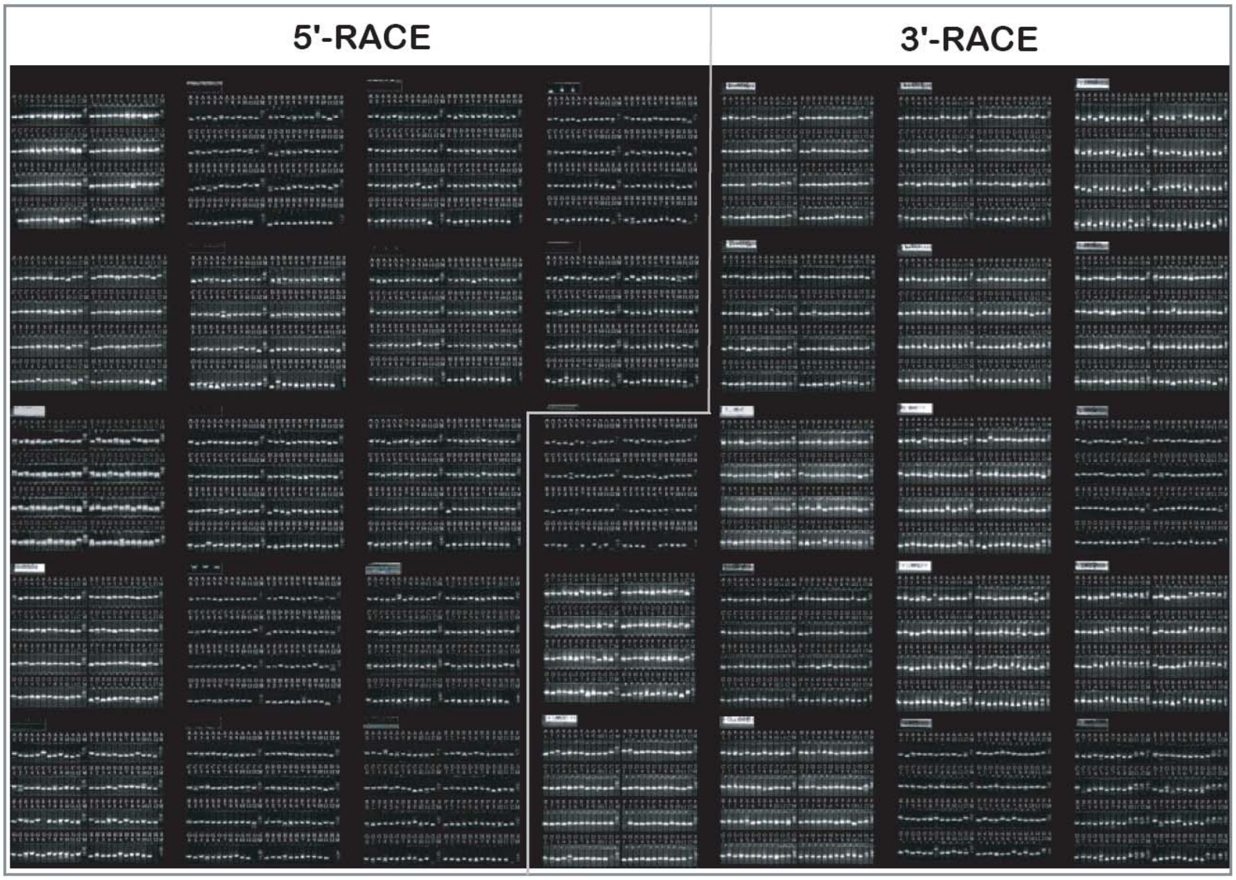
Supplemental File 1 - "Large-scale RACE approach for proactive experimental definition of *C. elegans* ORFeome"

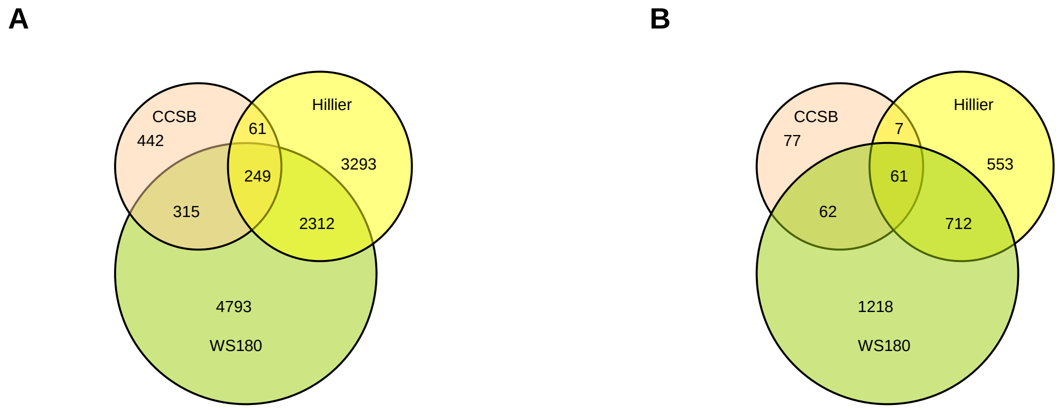
(Supplementa1 Figures 1 - 4, Supplementa1 Tables 1-3)



Supplemental Fig. 1. E-gel electrophoresis of 5’ and 3’ RACE PCR products (*A*) of the positive control set (PCS) (*B*) experimental reference sets (ERS). Worm RNA was reverse transcribed and used as template to generate the first set of RACE products. These amplicons were then re-amplified with the nested internal RACE primers to generate the final RACE products for cloning. (*C*) E-gel electrophoresis of full-length ORF RT-PCR verification products. PCR primers to amply full length ORFs were designed based on the annotated ORF models. Total RNA isolated from mixed stage N2 worms was reverse transcribed and used as template for PCR. PCR products shown in lanes A1 to C10 correspond to confirmed or predicted ORFs from the positive control set; C11 to H11 correspond to ORFs annotated from the experimental reference set (G12 and H12 are negative controls). Altogether, 81 of 84 re-annotated models yielded products (96% overall success).



Supplemental Fig. 2. E-gel electrophoresis of 5’ and 3’ RACE PCR products. Worm RNA was reverse transcribed and used as template to generate the first sets of RACE products. These amplicons were then re-amplified with the nested internal RACE primers to generate the final RACE products for cloning. Usually a positive signal is observed for both 5’ and 3’ RACE.

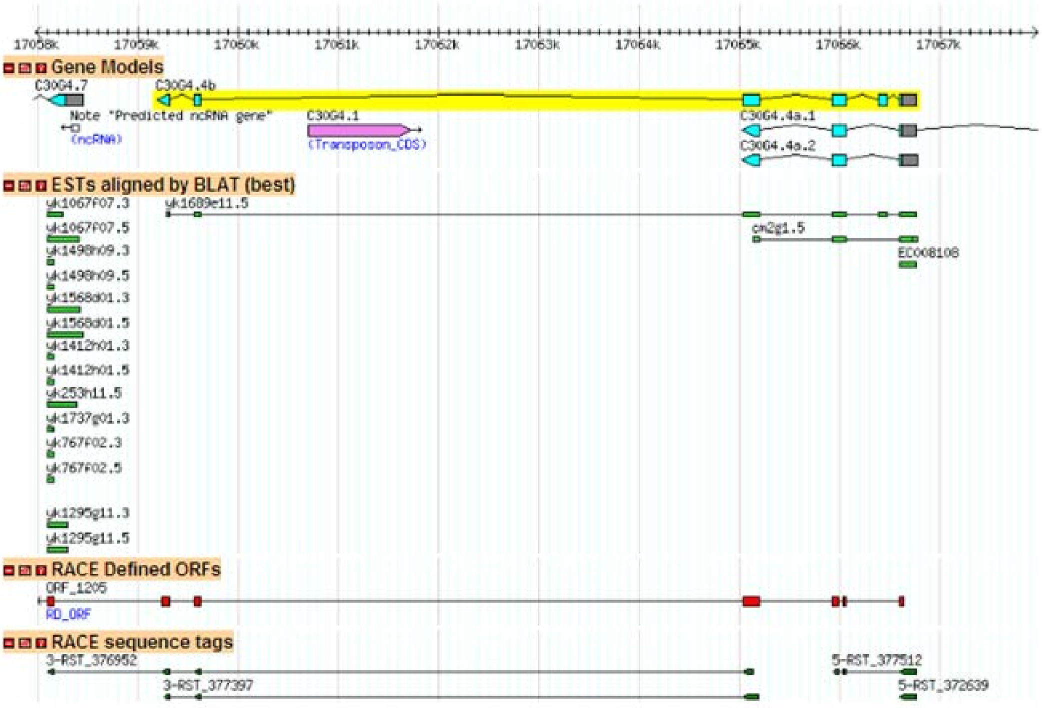


Supplemental Fig. 3. Overlap of SL1 (*A*) and SL2 (*B*) sites between the RACE-defined transcripts in this study, in the mid-L2 stage of *C. elegans* (Hillier et al., 2009) and SL1 and SL2 trans-splice acceptors from Wormbase WS180 annotation. Data from Hillier et al. was downloaded from modENCODE project website (http://www.modencode.org Mid-L2_20dC_14hrs_post-L1.genelets.ws180.gff3 Feb 25, 2009) and entered into WS180 GFF database. We re-aligned our RSTs and RACE-defined ORF (RD-ORF) sequences against WS180 genome using BLAT program, and loaded the alignment results into the WS180 GFF database to make the comparisons.

Supplemental Fig. 4. Representative examples of RACE-defined ORFs (*A*-*H*) with supporting RACE evidence. 5' RSTs, 3' RSTs and RACE-defined ORF models were aligned against WS180 genome. The alignment results were entered into a GFF database that contains WS180 annotations, and then displayed using a local server running WormBase Geneome Browser, along with gene models and ESTs from the annotation. RACE-defined ORFs are shown in red; RSTs, dark green; existing EST evidence when available, light green; WormBase transcript models are in cyan or pink.

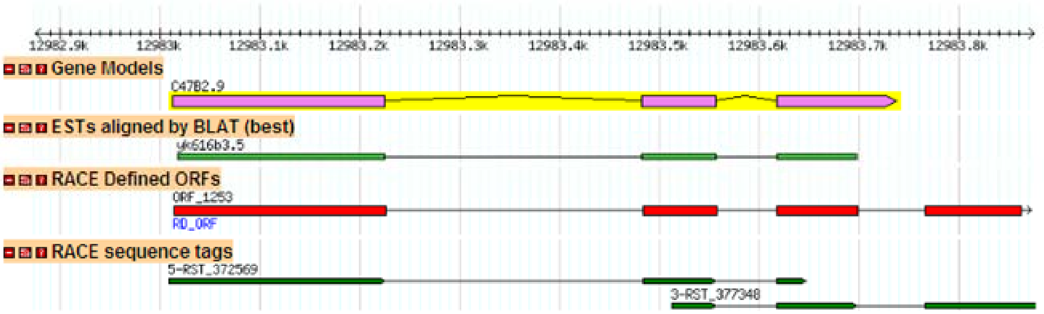
A

C30G4.7



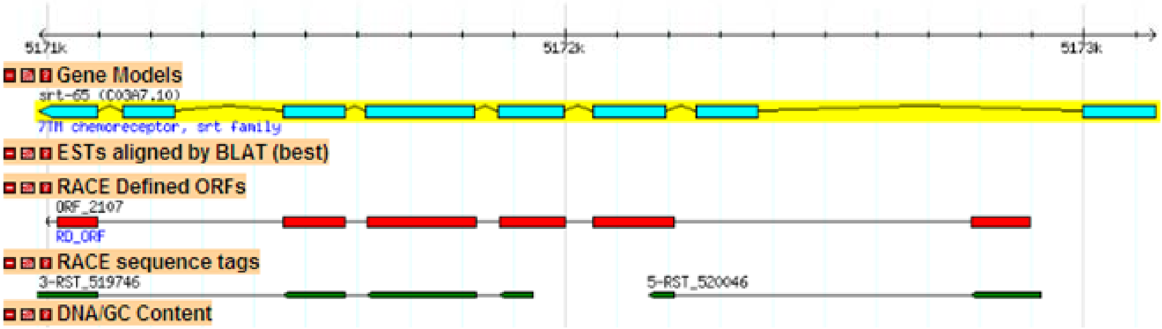
B

C47B2.9



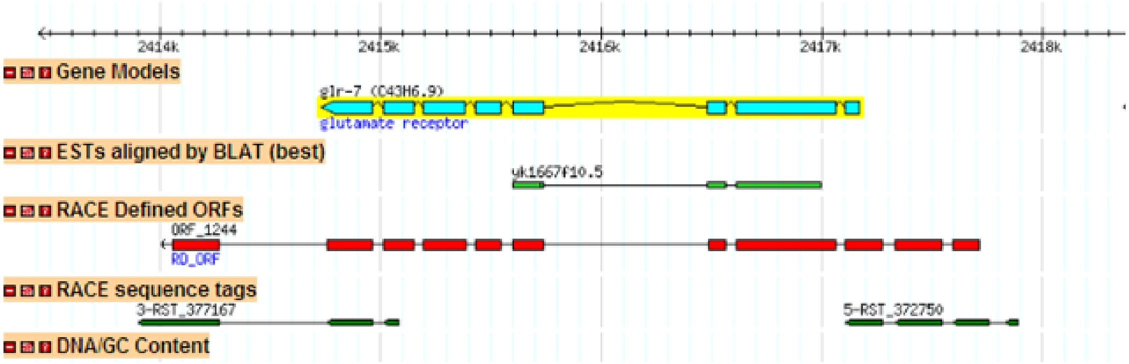
C

C03A7.10



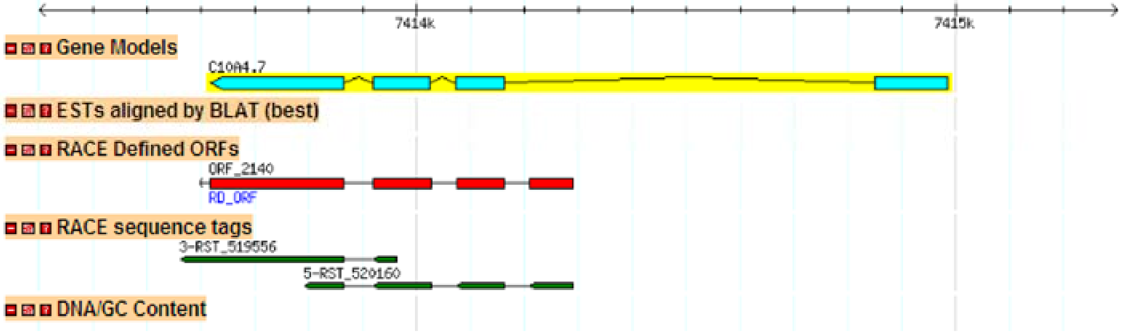
D

C43H6.9



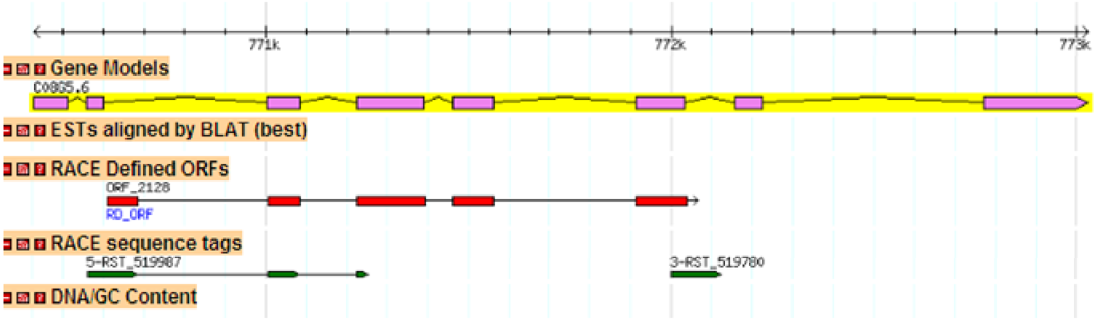
E

C10A4.7



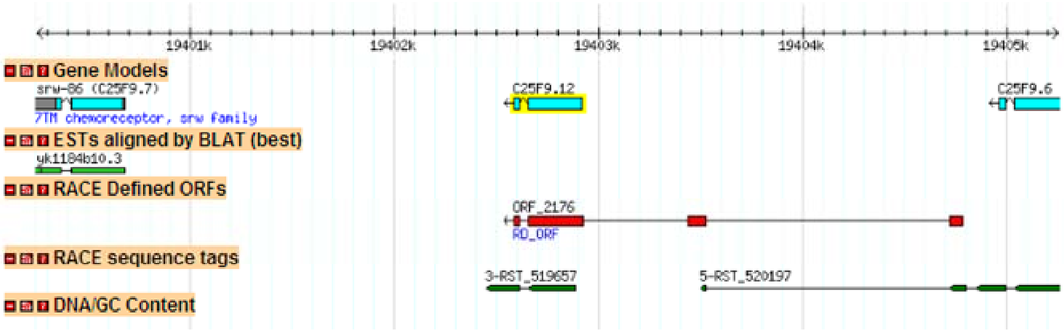
F

C08G5.6



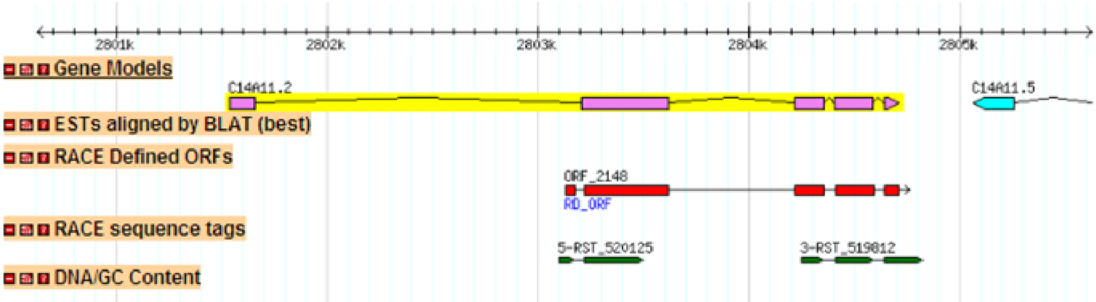
G

C25F9.12



H

C14A11.2



Supplemental Table 1 - Overlap of SL1/SL2 trans-splice sites defined by RACE with those from Hillier et al. Genome Res 2009.

Type	**SL1 count**	**SL2 count**
Match	310	68
Do not match	757	139

Note: Total RST sequences with SL1 =1067, 988 are unique
Total RST sequences with SL2 = 207, 197 are unique
Total Hillier et al. SL1 = 3603, 3585 of these are unique
Total Hillier et al. SL2 = 1333, 1330 of these are unique

Supplemental Table 2. Comparison between new RACE-defined exons listed in Supplemental File 6 (novel exons) and Hillier et al.

Category	**Count**
No overlap	36
Exact match	25
5' different (longer = 2, shorter = 10)	12
3' different (longer = 2, shorter = 3)	5
Both ends different	1
could not be aligned to WS180	4
3' may be shorter, could not be determined	2
Total	85

*Total new RACE exons reported in Supplemental File 6 = 85.

Supplemental Table 3. Overlap comparison between RACE-defined exons listed in Supplemental File 7 (modified exons) and Hillier et al.

Category	**Count**
Undefined	7
No overlap	88
Exact Match	46
Both ends different	3
5' different (longer = 30, shorter = 103)	133
3' different (longer = 12, shorter = 24)	36
Total*	313

*Total modified RACE-defined exons reported in supplemental file 7 = 313;