Supplementary Materials

Supplementary Methods

Genome alignment and definition of bins

The *C. elegans* and *C. briggsae* genomes were aligned with MUMmer 3.19 (Kurtz et al. 2004). In order to have higher genome coverage, a less stringent option, -b 1600 -c 10, was used. The long alignment was split into small bins. Each bin contains 150 aligned columns and every two adjacent bin has 75 overlapping columns. The reason for having overlap between consecutive bins is to increase the chance of having a large portion of each ncRNA gene contained in a bin. The bins at the ends of the alignment with less than 50 columns, and those with more than 100 gaps, were dropped. Subsequently, we obtained 439,815 bins from the plus and minus strands, which together cover 29,655,415 bases of the *C. elegans* genome on both strands.

For each bin, we derived nine sequence, structural and expression features. Using some annotations, we defined some bins as gold-standard examples of four sequence element classes, namely ncRNA, CDS, UTR and unexpressed intergenic. These features and gold-standard examples are detailed below.

Sequence and structural features at DNA, RNA and Protein level

As shown in a previous study (Harmanci et al. 2007), an improved version of the Dynalign program outperformed some other local RNA folding methods, including foldalign, consan and stemloc using some benchmark datasets. Thus, we used Dynalign to predict RNA folding free energy change from the pairwise sequence alignment between *C elegans* and *C briggsae*, where local structural alignment option was used (the gap penalty is not applied to gaps at either end of either sequence). The window size of 150 columns was adopted from a previous study that used Dynalign to predict structured RNA sequences (Uzilov et al. 2006). The bin size was selected to balance the trade-off between the accuracy of RNA secondary structure prediction and computational time. The default parameters of Dynalign were used, where the gap penalty is 0.4 and the single optimal structure is predicted. The free energy of RNA secondary structure was measured by the z-score of RNA folding $\Delta G°_{37}$,

$$z = \frac{\Delta G°_{\text{dynalign}} - <\Delta G°>}{\sigma},$$

where $<\Delta G°>$ is the average free energy change of shuffled sequences with preserved di-nucleotide distribution, and σ is the standard deviation. In order to save computing time, $<\Delta G°>$ and σ were calculated from the A, C, G, and U nucleotide contents and the length of the short sequence in the aligned windows using an SVM model (S. Zhao & D.H. Mathews, unpublished method). The average $\Delta G°_{37}$ z-score of intergenic regions in the gold-standard set was -1.1 with a standard deviation of 2.6. Any bin with a z-score less than -3.7 (Mean – 1S.D.) was defined as a highly-structured region.

RNA secondary structure conservation was measured by SCI, the structure conservation index, between *C. elegans* and *C. briggsae*. SCI is defined according to a previous study (Washietl et al. 2005) as follows:

$$SCI = \frac{\Delta G°_{\text{dynalign}}}{\Delta G°_{\text{seq1}} + \Delta G°_{\text{seq2}}},$$

where $\Delta G°_{\text{dynallign}}$ is the total free energy change of two sequences (sum of two free energy changes) with gap penalty from a structural alignment, and $\Delta G°_{\text{seq1}}$ and $\Delta G°_{\text{seq2}}$ are the free energy changes of the folding of the two individual sequences (Mathews et al. 1999; Mathews et al. 2004). Thus, if SCI is closer to 1, the secondary structures of the two sequences are better conserved.

Selecting a proper RNA secondary structure prediction method is an issue in ncRNA discovery. In this study, we used Dynalign to perform local structural alignment, where the gap penalty is not applied to gaps at either end of either sequence. We chose this method as it was shown to perform well in a benchmark study (Harmanci et al. 2007). We remark that the benchmark in (Harmanci et al. 2007) was based on global structural alignments, but not local structural alignments. Some other benchmarks using different RNA structure data sets have shown that other methods in some cases outperformed Dynalign (Havgaard et al. 2007; Taneda 2008; Taneda 2010). These methods, including the ones that apply a local structural alignment such as FOLDALIGN (Havgaard et al. 2007), could potentially provide a better solution to find local structures of short RNAs and improve the overall identification of ncRNAs.

DNA conservation was defined as the nucleotide identity between *C. elegans* and *C. briggsae* in each aligned bin. We also took the aligned DNA region from each bin and used tblastx (Camacho et al. 2009) to search the protein-level alignment from all 6 frames. The score from the best hit was normalized by the DNA identity to find the protein sequence conservation score.

Expression datasets

Eleven small RNA sequencing, six poly-A+ RNA sequencing and forty-one tiling array datasets were collected from the modENCODE consortium (Gerstein et al. 2010). Small RNA sequencing experiments were conducted in six developmental stages, (embryo, L1, L2, L3, L4 and young adult), a young adult male stage (GEO:GSE13339) (Kato et al. 2009), and four aging stages (Day 0, 5, 8 and 12) post L4 molt (GEO:GSE18634) (Kato and Slack, unpublished). Two sets of poly-A+ RNA sequencing data were produced in N2 strains at embryo and starved L1 stages (GEO: GSE16552) (Zhong et al. 2010), and the remaining four sets were from different developmental stages (L2, L3, L4 and young adult) of N2 strains (Gerstein et al. 2010). The tiling array datasets consist of two types: 29 total RNA arrays and 12 poly-A+ RNA arrays (Table S3 in (Gerstein et al. 2010)). Total RNA tiling array datasets include life stages from early embryo and late embryo to L1, L2, L3, L4, and young adult. There are also datasets from a male sample, pathogenic samples in which worms were grown on plates seeded with *S. marcescens*, *E. faecalis*, and *P. luminescens*, and a sample grown on a plate seeded with the nonpathogenic bacteria OP50, which served as a negative control. The poly-A+ tiling array datasets include gonad only, L2, and a combination of tissue and life stage specific samples in which various neuronal, intestinal, excretory, body wall muscle, and hypodermis cells were extracted from embryo, L2, L3, and L4 worms. All the tiling array data were normalized to the same medium on the oligo level for each sample. All sequencing data were normalized as DCPM (Depth of Coverage Per Million mapped reads), which we adapted from a previous study (Hillier et al. 2009).

Definition of gold-standard set

The gold-standard annotations of CDSs (coding sequences), UTRs (untranslated regions), exonic, intronic and intergenic regions were derived from a published result (Hillier et al. 2009) based on the annotations in Wormbase, and revised by Genefinder (P. Green, unpublished) and Twinscan (Korf et al. 2001). The intergenic regions in the gold-standard set were required to be located at least 200nt upstream from the start of any CDS, and 700nt downstream from the end of any CDS (Hillier et al. 2009) from any confirmed or predicted genes. The CDSs were required to be fully confirmed in Wormbase. The gold-standard annotations of known ncRNAs were selected from Wormbase. We included 158 miRNAs, 19 rRNAs, 94 snRNAs, 4 snlRNAs, 1 scRNAs, and 139 snoRNAs. As there was a large number of tRNAs predicted from tRNAscan-SE (Lowe and Eddy 1997), only a random 10% of them (61 tRNAs) were added to the gold-standard set in order to balance the representations of the different types of ncRNAs. The regions annotated as predicted ncRNAs in Wormbase, as well as small siRNAs such as 21U-RNA, were excluded from the gold-standard set. When these selected known ncRNAs were overlapped with the conserved regions from the pairwise alignments between *C. elegans* and *C. briggsae*, only 219 of them (41 miRNAs, 4 rRNAs, 41 tRNAs, 47 snRNAs, 1 scRNAs, 85 snoRNAs) were covered.

A bin was classified as ncRNA if more than 50% of it overlapped with the known ncRNAs in the gold-standard set. A bin was defined as a CDS (annotated as coding sequence and exon simultaneously), UTR or intergenic region, if more than 90% of the nucleotides of the bin overlapped with the corresponding gold-standard annotations. Bins that overlapped with both known ncRNA and UTR were treated as known ncRNA bins. Since ncRNAs can appear on both DNA strands, we considered the two strands separately, and determined the values of strand-specific features, such as RNA secondary structure stability and small RNA sequencing signals.

The bins annotated as intergenic regions in the gold-standard set were intended to be used as a negative control. Since some intergenic regions could also contain unannotated expressed genomic elements, we used a threshold on the expression values of these bins to filter potential ncRNAs, CDSs and UTRs. There are two competing factors in determining an optimal threshold: purity and coverage. If a low threshold is used, more bins would be filtered and the intergenic class would be purer (with fewer unannotated ncRNAs, CDSs and UTRs), but the included examples would be the more extreme ones, rendering the learned models less informative. We tried different threshold values, using the unit of number of standard deviations from the mean expression of all intergenic bins, as well as a setting that includes all intergenic bins without thresholding (Supplementary Figure 2). As expected, prediction accuracy became lower with larger thresholds, as the gold-standard "intergenic regions" examples contain more unannotated expressed elements, yet the accuracy difference between the different thresholds was only marginal for the better-performing methods. We thus decided to use the mean expression value as the threshold for defining the class of unexpressed intergenic regions, as it appears to be a good tradeoff between the purity and coverage of the real intergenic regions. In all our discussions the "intergenic regions" class refers to the set of intergenic regions derived from Hillier et al. (Hillier et al. 2009) and preprocessed by our alignment procedure, while the "unexpressed intergenic" class refers to the set of such intergenic regions with expression levels lower than the threshold in all expression datasets. The latter set was used in the training of the machine learning models.

We used a permissive annotation set (Hillier et al. 2009) to filter bins that were previously found to have a certain chance of belonging to some genomic element class. In particular, none of our novel ncRNA candidates are allowed to overlap with any ncRNA genes or confirmed, unconfirmed or predicted exons from coding genes in the annotation set. On the other hand, since many known ncRNAs are found in repeat regions, pseudogenes (Sasidharan and Gerstein 2008), introns or at the antisense strand of exons, we allowed novel ncRNA candidates to reside in these regions.

Overall, our gold-standard set contains 659 ncRNA, 34,859 CDS, 8,591 UTR and 6,697 unexpressed intergenic bins.

More details of machine learning methods

For each of the four genomic element classes (i.e. known ncRNA, CDS, UTR and unexpressed intergenic region), 2/3 of the gold-standard examples (412 ncRNA, 23,291 CDS, 5,673 UTR and 4,495 unexpressed intergenic bins) were used for model training and testing (the "cross-validation set"), and the remaining 1/3 (247 ncRNA, 11,568 CDS, 2,918 UTR and 2,202 unexpressed intergenic bins) were used for evaluating the final model (the "independent validation set") (Supplementary Figure 3). Training and testing were performed using 10-fold cross-validation. For each test, 9/10 of the examples in the cross-validation set were used to train a model, and the model was then tested with the remaining 1/10 of the examples, with the accuracy measured by AUC, the area under the receiver-operator characteristic curve. Since there are four classes, each time we set one class as positive and the other three as negative. We thus received four AUC values in each fold. The average of all AUC values in the 10 folds was used to indicate the overall accuracy of a machine learning method.

We repeated the procedure with 5 different machine learning methods (Naive Bayes, Bayes Net, Decision Tree, Random Forest, Logistic Regression, and SVM with linear, polynomial or RBF kernel) using the Weka data mining package (Hall et al. 2009). Random forest models involved 100 decision trees to ensure robustness. SVM models were run using the parameter values "-C 1.0 -N 1 -L 0.001 -P 0.1" with gamma set to the inverse of the number of features. We used the default values for all other algorithm parameters.

The method with the highest average cross-validation accuracy (Random Forest) was picked as the method of choice. This method was then evaluated using the left-out validation set that was not used in the cross-validation process. The average AUC of the four classes on the validation set was used as the final accuracy of the whole procedure.

When a method predicts a genomic region as an ncRNA, we define it as a "true positive" if it is among the ncRNAs in the gold-standard set, and as a "false positive" if otherwise. We define these terms in the machine learning sense for the sake of evaluating prediction accuracy, but we note the possibility that some regions not defined as ncRNAs in the gold-standard set may actually be ncRNAs (e.g. novel genes in the unexpressed intergenic regions).

The basic assumption behind our machine learning work is that we can predict the sequence class of a region because it shares some similarities to known regions in the class. In this sense, it is important for examples in the training and testing sets to be similar, for otherwise the learned statistical models would not be able to make correct predictions. On the other hand, in order for the models to be generally useful to predict the identity of unannotated regions, the evaluation procedure should not be influenced by factors that could artificially raise the prediction accuracy but which cannot be applied in the general setting. For instance, if two overlapping bins have very similar feature values and they belong to the same sequence element class, yet one is in the training set and the other is in the testing set, it is easy to predict the identity of the one in the testing set simply by referencing the one in the training set. In the general setting, such predictions are not very useful. To make sure that our good prediction accuracy was not due to bins in the cross-validation (training/testing) set that overlap with bins in the left-out validation set, we repeated the predictions using a new random split that contains no such cases. We did this by the following procedure: for each bin, we defined a "connected set" that consists of the bin, all bins that overlap with it, all bins that overlap with these bins, and so on. When a bin is randomly chosen to be added to the left-out validation set, all bins in its connected set are also added to the left-out validation set. The resulting prediction accuracy is still very high (AUROC > 0.99), confirming that the good accuracy was not caused by the overlapping bins.

Using incRNA to predict novel ncRNA candidates

We then trained a new Random Forest model using all exmples in the cross-validation set, and applied the model to give an "ncRNA score" for each of the 439,515 bins, which indicates the likelihood that the bin lies in an ncRNA gene. It also gives a CDS score, a UTR score, and an intergenic region score in similar ways.

We found that all known ncRNA bins had predicted ncRNA scores of at least 0.69, while all the other elements had scores of 0.18 at most (Figure 3d). We called these values P_{high} and P_{low}, respectively, and used them as thresholds for defining novel ncRNA candidates. Specifically, we defined each unannotated bin with an ncRNA score of P_{high} or higher as a high-confidence candidate ncRNA bin and each unannotated bin with a ncRNA score between P_{low} and P_{high} as a medium-confidence candidate ncRNA bin (Figure 3e, see also Figure 2a). Altogether, the two sets contain 10,994 bins (covering 1,045,795 bases in total), among which 1,413 are high-confidence predictions and 9,581 are medium-confidence predictions.To estimate the accuracy of these candidate ncRNA bins, we examined the ncNRA scores of the bins in the independent validation set, which were not involved in the whole model training and selection process. We found that all the bins with ncRNA scores higher than P_{high}, and 221 out of 242 bins with ncRNA scores higher than P_{low}, were known ncRNA bins, corresponding to positive predictive values (PPVs) of 100% and 91%, respectively. On the other hand, since there are a total of 247 known ncRNA bins in the validation set, the two thresholds lead to prediction sensitivities of 66% and 89%, respectively.

Furthermore, in order to explore how our method would perform on random sequences, we have constructed a model that involves only sequence features, and applied the model to random sequences sampled from unexpressed intergenic regions not involved in the training process. Using the same procedure to define P_{low} as described in the manuscript, we found that only 0.5% of the unexpressed intergenic regions were predicted as ncRNA candidates. This result shows that considering sequence features alone, very few random sequences would be predicted as ncRNA candidates. The estimated 9% FDR of our full model is mainly due to CDS and UTR bins that have expression and sequence features highly resembling.

Additional conservation information from the multiple sequence alignment of five nematodes

After the training and prediction, we also overlapped the candidate ncRNA bins with a five-way Multiz alignment (Siepel et al. 2005) between *C. elegans, C. brenneri, C. remanei, C. briggsae* and *P. pacificus.* The alignment was downloaded from the UCSC genome browser and the regions with a score less than the median value (0.5) were removed. 94% (10,377 of the 10,994) candidate ncRNA bins were found to overlap with the remaining high-scored alignments (listed in Supplementary File 1).

Details of Northern blot

Approximately 20.0 μg of total RNAs were purified from the late embryonic stage of N2 wild-type animals using the *miRVana* miRNA Isolation Kit (Ambion), and were used for a Northern blot analysis. RNAs were then separated on a 5% TBE-Urea poly acrylamide gel and blotted as described previously (Esquela-Kerscher et al., DevDyn. 234:868–877, 2005). 25 nucleotide oligo-DNAs corresponding to parts of ncRNA regions, and validated by RT-PCR, were selected based on non-homology to any other genomic regions, and obtained from IDT (Integrated DNA Technologies). These oligo-DNAs were labeled using the StarFire labeling kit (IDT) according to the manufacturer's instructions, and were used as probes in the hybridization. Additionally, the hybridization process was further repeated in an independent experiment using the probes of approximately 120 nucleotide RT-PCR products (Supplementary Table 2), which were labeled by the general random prime method. Five out of fifteen ncRNA candidates (used for RT-PCR) were used for Northern plot. The sequences of IDT StarFire probes were as follows (also see details of all 15 ncRNA candidates in Supplementary Table 2):

ncRNA4: ACCGAGCTTCCCCTAGTGTCCAGTA;
ncRNA5: GAGGATAGTGGACACTGTTTGTGAT;
ncRNA6: GTACTAGTGACCTGATGCGACGAAT;
ncRNA7: GTCTGCGGGTCTCTGATCCTACTAC;
ncRNA8: TTTCACCTTCTCGCGGTCACTTTCC.

Categories of candidate ncRNA bins based on existing annotation

Among the 10,994 candidate ncRNA bins, 41 are overlapped with unconfirmed non-coding RNA annotations in Wormbase, 1,172 are antisense to annotated ncRNA regions, 68 are overlapped with pseudogenes, 1,479 bins are overlapped with introns at the sense strand (494 of them are confirmed intronic regions, annotated as intronic_gold), 2,966 are overlapped with introns at the anti-sense strand and 1,966 are antisense to exons. 883 bins of the remainder are intergenic but close to CDSs. As a result, only 2,469 candidate ncRNA bins (merged into 1,678 ncRNA fragments) are located inside the intergenic regions of our gold standard annotations, as previously defined (annotated as intergenic_gold). The ncRNAs in all these different categories are labeled differently in Supplementary File 1. We refer to the 1,678 ncRNA fragments as intergenic novel ncRNA candidates in this paper, and list them separately in Supplementary File 2. These 1,678 fragments come from 1,223 genomic loci after strands are dissolved. In addition to labeling these genomic locations, 2,271 ncRNA bins (374 of them are intergenic), overlapping with repeated regions, are also labeled in Supplementary File 1. Therefore, about 20% of our novel ncRNA candidates (~15% of intergenic novel ncRNA candidates) are repeat-associated.

Fine-ranking of novel ncRNA candidates based on multiple information

In order to accommodate different potential uses of our predictions and the selection of candidates for further study, we ranked the 10,994 candidate ncRNA bins (merged into 7,237 fragments) into 9 levels (from -3 to 5) using supplemental information in addition to our predicted ncRNA scores. The default level is 0. A bin gets an increment of one level for each match with the following conditions:

- It is predicted as a high-confidence candidate ncRNA bin

- It overlaps with a conserved region from multiple-species alignment (conservation score>0.5)
- It has a nearby POL II binding site in at least one stage
- It has a nearby transcription factor binding site in at least one stage
- It is defined as an intergenic candidate ncRNA bin, using our gold standard annotation

A bin is docked one level for each match with the following conditions:

- It is overlapped with an intron, a pseudogene, or is antisense to an exon
- It is overlapped with a repeat region

A bin is docked two levels if it matches the following condition:

- It is antisense to an annotated ncRNA (most are predicted and unconfirmed)

The level of a candidate ncRNA fragment (merged bins) is the average level of all the candidate ncRNA bins inside it (Supplementary File 1).

Legends of Supplementary Figures

Supp. Figure 1. (a) Prediction accuracy of the chosen machine learning methods at different thresholds for defining the unexpressed intergenic region class of the gold-standard set. In each case, members of the class are defined as bins with an expression value lower than a certain standard deviation (SD) from the mean in all four expression features. The rightmost column corresponds to using all intergenic regions without thresholding. The resulting number of bins in each case is shown in parentheses. **(b)** The prediction accuracies of RNAz were compared between the pairwise alignment from MUMmer and the five-way alignment from MultiZ. To make a fair comparison, RNAz was applied to the same regions appearing in both kinds of alignments to predict ncRNAs. We used the default parameters of RNAz, except that we used the same window size (150 aligned columns) as the one we used in our machine learning model. In general, the prediction performance of RNAz for both kinds of alignments is similar in regions with high sequence identity values.

Supp. Figure 2. Distributions of feature values for different types of known ncRNAs. The distributions of the values of the 9 features are shown for 6 types of ncRNAs (bins). **(a)** Box plots of individual features. **(b)** Two-dimensional scatter-plot of the maximum small RNA-seq signal against the maximum poly-A+ RNA-seq signal. **(c)** Two-dimensional scatter-plot of the predicted protein sequence conservation against DNA conservation.

Supp. Figure 3. Numbers of bins of the four classes of genomic elements in our gold-standard set, and the sizes of our four sets of predictions for the originally unannotated bins.

Supp. Figure 4. The length distributions of candidate ncRNA fragments and known ncRNA transcripts.

Supp. Figure 5. Northern blot of five novel ncRNA candidates and multiple signal tracks of two novel ncRNA candidates. (**a**) Five ncRNA candidates were selected (labeled in Supplementary Table 2, out of 15 candidates for RT-PCR) for a Northern blot assay. These five candidates are manually selected because of strong signals on RT-PCR gels. Three candidates (ncRNA 4, 7 and

8) were detected. One (ncRNA 4) is between 100 and 200 nt, which could be processed from a longer transcript. The others (ncRNA 7 and 8) are larger than 500 nt, which could be the precursors of shorter products. Ethidium bromide staining of rRNAs is shown as a loading control. (**b,c**) Two ncRNA candidates are located within transcribed regions supported by multiple signals. The heights of PHA-4 binding, POL II binding and input signals are normalized by their total numbers of mapped reads from the corresponding ChIP-seq experiments. The values of tiling array are the log2-transformed and normalized signals. The values of RNA sequencing are the numbers of mapped reads at every single nucleotide.

Supp. Figure 6. Structural properties of the intergenic novel ncRNA candidates. The percentages of highly structured ncRNAs are shown for the high-confidence and medium-confidence candidate ncRNA bins from intergenic regions (see Methods). Among the highly structured bins, the percentages that overlap with structural homologues with Rfam families are also shown.

Supp. Figure 7. Comparison of feature values of known ncRNA bins, high-confidence candidate ncRNA bins and medium-confidence candidate ncRNA bins. **(a)** Box plots of individual features. **(b)** Two-dimensional scatter-plot of the predicted secondary structure free energy against DNA conservation. **(c)** Two-dimensional scatter-plot of the maximum poly-A+ RNA tiling array signal against the predicted secondary structure conservation.

Supp. Figure 8. Saturation plot of expressed candidate ncRNA bins in different expression datasets. The fractions of expressed regions (with expression signals stronger than the average signal of gold-standard intergenic regions) are computed using random samples of all combinations of the 58 RNA-seq and tiling array datasets. The x-axis corresponds to the number of datasets considered, and each point at a given number of datasets corresponds to a different combination of datasets.

Supp. Figure 9. Binding signals of POL II and 22 transcription factors around their genomic regions. (a) and (b) Fractions of intergenic candidate ncRNA fragments potentially targeted by (a) POL II and (b) 22 transcription factors in a total of 27 experiments. The total fractions targeted by POL II and any of the transcription factors in any of the stages are also shown. In (b), each bar is labeled by the name of the transcription factor followed by the stage at which the binding experiment was performed. The bindings on random genome locations with the same size are also shown in (a) and (b). Abbreviations: EMB – embryo; YA – young adult.

Supplementary Tables

Supp. Table 1a. Performance of *incRNA* on the bins derived from the genome alignment between *C. elegans* and *C. briggsae*, including three different ways to define element classes in the gold-standard set

Class definition 1		Class definition 2		Class definition 3	
Element class	AUC[*]	Element class	AUC[*]	Element class	AUC[*]
Known ncRNA	0.9860	Known ncRNA	0.9732	Known ncRNA	0.9692
CDS	0.9992	CDS	0.9730	CDS	0.9715
Unexpressed intergenic region	0.9997	3’ UTR	0.9375	5’ and 3’ UTR	0.9304
		Unexpressed intergenic region	0.9982	Unexpressed intergenic region	0.9987

[*] AUC is the area under the receiver operator characteristics curve for each class.

Supp. Table 1b. Performance of *incRNA* binned in relation to sequence identity and GC content

GC%	DNA identity	AUC
Low	Low	0.9747
Low	High	0.9660
High	Low	0.9745
High	High	0.9808

We have binned our independent validation set and evaluated the prediction accuracy of the examples at different levels of sequence identity and GC content. Specifically, we divided sequence identity values and GC content values each into two levels (lower than median, higher than median), forming a total of 4 combinations. For each combination, we took positive (ncRNA) examples in the independent validation set with this combination of sequence identity and GC content, and all negative examples, to compute a prediction accuracy of the model.

Supp. Table 2. The fifteen novel ncRNA candidates tested by RT-PCR in late embryo

Name	Chr.	Start [a]	End	PCR length [b]	ncRNA score [c] (plus strand)	ncRNA score (minus strand)	Conservation Score[d]
ncRNA1[f]	chrI	556919	557049	107	0.2	0.2	0.65
ncRNA2	chrI	7611289	7611425	100	0.3	0.3	0.53
ncRNA3[f]	chrII	3930125	3930260	125	0.2	0.2	0.64
ncRNA4[e]	chrII	7569982	7570327	166	0.3	0.3	0.66
ncRNA5[e]	chrII	8750286	8750829	146	0.3	0.3	0.56
ncRNA6[e]	chrII	8750877	8751148	120	0.4	0.4	0.62
ncRNA7[e]	chrIII	5964548	5964705	124	0.2	0.2	0.55
ncRNA8[e]	chrIV	3851506	3851664	126	0.3	0.3	0.50
ncRNA9	chrIV	9196924	9197103	125	0.4	0.4	0.54
ncRNA10	chrV	14505342	14505801	136	0.2	0.2	0.61
ncRNA11	chrV	11525384	11525537	122	0.6	0.6	0.54
ncRNA12	chrV	14504822	14505114	127	0.3	0.3	0.85
ncRNA13	chrV	6879489	6879644	120	0.2	0.2	0.57
ncRNA14	chrX	5818466	5818828	132	0.4	0.4	0.72
ncRNA15	chrX	3834195	3834401	130	0.4	0.4	0.50

[a] The coordinates of candidate ncRNA TARs (Wormbase 170).

[b] The lengths of PCR products are determined by the distance between the 5' and 3' primers.

[c] If multiple candidate ncRNA bins are overlapped with a TAR, the ncRNA score for the TAR is the maximum score among them.

[d] The maximum score from overlapped 5-way Multiz alignments (*C. elegans, C. brenneri, C. remanei, C. briggsae* and *P. pacificus*).

[e] The five novel ncRNA candidates with strongest signals on RT-PCR gel are selected manually for a Northern Blot assay.

[f] These two ncRNA candidates were overlapped with repeat and inverted repeat regions.

Supp. Table 3. Summary of feature values for four of our prediction sets

	High-confidence novel ncRNA candidates [a]	Medium-confidence novel ncRNA candidates [b]	CDS-like ambiguous regions [c]	UTR-like ambiguous regions [c]
DNA Conservation	High	Medium	High	Low
Protein Conservation	No	No	Yes	No
RNA Secondary Structure Conservation	Yes	Yes	Yes	No
RNA Secondary Structure Free Energy	Low	Low	Medium	Medium
Poly-A+ RNA-seq	Low	Low	High	Medium
Small RNA-seq	High	Medium	Medium	Medium
Poly-A+ RNA Tiling Array	Low	Low	High	Medium
Total RNA Array Tiling Array	Low	Low	High	Medium

[a] Bins predicted by Random Forest to be novel ncRNAs with ncRNA scores at least P_{high}

[b] Bins predicted by Random Forest to be novel ncRNAs with ncRNA scores between P_{low} and P_{high}

[c] Bins predicted by Random Forest to have ncRNA scores lower than P_{low}, and which are relatively similar to CDSs or UTRs

Supp. Table 4. Statistics of expression pattern classes in terms of high-confidence and medium-confidence candidate ncRNA bins

	All	Universal expression class		Differential expression class		Undetectable expression class	
	# of bins[1]	# of bins	%	# of bins	%	# of bins	%
High-confidence candidate ncRNA bins	1,413 (171)[2]	67 (13)	4.7% (7.6%)	594 (98)	42.0% (57.3%)	752 (60)	53.2% (35.1%)
Medium-confidence candidate ncRNA bins	9,581 (2,298)	79 (8)	0.8% (0.3%)	6,164 (1,528)	64.3% (66.5%)	3,338 (762)	34.8% (33.2%)

The expression values are determined by the normalized small RNA sequencing data only.

[1] 10,994 candidate ncRNA bins predicted from the genome alignment between *C. elegans* and *C. briggsae*.

[2] 2,469 candidate ncRNA bins are inside intergenic regions (gold standard set). The values in parentheses are calculated for intergenic candidate ncRNA bins.

Supp. Table 5. Nomenclature of known and predicted ncRNAs

	Genes	Bins		Merged bins	
Known ncRNAs (Gold-standard set)	219[1] Known ncRNA transcripts	659[1] Known ncRNA bins		na	
Novel ncRNA candidates (7K-set)	na	10,994 Candidate ncRNA bins		**7,237** Candidate ncRNA fragments	
		1,413 High-confidence	9,581 Medium-confidence	1,678[2] Intergenic	5,559 Non-intergenic

[1] Gold-standard set of ncRNAs inside conserved regions from *C. elegans* and *C briggsae* alignment.

[2] 730 of them overlapped with longer TARs (>100nt).

Supplementary File 1

Prediction scores, structural features, sequence features and expression values for candidate ncRNA fragments/bins.

Content of each column:

1.ID: Candidate ncRNA fragment ID.

2.Rank_of_frag: Rank of each candidate ncRNA fragment averaged from bins.

3.Rank_of_bin: Rank of each candidate ncRNA bin decided by genomic location, predicted ncRNA score from integrative model, binding of POL II and transcription factors, overlapping with repeat regions and DNA conservation of five nematodes (see methods).

4-7. Novel ncRNA candidates' coordinates (chromosome, start, end, strand) at *C. elegans* genome (Wormbase 170).

8. Bin_position: Bin number and coordinate (chromosome, start, end, strand) of each fragment, which is merged from overlapped small bins.

9-11: Our prediction scores: Probability of being ncRNA or coding sequence (CDS or coding exon) or UTR from our machine learning method (Random Forest); range: 0-1.

12.Confidence: annotation of high-confidence candidate ncRNA bin (ncRNA score>=0.69) and medium confidence candidate ncRNA bin (ncRNA score>=0.19&&<0.69).

13.Rfam: Predicted secondary structure family from Rfam/INFERNAL.

14.Genomic location: Annotated type using gold-standard annotations and permissive annotations.

15.Overlapped_repeat: Overlapped repeat regions

16.Overlapped_exonic_Annotation(Sense): Overlapped exonic annotations at sense strand

17.Overlapped_intronic_Annotation(Sense): Overlapped intronic annotations at sense strand

18.Overlapped_exonic_Annotation(Antisense): Overlapped exonic annotations at antisense strand

19.Overlapped_intronic_Annotation(Antisense): Overlapped intronic annotations at antisense strand

20.length: Length of each bin.

21.GC%: GC content for each bin.

22.DNA_identities: DNA identity when aligned with *C. briggsae*; range: 0-1.

23.Overlapped_with_multiz5way_conservation: Conserved regions from 5 way Multiz alignments with scores.

24.RNA_secondary_structure_zscore: RNA secondary structure stability z-score calculated from Dynalign for each bin; negative score is favored.

25.RNA_secondary_structure_SCI: Structure Conservation Index for RNA secondary structure for each bin.

26.tblastx_score_scaled: tblastx score for 6-frame protein-translation alignment and it is normalized with DNA identity.

27.polyA_RNAseq_max_all: Maximum value of poly-A+ RNA-seq (DCPM) from all 6 poly-A+ RNA-seq samples.

28.small_RNAseq_max_all: Maximum value of small RNA-seq (DCPM) from all 11 small RNA-seq samples.

29.Array_max_all: Maximum value of 41 tiling array samples. It is the log2 values of a medium normalized signal.

30.Array_max_totalRNA: Maximum value of 27 total RNA tiling array samples. It is the log2 values of a medium normalized signal.

31.Array_max_polyA: Maximum value of 12 poly-A+ RNA tiling array samples. It is the log2 values of a medium normalized signal.

32-37: Poly-A+ RNA-seq values (DCPM) of 6 poly-A+ RNA-seq samples.

38-48: Small RNA-seq values (DCPM) of 11 small RNA-seq samples.

49-77: Expression values of 27 total RNA tiling array samples. It is the log2 values of a medium normalized signal.

78-89: Expression values of 12 poly-A+ RNA tiling array samples. It is the log2 values of a medium normalized signal.

90.RNAz: ncRNA prediction score of a previously published program RNAz; range:0-1000.

91.Expression_Class: Three classes from expression pattern analysis (see main text).

Supplementary File 2

Intergenic candidate ncRNA loci/fragments targeted by POL II and transcription factors across different developing stages.

References

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Gerstein MB Lu ZJ Van Nostrand EL Cheng C Arshinoff BI Liu T Yip KY Robilotto R Rechtsteiner A Ikegami K et al. 2010. Integrative Analysis of the Caenorhabditis elegans Genome by the modENCODE Project. *Science* **330**(6012): 1775-1787.

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations* **11**(1).

Harmanci AO, Sharma G, Mathews DH. 2007. Efficient pairwise RNA structure prediction using probabilistic alignment constraints in Dynalign. *BMC Bioinformatics* **8**: 130.

Havgaard JH, Torarinsson E, Gorodkin J. 2007. Fast pairwise structural RNA alignments by pruning of the dynamical programming matrix. *PLoS Comput Biol* **3**(10): 1896-1908.

Hillier LW, Reinke V, Green P, Hirst M, Marra MA, Waterston RH. 2009. Massively parallel sequencing of the polyadenylated transcriptome of C. elegans. *Genome Res* **19**(4): 657-666.

Kato M, de Lencastre A, Pincus Z, Slack FJ. 2009. Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during Caenorhabditis elegans development. *Genome Biol* **10**(5): R54.

Korf I, Flicek P, Duan D, Brent MR. 2001. Integrating genomic homology into gene structure prediction. *Bioinformatics* **17 Suppl 1**: S140-148.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* **5**(2): R12.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**(5): 955-964.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* **101**: 7287-7292.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA Secondary Structure. *J Mol Biol* **288**: 911-940.

Sasidharan R, Gerstein M. 2008. Genomics: Protein fossils live on as RNA. *Nature* **453**(7196): 729-731.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**(8): 1034-1050.

Taneda A. 2008. An efficient genetic algorithm for structural RNA pairwise alignment and its application to non-coding RNA discovery in yeast. *BMC Bioinformatics* **9**: 521.

Taneda, A. 2010. Multi-objective pairwise RNA sequence alignment. *Bioinformatics* **26**(19): 2383-2390.

Uzilov AV, Keegan JM, Mathews DH. 2006. Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. *BMC Bioinformatics* **7**(1): 173.

Washietl S, Hofacker IL, Stadler PF. 2005. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci USA* **102**(7): 2454-2459.

Zhong M, Niu W, Lu ZJ, Sarov M, Murray JI, Janette J, Raha D, Sheaffer KL, Lam HYK, Preston E et al. 2010. Genome-Wide Identification of Binding Sites Defines Distinct Functions for <italic>Caenorhabditis elegans</italic> PHA-4/FOXA in Development and Environmental Response. *PLoS Genet* **6**(2): e1000848.

Details of RT-PCR and sequencing validation for 15 novel ncRNA candidates

The candidates picked for validation are listed below, with their chromosome's start and end locations, TAR sizes, primer sequences and product size. The coordinates are based on Wormbase WS170.
Each RT-PCR product was sequenced (W.M. Keck Facility) using forward and reverse primers separately. The low quality ends were truncated and the rest of the sequences were aligned with the whole candidate ncRNA TARs derived from the reference genome. The whole length of the PCR product can be mapped accurately by combining the two sequencing results from both primers.

>ncRNA1 I:556919,557049: 131bp
tttgttcaggattttaggaatttctgcgaccttctcactcatgtcctccagccccgcctaagcctatgccttaactcaagcctaagcctaagcctaagcctaacctaaatcgcgtcagagataacgttcgc

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	CTGCGACCTTCTCACTCATGT	Plus	21	24	44	54.27	52.38%
Reverse primer	CGAACGTTATCTCTGACGCGA	Minus	21	130	110	54.91	52.38%

Internal oligo: Plus
Product length: 107

Sequencing result:

Forward primer

```
TTTGTTCAGGATTTTAGGAATTTCTGCGACCTTCTCACTCATGTCCTCCAGCCCCGCCTA
------------------------------------------------------------


AGCCTATGCCTTAACTCAAGCCTAAGCCTAAGCCTAAGCCTAACCTAAATCGCGTCAGAG
---------GCTTACTCA-GCCTAAGCCTAAGCCTAAGCCTAACCTAAATCGCGTCAGAG
           * ***** *****************************************


ATAACGTTCGC
ATAACGTTCG-
**********
```

Reverse primer

```
TTTGTTCAGGATTTTAGGAATTTCTGCGACCTTCTCACTCATGTCCTCCAGCCCCGCCTA
```

```
-----------------------CTGCGACCTTCTCACTCATGTCCTCCAGCCCCGCCTA
                       *************************************

AGCCTATGCCTTAACTCAAGCCTAAGCCTAAGCCTAAGCCTAACCTAAATCGCGTCAGAG
AGCCTATGCCT-AACTCAAGCCTAANC---------------------------------
*********** ************* *

ATAACGTTCGC
-----------
```

```
>ncRNA2        I:7611289,7611425: 137bp
cggtttatttgtccttttgtccctctattttcttcttctctgttttgcatctctcttaggtatctagaccctatccaccaattccttatattgcccaggttattt
tcagtttttttttcgttttgaaatgagtcatc
```

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	CGGTTTATTTGTCCTTTTGTCCCT	Plus	24	1	24	53.78	41.67%
Reverse primer	ACCTGGGCAATATAAGGAATTGGT	Minus	24	100	77	53.96	41.67%

Internal oligo: Plus
Product length: 100

Sequencing result:

Forward primer

```
CGGTTTATTTGTCCTTTTGTCCCTCTATTTTCTTCTTCTCTGTTTTGCATCTCTCTTAGG
-------------------------------------------------TCTCTCTTAGG
                                                 ***********

TATCTAGACCCTATCCACCAATTCCTTATATTGCCCAGGTTATTTTCAGTTTTTTTTTCG
TATCTAGACCCTATCCACCAATTCCTTATATTGCCCAGG---------------------
***************************************

TTTTGAAATGAGTCATC
-----------------
```

Reverse primer

```
CGGTTTATTTGTCCTTTTGTCCCTCTATTTTCTTCTTCTCTGTTTTGCATCTCTCTTAGG
-GGTTTATTTGTCCTTTTGTCCCTCTATTTC--TCTTCTCNNTNTGCACTCTNTAGT---
 *****************************   *******  * *    *** *  *

TATCTAGACCCTATCCACCAATTCCTTATATTGCCCAGGTTATTTTCAGTTTTTTTTTCG
------------------------------------------------------------

TTTTGAAATGAGTCATC
-----------------
```

>ncRNA3 II:3930125,3930260: 136bp
aaaagcctaagcttctgcctaaaggcctaagcctaagcctgagcctaagcctaagcatagcctaagcctaagcctaagcataaccaagcct
aaatagctaacgctcgccactgacgccaaagcctaagcccaagac

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	AAAGCCTAAGCTTCTGCCTAAAG	Plus	23	2	24	53.07	43.48%
Reverse primer	TAGGCTTTGGCGTCAGTGG	Minus	19	126	108	54.24	57.89%

Internal oligo: Plus
Product length: 125
Sequencing result:

Forward primer

```
AAAAGCCTAAGCTTCTGCCTAAAGGCCTAAGCCTAAGCCTGAGCCTAAGCCTAAGCATAG
---------------------------------------------CTAGCCTAAGCATAG
                                               *************


CCTAAGCCTAAGCCTAAGCATAACCAAGCCTAAATAGCTAACGCTCGCCACTGACGCCAA
CCTAAGCCTAAGCCTAAGCATAACCAAGCCTAAATAGCTAACGCTCGCCNCNGACGCNAA
************************************************* * ***** **


AGCCTAAGCCCAAGAC
AGCCT-----------
*****
```

Reverse primer

```
AAAAGCCTAAGCTTCTGCCTAAAGGCCTAAGCCTAAGCCTGAGCCTAAGCCTAAGCATAG
-AAAGCCTAAGCTTCTGCCTAAAGGCCTAAGCCTAAGCCTGAGCCTAAGCCTAAGCATAG
 ***********************************************************


CCTAAGCCTAAGCCTAAGCATAACCAAGCCTAAATAGCTAACGCTCGCCACTGACGCCAA
CCTAAGCCTAAGCCTAAG------------------------------------------
******************


AGCCTAAGCCCAAGAC
----------------
```

>ncRNA4   II:7569982,7570327: 346bp

actattaccacttcccgctccactattggggcaccaaatttgtttgcacccactcgaccgagcttcccctagtgtccagtaattaagctcttgttg
gccctcgcccatcatttttcttggccttcatcaaaatgaaaggagttttgactttttatattcaagtcaagaagctactctcactactactgcacatt
cataaaaacagattgaatttcgcgcgcacggctcccttcgtctccttcttgagccaccggactaatttgtactaccgccgtctcactttccacgc
tgccacgacgttctaattgaaagtgacctatcagctttcaaagcagcttttccgcgt

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	ACCGAGCTTCCCCTAGTGTC	Plus	20	57	76	55.08	60.00%
Reverse primer	TGCGCGCGAAATTCAATCTG	Minus	20	222	203	54.94	50.00%

Internal oligo: Plus

Product length: 166

Sequencing result:

Forward primer

```
ACTATTACCACTTCCCGCTCCACTATTGGGGCACCAAATTTGTTTGCACCCACTCGACCG
------------------------------------------------------------


AGCTTCCCCTAGTGTCCAGTAATTAAGCTCTTGTTGGCCCTCGCCCATCATTTTTCTTGG
---------------------------------------CTCNNNNN-CATTTTTCTTGG
                                       ***      ************

CCTTCATCAAAATGAAAGGAGTTTTGACTTTTTATATTCAAGTCAAGAAGCTACTCTCAC
CCTTCATCAAAATGAAAGGAGTTTTGACTTTTTATATTCAAGTCAAGAAGCTACTCTCAC
************************************************************

TACTACTGCACATTCATAAAAACAGATTGAATTTCGCGCGCACGGCTCCCTTCGTCTCCT
TACTACTGCACATTCATAAAAACAGATTGAATTTCCCGCGC-------------------
*********************************** *****
```

```
TCTTGAGCCACCGGACTAATTTGTACTACCGCCGTCTCACTTTCCACGCTGCCACGACGT
------------------------------------------------------------

TCTAATTGAAAGTGACCTATCAGCTTTCAAAGCAGCTTTTCCGCGT
----------------------------------------------
```

Reverse primer

```
ACTATTACCACTTCCCGCTCCACTATTGGGGCACCAAATTTGTTTGCACCCACTCGACCG
------------------------------------------------------CTTCCG
                                                      *  ***

AGCTTCCCCTAGTGTCCAGTAATTAAGCTCTTGTTGGCCCTCGCCCATCATTTTTCTTGG
AGCTTCCCCTAGTGTCCAGTAATTAAGCTCTTGTTGGCCCTCGCCCATCATTTTTCTTGG
************************************************************

CCTTCATCAAAATGAAAGGAGTTTTGACTTTTTATATTCAAGTCAAGAAGCTACTCTCAC
CCTTCATCAAAATGAAAGGAGTTTTGACTTTTTATATTCAAGTCAAGAAGNT-CTCTCAT
************************************************** * ******

TACTACTGCACATTCATAAAAACAGATTGAATTTCGCGCGCACGGCTCCCTTCGTCTCCT
------------------------------------------------------------


TCTTGAGCCACCGGACTAATTTGTACTACCGCCGTCTCACTTTCCACGCTGCCACGACGT
------------------------------------------------------------


TCTAATTGAAAGTGACCTATCAGCTTTCAAAGCAGCTTTTCCGCGT
----------------------------------------------
```

>ncRNA5 II:8750286,8750829: 544bp

tctgttactctacctaaccgtattatccgctctccatgtaaaactactgaagtagacacagcaagggacaggcacacactcactggtcatcaaa
atcatcaaatgatgcagtgaatttggctgttctggtttgcagacaaaaaagttttggaagcggtgcgggaggatagtggacactgtttgtgattc
acttctcgtcattctccgagccttacaaacaaacaaaaatagctctcttttcgctgacgattgtcccccaacacgggcggcggttattttggct
ctcgccactctttttcaaggtgcacaatcaaaaacaaacaagtccagcccgatttcgcgaatttgttcatctagttcaacgcatttttttcattttga
atacatttgcttgtatcgtttccgagtttcttgaacaattactttcagaattgagtatcatttttgtgttcagcagtgaaaaataaccaaaatttctcta
tcaggcttgcactcacattggtcatattttacagtcctgacctagagtgcgtttatgaat

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	GACAGGCACACACTCACTGG	Plus	20	67	86	55.14	60.00%
Reverse primer	AGGCTCGGAGAATGACGAGA	Minus	20	212	193	54.54	55.00%

Internal oligo: Plus

Product length: 146

Sequencing result:

Forward primer

```
TCTGTTACTCTACCTAACCGTATTATCCGCTCTCCATGTAAAACTACTGAAGTAGACACA
------------------------------------------------------------


GCAAGGGACAGGCACACACTCACTGGTCATCAAAATCATCAAATGATGCAGTGAATTTGG
--------------------------------------------------GNTGATTTGG
                                                  *   ******

CTGTTCTGGTTTGCAGACAAAAAAGTTTTGGAAGCGGTGCGGGAGGATAGTGGACACTGT
------TGNNTGGTTGAGACAAAAATTTTGGAAGCGGTGCGGGAGGATAGTGGACACTGT
      **  * *  ** * **** ***********************************

TTGTGATTCACTTCTCGTCATTCTCCGAGCCTTACAAACAAACAAAAATAGCTCTCTCTT
TTGTGATTCACTTCTCGTCATTCTCCGAGCCTCG--------------------------
********************************
```

```
TTCGCTGACGATTGTCCCCCAACACGGGCGGCGGTTATTTTGGCTCTCGCCACTCTTTTT
------------------------------------------------------------


CAAGGTGCACAATCAAAAACAAACAAGTCCAGCCCGATTTCGCGAATTTGTTCATCTAGT
------------------------------------------------------------


TCAACGCATTTTTTTCATTTTGAATACATTTGCTTGTATCGTTTCCGAGTTTCTTGAACA
------------------------------------------------------------


ATTACTTTCAGAATTGAGTATCATTTTTTGTGTTCAGCAGTGAAAAATAACCAAAATTTC
------------------------------------------------------------


TCTATCAGGCTTGCACTCACATTGGTCATATTTTACAGTCCTGACCTAGAGTGCGTTTAT
------------------------------------------------------------

GAAT
----
```

Reverse primer

```
TCTGTTACTCTACCTAACCGTATTATCCGCTCTCCATGTAAAACTACTGAAGTAGACACA
------------------------------------------------------------

GCAAGGGACAGGCACACACTCACTGGTCATCAAAATCATCAAATGATGCAGTGAATTTGG
------GACAGGCACACACTCACTGGTCATCAAAATCATCAAATGATGCAGTGAATTTGG
      ******************************************************
```

```
CTGTTCTGGTTTGCAGACAAAAAAGTTTTGGAAGCGGTGCGGGAGGATAGTGGACACTGT
CTGTTCTGGTTTGCAGACAAAAAAGTTT-GGAAGCGGTGCGG-AGGAT------------
**************************** ************* *****

TTGTGATTCACTTCTCGTCATTCTCCGAGCCTTACAAACAAACAAAAATAGCTCTCTCTT
------------------------------------------------------------


TTCGCTGACGATTGTCCCCCAACACGGGCGGCGGTTATTTTGGCTCTCGCCACTCTTTTT
------------------------------------------------------------


CAAGGTGCACAATCAAAAACAAACAAGTCCAGCCCGATTTCGCGAATTTGTTCATCTAGT
------------------------------------------------------------


TCAACGCATTTTTTTCATTTTGAATACATTTGCTTGTATCGTTTCCGAGTTTCTTGAACA
------------------------------------------------------------


ATTACTTTCAGAATTGAGTATCATTTTTTGTGTTCAGCAGTGAAAAATAACCAAAATTTC
------------------------------------------------------------


TCTATCAGGCTTGCACTCACATTGGTCATATTTTACAGTCCTGACCTAGAGTGCGTTTAT
------------------------------------------------------------


GAAT
----
```

>ncRNA6 II:8750877,8751148: 272bp
tctttttcaactttatcttttagccgtatctcattattcactttccacccagtagcttcatcctcttgtcttccaattctccgcaaagttcacctacccttttgtactagtgacctgatgcgacgaattcaaccattgacctcctcctccttctcacacttagacacacatacatccgtaacccgaagcagcgaaagttaagcatgaaagcgaaaggaaagtgaaaacaattaggaaaagtcggcatcatcatggatcggaggaggcgacgc

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	TCCGCAAAGTTCACCTACCC	Plus	20	80	99	54.13	55.00%
Reverse primer	AACTTTCGCTGCTTCGGGTT	Minus	20	199	180	55.02	50.00%

Internal oligo: Plus
Product length: 120
Sequencing result:

Forward primer

```
TCTTTTTCAACTTTATCTTTTAGCCGTATCTCATTATTCACTTTCCACCCAGTAGCTTCA
------------------------------------------------------------


TCCTCTTGTCTTCCAATTCTCCGCAAAGTTCACCTACCCTTTTGTACTAGTGACCTGATG
-----------------------------------------------------------G
                                                           *

CGACGAATTCAACCATTGACCTCCTCCTCCTTCTCACACTTAGACACACATACATCCGTA
CGACGATTC--ACCATTGACCTCCTCCTCCTTCTCACACTTAGACACACATACATCCGTA
****** *   *************************************************

ACCCGAAGCAGCGAAAGTTAAGCATGAAAGCGAAAGGAAAGTGAAAACAATTAGGAAAAG
ACCCGAAGCAGCGAAAGTT-----------------------------------------
*******************

TCGGCATCATCATGGATCGGAGGAGGCGACGC
--------------------------------
```

Reverse primer

```
TCTTTTTCAACTTTATCTTTTAGCCGTATCTCATTATTCACTTTCCACCCAGTAGCTTCA
------------------------------------------------------------
```

```
TCCTCTTGTCTTCCAATTCTCCGCAAAGTTCACCTACCCTTTTGTACTAGTGACCTGATG
--------------------CCGCAAAGTTCACCTACCCTTTTGTACTAGTGACCTGATG
                    ****************************************

CGACGAATTCAACCATTGACCTCCTCCTCCTTCTCACACTTAGACACACATACATCCGTA
CGACGAATTCAACCATTGACCTCNTCCT--NTCTCACACT--------------------
*********************** ****   *********

ACCCGAAGCAGCGAAAGTTAAGCATGAAAGCGAAAGGAAAGTGAAAACAATTAGGAAAAG
------------------------------------------------------------


TCGGCATCATCATGGATCGGAGGAGGCGACGC
--------------------------------
```

>ncRNA7 III:5964548,5964705: 158bp
tacacacgtgtctgcgggtctctgatcctactactctctcctcccgtgcttcgtgttcttcccgccgggagctcaaaaacgccgccgccgc
cgctgttcctctaactctcttgcatcgacgggtgcttttcgttgtttcttattgttcttgctca

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	GTCTGCGGGTCTCTGATCCT	Plus	20	10	29	55.17	60.00%
Reverse primer	AAAAGCACCCGTCGATGCAA	Minus	20	133	114	55.38	50.00%

Internal oligo: Plus
Product length: 124

Sequencing result:

Forward primer

```
TACACACGTGTCTGCGGGTCTCTGATCCTACTACTCTCTCCTCCCGTGCTTCGTGTTCTT
------------------------------------------------GCTCGTGTTCTT
                                                  **********


CTTCCCGCCGGGAGCTCAAAAACGCCGCCGCCGCCGCTGTTCCTCTAACTCTCTTGCATC
CTTCC-GCCGGGAGCTCNAAA-CGCCGCCGCCGCCGCTGTTCCTCTAACTCTCTTGNNNN
***** *********** *** **********************************


GACGGGTGCTTTTCGTTGTTTCTTATTGTTCTTGCTCA
GACGGGTGCTTTTG------------------------
*************
```

Reverse primer

```
TACACACGTGTCTGCGGGTCTCTGATCCTACTACTCTCTCCTCCCGTGCTTCGTGTTCTT
---------GTCTGCGGGTCTCTGATCCTACTACTCTCTCCTCCCGTGCTTCGTGTTCTT
         ***************************************************


CTTCCCGCCGGGAGCTCAAAAACGCCGCCGCCGCCGCTGTTCCTCTAACTCTCTTGCATC
CTTCCCGCCGG-AGCTCAAAAACGCC----------------------------------
*********** **************


GACGGGTGCTTTTCGTTGTTTCTTATTGTTCTTGCTCA
--------------------------------------
```

>ncRNA8   IV:3851506,3851664: 159bp
agttgtgcaatattttcgaggctagtttggcaaagctggacacaattttcccaaaaaatacggaaattcacaatttacctcaccttcctccttcca
actttcaccttctcgcggtcactttcccgccgtagccctcctccaccatccaaagtttaatag

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	TAGTTTGGCAAAGCTGGACACA	Plus	22	23	44	54.75	45.45%
Reverse primer	TGGATGGTGGAGGAGGGCTA	Minus	20	148	129	55.78	60.00%

Internal oligo: Plus
Product length: 126

Sequencing result:

Forward primer

```
AGTTGTGCAATATTTTCGAGGCTAGTTTGGCAAAGCTGGACACAATTTTCCCAAAAAATA

------------------------------------------------------------




CGGAAATTCACAATTTACCTCACCTTCCTCCTTCCAACTTTCACCTTCTCGCGGTCACTT

-------------TTTACCTCACCTTCCTCCTTCCAACTTTCACCTTCTCGCGGTCACTT

             ***********************************************


TCCCGCCGTAGCCCTCCTCCACCATCCAAAGTTTAATAG

TCCCGCCGTAGCCCTCCTCCACCATCC------------

***************************
```

Reverse primer

```
AGTTGTGCAATATTTTCGAGGCTAGTTTGGCAAAGCTGGACACAATTTTCCCAAAAAATA
----------------------AGTTTGGCAAAGCTGGACACAATTTTCCCAAAAAATA
                      **************************************

CGGAAATTCACAATTTACCTCACCTTCCTCCTTCCAACTTTCACCTTCTCGCGGTCACTT
CGGAAATTCACAATTTACCTCACCTCC--TCCTCCAANT--CACCT--------------
************************* *   * ***** *  *****

TCCCGCCGTAGCCCTCCTCCACCATCCAAAGTTTAATAG
---------------------------------------
```

>ncRNA9 IV:9196924,9197103: 180bp
ccgcagaagcagcagcagtcgtcgccatccttctccgtctccaatgtctcgataagcaacgagagtgaagagagcccaaagaaactgtcta
cactctctcgtgtctctcgttgctcactcgtctcggcgagaaaacgaacgaaacgaagcgatgaagagcagcagctcagtcctcgagca

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	GCAACGAGAGTGAAGAGAGCC	Plus	21	56	76	55.28	57.14%
Reverse primer	TGCTCGAGGACTGAGCTGC	Minus	19	180	162	56.26	63.16%

Internal oligo: Plus
Product length: 125
Sequencing result:

Forward primer

```
CCGCAGAAGCAGCAGCAGTCGTCGCCATCCTTCTCCGTCTCCAATGTCTCGATAAGCAAC
------------------------------------------------------------



GAGAGTGAAGAGAGCCCAAAGAAACTGTCTACACTCTCTCGTGTCTCTCGTTGCTCACTC
------------------------------------------GTCTCTCGTTGCTCACTC
                                          ******************


GTCTCGGCGAGAAAACGAACGAAACGAAGCGATGAAGAGCAGCAGCTCAGTCCTCGAGCA
GTCTCGGCGAGAAAACGAACGAAACGAAGCGATGAAGAGCAGCAGCTCAGTCCTCGAGC-
***********************************************************
```

Reverse primer

```
CCGCAGAAGCAGCAGCAGTCGTCGCCATCCTTCTCCGTCTCCAATGTCTCGATAAGCAAC
-------------------------------------------------------GCAAC
                                                       *****

GAGAGTGAAGAGAGCCCAAAGAAACTGTCTACACTCTCTCGTGTCTCTCGTTGCTCACTC
GAGAGTGAAGAGAGCCCAAAGAAACTGTCTACACTCTCTCGTGTCTCTCNT-GCTCACTC
************************************************* * ********

GTCTCGGCGAGAAAACGAACGAAACGAAGCGATGAAGAGCAGCAGCTCAGTCCTCGAGCA
GTCTCGGCGAGA------------------------------------------------
************
```

> ncRNA10   V:14505342,14505801:460bp

```
gtttgataatttcaaaaaattctgtattatttggtaacaaaccctaggtctacaatacctctttgagcgcaaaagttgaaaactagaagtttcaaata
ccgtatttcctctattagttctaaactcattgttcaactttgaaacttttaaattgaagtagtagggtaggaaatactatcagtgtctatgcagaaac
ctataaagttgtactttcactttttcgatctcgcgttaacttggaccaaaagctaccttaattttaagtgaaatccccagacgtggatttccctcttaa
tggtcaacttgatagatctaggtaatccccacttttcagtacttttcccatcaacgccctctcattcttttccaatttccgtgtcaattattccaattatt
catgttcatatcaatctctcttatcacctctaatcgtcgacaccactgtcttcttttcattttt
```

Sequence (5'->3')	Strand on template	Length	Start	Stop	Tm	GC%
Forward primer						
TCGATCTCGCGTTAACTTGGAC	Plus	22	221	243	54.67	50.00%
Reverse primer						
ATGAGAGGGCGTTGATGGGA	Minus	20	356	341	55.26	55.00%

Internal oligo Plus

Product length 136

Sequencing result:

Forward primer

```
GTTTGATAATTTCAAAAAATTCTGTATTATTTGGTAACAAACCCTAGGTCTACAATACCT
------------------------------------------------------------


CTTTGAGCGCAAAAGTTGAAAACTAGAAGTTTCAAATACCGTATTTCCTCTATTAGTTCT
------------------------------------------------------------


AAACTCATTGTTCAACTTTGAAACTTTTAAATTGAAGTAGTAGGGTAGGAAATACTATCA
------------------------------------------------------------


GTGTCTATGCAGAAACCTATAAAGTTGTACTTTCACTTTTTCGATCTCGCGTTAACTTGG
------------------------------------------------------------


ACCAAAAGCTACCTTAATTTTAAGTGAAATCCCCAGACGTGGATTTCCCTCTTAATGGTC
--------------------TNANTGA---TCNCAGACGTGGATTTCCCTCTTAATGGTC
                    * * ***    * ***************************
```

```
AACTTGATAGATCTAGGTAATCCCCACTTTTCAGTACTTTTCCCATCAACGCCCTCTCAT
AACTTGATAGATCTAGGTAATCCCCACTTTTCAGTACTTTTCCCATCAACGCCCTCTC--
**********************************************************

TCTTTTCCAATTTCCGTGTCAATTATTCCAATTATTCATGTTCATATCAATCTCTCTTAT
------------------------------------------------------------

CACCTCTAATCGTCGACACCACTGTCTTCTTTTCATTTTT
----------------------------------------
```

Reverse primer

```
GTTTGATAATTTCAAAAAATTCTGTATTATTTGGTAACAAACCCTAGGTCTACAATACCT
------------------------------------------------------------

CTTTGAGCGCAAAAGTTGAAAACTAGAAGTTTCAAATACCGTATTTCCTCTATTAGTTCT
------------------------------------------------------------

AAACTCATTGTTCAACTTTGAAACTTTTAAATTGAAGTAGTAGGGTAGGAAATACTATCA
------------------------------------------------------------

GTGTCTATGCAGAAACCTATAAAGTTGTACTTTCACTTTTTCGATCTCGCGTTAACTTGG
-----------------------------------------CGATCTCGCGTTAACTTGG
                                         *******************

ACCAAAAGCTACCTTAATTTTAAGTGAAATCCCCAGACGTGGATTTCCCTCTTAATGGTC
ACCAAAAGCTACCTTAATTTTAAGTGAAATCCCCAGACGTGGATTTCCCTCTTAATGGTC
************************************************************

AACTTGATAGATCTAGGTAATCCCCACTTTTCAGTACTTTTCCCATCAACGCCCTCTCAT
AACT-GATAGATCTAG--------------------------------------------
**** ***********

TCTTTTCCAATTTCCGTGTCAATTATTCCAATTATTCATGTTCATATCAATCTCTCTTAT
------------------------------------------------------------

CACCTCTAATCGTCGACACCACTGTCTTCTTTTCATTTTT
----------------------------------------
```

>ncRNA11 V:11525384,11525537:154bp
tcatcttctccaaaggtgttcttgactcactcactctcttttctcgtagtacaccaccacaccaatcaggttatttctttcacgcaccgccacgtcg
gctcctcccactttttttgcgcaatcgtcgtgccctccgcgcctgctgctgcagacag

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	AGGTGTTCTTGACTCACTCACTC	Plus	23	14	36	54.03	47.83%
Reverse primer	CGGAGGGCACGACGATTG	Minus	18	135	118	55.33	66.67%

Internal oligo: Plus
Product length: 122
Sequencing result:

Forward primer

```
TCATCTTCTCCAAAGGTGTTCTTGACTCACTCACTCTCTTTTCTCGTAGTACACCACCAC
------------------------------------------------------------



ACCAATCAGGTTATTTCTTTCACGCACCGCCACGTCGGCTCCTCCCACTTTTTTGCGCAA
-------AGGTTATTTCTTTCACGCACCGNNACGTCGGCTCCTCCCACTTTTTTGCGCAA
       **********************  *****************************


TCGTCGTGCCCTCCGCGCCTGCTGCTGCAGACAG
TCGTCGTGCCCTCCG-------------------
***************
```

Reverse primer

```
TCATCTTCTCCAAAGGTGTTCTTGACTCACTCACTCTCTTTTCTCGTAGTACACCACCAC
-------------ANGGTGTTCTTGACTCACTCACTCTCTTTTCTCGTAGTACACCACCAC
             * *********************************************

ACCAATCAGGTTATTTCTTTCACGCACCGCCACGTCGGCTCCTCCCACTTTTTTGCGCAA
ACCAATCANGTTATT--CTTCACGCACC--------------------------------
******** ******   **********

TCGTCGTGCCCTCCGCGCCTGCTGCTGCAGACAG
----------------------------------
```

>ncRNA12   V:14504822,14505114: 293bp

cctctattaaaccattgcttccaattttctacacctcttattctttctctatctatctcacactcattt*cccccacaccaaaatgacga*ctcatcgaca
cattttcatcgtagtcagcccggggcggtgggaggagggcgccccccgagtaataccataaaacttttgtgtgtgtccggtgtgagcgacta
gaggaaggaagaggcaaaaaatgaaaagagcggcggcgattctttgctattactgaaggatgaagcgtgtacgtatctatttagacttcttgt
gttctcgat

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	CCCCCACACCAAAATGACGA	Plus	20	70	89	54.67	55.00%
Reverse primer	TCCTCTAGTCGCTCACACCG	Minus	20	196	177	55.26	60.00%

Internal oligo: Plus
Product length: 127
Sequencing result:

Forward primer

```
CCTCTATTAAACCATTGCTTCCAATTTTCTACACCTCTTATTCTTTCTCTATCTATCTCA
------------------------------------------------------------



CACTCATTTCCCCCACACCAAAATGACGACTCATCGACACATTTTCATCGTAGTCAGCCC
---------------------------------------------------------CCC
                                                         ***

GGGGCGGTGGGAGGAGGGCGCCCCCCGAGTAATACCATAAAACTTTTGTGTGTGTCCGGT
GGGGCGGTGGGAGGAGGGCGCCCCCCGAGTAATACCATAAAACTTTTGTGTGTGTCCGGT
************************************************************

GTGAGCGACTAGAGGAAGGAAGAGGCAAAAAATGAAAAGAGCGGCGGCGATTCTTTGCTA
GTGAGCGACTAGAGG---------------------------------------------
***************

TTACTGAAGGATGAAGCGTGTACGTATCTATTTAGACTTCTTGTGTTCTCGAT
-----------------------------------------------------
```

Reverse primer

```
CCTCTATTAAACCATTGCTTCCAATTTTCTACACCTCTTATTCTTTCTCTATCTATCTCA
------------------------------------------------------------


CACTCATTTCCCCCACACCAAAATGACGACTCATCGACACATTTTCATCGTAGTCAGCCC
---------CCCCCACACCAAAATGACGACTCATCGACACATTTTCATCGTAGTCAGCCC
         ***************************************************

GGGGCGGTGGGAGGAGGGCGCCCCCCGAGTAATACCATAAAACTTTTGTGTGTGTCCGGT
GGGGCGGTGGGAGGAGGGCGCCNNCCGAGT------------------------------
**********************  ******

GTGAGCGACTAGAGGAAGGAAGAGGCAAAAAATGAAAAGAGCGGCGGCGATTCTTTGCTA
------------------------------------------------------------


TTACTGAAGGATGAAGCGTGTACGTATCTATTTAGACTTCTTGTGTTCTCGAT
-----------------------------------------------------
```

>ncRNA13 V:6879489,6879644: 156bp
ccaccagaccagtagcatctgtaaagtgcgcggccttcctccacccatcactctctcgtaacctcaatgaaatagtggcgcgttcgatgagg
gtatcctcctgcgtctgtggtacacaaaacgctattttttgcctgcaacccgggggccctcttt

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	CACCAGACCAGTAGCATCTGT	Plus	21	2	22	53.57	52.38%
Reverse primer	TTTTGTGTACCACAGACGCAGG	Minus	22	121	100	55.50	50.00%

Internal oligo: Plus
Product length: 120
Sequencing result:

Forward primer

```
CCACCAGACCAGTAGCATCTGTAAAGTGCGCGGCCTTCCTCCACCCATCACTCTCTCGTA

--------------------------------------------------CTCTCTCGTA

                                                  **********


ACCTCAATGAAATAGTGGCGCGTTCGATGAGGGTATCCTCCTGCGTCTGTGGTACACAAA

ACCTCAATGAAATAGTGGCGCGTTCGATGAGGGTATCCTCCTGCGTCTGTGGTACAC---

*********************************************************


ACGCTATTTTTTGCCTGCAACCCGGGGGCCCTCTTT

------------------------------------
```

Reverse primer

```
CCACCAGACCAGTAGCATCTGTAAAGTGCGCGGCCTTCCTCCACCCATCACTCTCTCGTA
-CACCAGACCAGTAGCATCTGTAAAGTGCGCGGCCTTCCTCCACCCATCACTCTCTCGTA
 ***********************************************************

ACCTCAATGAAATAGTGGCGCGTTCGATGAGGGTATCCTCCTGCGTCTGTGGTACACAAA
ACCTCAA-GAAATAGNG-CGCG--------------------------------------
******* ******* * ****

ACGCTATTTTTTGCCTGCAACCCGGGGGCCCTCTTT
------------------------------------
```

>ncRNA14 X:5818466,5818828: 363bp

gcgtatatatcgttgatgtgttctctcgattgtttactcgatgcactatgtttctactaacaataacggatttttaggagacagctccataagccccg
cccttcccccctttaacgccagcctcttgcggccagccccacaggcctagaacactactgcataataaaaatgaccactggccgtttttttccttc
tctcactcccttcgagtagttaaccaaagtccttcctcgtttttcattctatttttgtcgcatctttttttcaacttttccagaattctttttgcatgacct
ctcgtttctttttttgcattctatattctaattggtcttattcaaaatcatctcattttttatat

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	AGACAGCTCCATAAGCCCCG	Plus	20	79	98	55.80	60.00%
Reverse primer	ACTCGAAGGGAGTGAGAGAAGG	Minus	22	210	189	54.94	54.55%

Internal oligo: Plus

Product length: 132

Sequencing result:

Forward primer

```
GCGTATATATCGTTGATGTGTTCTCTCGATTGTTTACTCGATGCACTATGTTTCTACTAA
------------------------------------------------------------

CAATAACGGATTTTTAGGAGACAGCTCCATAAGCCCCGCCCTTCCCCCTTTAACGCCAGC
------------------------------------------------------------


CTCTTGCGGCCAGCCCCACAGGCCTAGAACACTACTGCATAATAAAAATGACCACTGGCC
CTCTTGCGGN-AGCCCCACAGGCCTAGAACACTACTGCATAATAAAAATGACCACTGGCC
*********  *************************************************

GTTTTTTTCCTTCTCTCACTCCCTTCGAGTAGTTAACCAAAGTCCTTCCTCGTTTTTTCA
GTTTTTTTCCTTCTCTCCCTCCCTTCGAGT------------------------------
***************** ************

TTCTATTTTTGTCGCATCTTTTTTTTCAACTTTTCCAGAATTCTTTTTTGCATGACCTCT
------------------------------------------------------------
```

```
CGTTTCTTTTTTTGCATTCTATATTCTAATTGGTCTTATTCAAAATCATCTCATTTTTTA
------------------------------------------------------------


TAT
---
```

Reverse primer

```
GCGTATATATCGTTGATGTGTTCTCTCGATTGTTTACTCGATGCACTATGTTTCTACTAA
------------------------------------------------------------

CAATAACGGATTTTTAGGAGACAGCTCCATAAGCCCCGCCCTTCCCCCTTTAACGCCAGC
-----------------GAGACAGCTCCATAAGCCCCGCCCTTCCCCCTTTAACGCCAGC
                 *******************************************

CTCTTGCGGCC-AGCCCCACA-GGCCTAGAACACTACTGCATAATAAAAATGACCACTGG
CTCTTGCGGCCGAGCCCCACNAGGCCTAGAACACTACTGCATAATA--------------
*********** ********  ************************

CCGTTTTTTTCCTTCTCTCACTCCCTTCGAGTAGTTAACCAAAGTCCTTCCTCGTTTTTT
------------------------------------------------------------

CATTCTATTTTTGTCGCATCTTTTTTTTCAACTTTTCCAGAATTCTTTTTTGCATGACCT
------------------------------------------------------------

CTCGTTTCTTTTTTTGCATTCTATATTCTAATTGGTCTTATTCAAAATCATCTCATTTTT
------------------------------------------------------------

TATAT
-----
```

>ncRNA15 X:3834195,3834401: 207bp

aaaaaaactcgcgcatcgataacgtgaaaaggttgcttggtgctgagaaaagtagatagagaactctccccagacgaataccggaaaaga
aacaaacatcgcacaattggcagaaatacgtcagaaagccaaggacaaacacgttgttgcggcggctacagaaaacaaataatatggtgc
agcagcaacagatcggagcaacaaaaa

Sequence (5'->3')		Strand on template	Length	Start	Stop	Tm	GC%
Forward primer	CGTGAAAAGGTTGCTTGGTGC	Plus	21	23	43	55.43	52.38%
Reverse primer	CCGCAACAACGTGTTTGTCC	Minus	20	152	133	55.27	55.00%

Internal oligo: Plus
Product length: 130
Sequencing result:

Forward primer

```
AAAAAAACTCGCGCATCGATAACGTGAAAAGGTTGCTTGGTGCTGAGAAAAGTAGATAGA

------------------------------------------------------------


GAACTCTCCCCAGACGAATACCGGAAAAGAAACAAACATCGCACAATTGGCAGAAATACG

-------------ACGA-TACCGGAAAG--AACAAACATCGCACAATTGGCAGAAATACG

             **** *********   ******************************

TCAGAAAGCCAAGGACAAACACGTTGTTGCGGCGGCTACAGAAAACAAATAATATGGTGC

TCAGAAAGCCAAGGACAAACACGTTGTTGCGG----------------------------

********************************


AGCAGCAACAGATCGGAGCAACAAAAA

---------------------------
```

Reverse primer

```
AAAAAAACTCGCGCATCGATAACGTGAAAAGGTTGCTTGGTGCTGAGAAAAGTAGATAGA
-----------------------GTGAAAAGGTTGCTTGGTGCTGAGAAAAGTAGATAGA
                       *************************************

GAACTCTCCCCAGACGAATACCGGAAAAGAAACAAACATCGCACAATTGGCAGAAATACG
GAACTCTCCCCAGACGAATACCGGAAAAGAAACAAACATCGCACAATGC-----------
***********************************************

TCAGAAAGCCAAGGACAAACACGTTGTTGCGGCGGCTACAGAAAACAAATAATATGGTGC
------------------------------------------------------------


AGCAGCAACAGATCGGAGCAACAAAAA
---------------------------
```