Published online 5 November 2007 *Nucleic Acids Research, 2008, Vol. 36, Database issue* ***D57–D62***
doi:10.1093/nar/gkm946

UTRome.org: a platform for 3′UTR biology in *C. elegans*

Marco Mangone, Philip MacMenamin, Charles Zegar, Fabio Piano and Kristin C. Gunsalus*

Department of Biology and Center for Genomics and Systems Biology, New York University, 100 Washington Square East, New York, NY 10003, USA

Received August 16, 2007; Revised October 11, 2007; Accepted October 16, 2007

ABSTRACT

Three-prime untranslated regions (3′UTRs) are widely recognized as important post-transcriptional regulatory regions of mRNAs. RNA-binding proteins and small non-coding RNAs such as microRNAs (miRNAs) bind to functional elements within 3′UTRs to influence mRNA stability, translation and localization. These interactions play many important roles in development, metabolism and disease. However, even in the most well-annotated metazoan genomes, 3′UTRs and their functional elements are not well defined. Comprehensive and accurate genome-wide annotation of 3′UTRs and their functional elements is thus critical. We have developed an open-access database, available at http://www.UTRome.org, to provide a rich and comprehensive resource for 3′UTR biology in the well-characterized, experimentally tractable model system *Caenorhabditis elegans*. UTRome.org combines data from public repositories and a large-scale effort we are undertaking to characterize 3′UTRs and their functional elements in *C. elegans*, including 3′UTR sequences, graphical displays, predicted and validated functional elements, secondary structure predictions and detailed data from our cloning pipeline. UTRome.org will grow substantially over time to encompass individual 3′UTR isoforms for the majority of genes, new and revised functional elements, and *in vivo* data on 3′UTR function as they become available. The UTRome database thus represents a powerful tool to better understand the biology of 3′UTRs.

INTRODUCTION

Three-prime untranslated regions (3′UTRs) are untranslated portions of mRNAs located at the 3′ flanking end of open reading frames (ORFs). These regions are implicated in post-transcriptional regulation of gene activity through interaction with regulatory RNA-binding proteins and small non-coding RNAs such as miRNAs, which can influence protein activity by altering mRNA stability, translational efficiency or localization (1–6). Regulation at the level of 3′UTRs, by both regulatory proteins and small RNAs, plays essential roles in diverse developmental and metabolic processes and is also implicated in disease (1–6). miRNAs, which bind to short complementary sequences in 3′UTRs of metazoans, represent one of the best studied families of 3′UTR regulators (4,5). Based on bioinformatic analysis of predicted miRNA-binding sites in 3′UTRs, it has been proposed that each miRNA controls a network of proteins *in vivo*, and that collectively thousands of transcripts are likely to be regulated by miRNAs (7).

Due to the critical role that 3′UTRs play in living cells, it is important to study these regions in detail to uncover and characterize as many embedded regulatory elements as possible. However, 3′UTRs are still incompletely annotated in metazoan genomes, including humans (7). Even in *Caenorhabditis elegans*, one of the best annotated metazoan genomes, only about half of known transcripts have an annotated 3′UTR (8,9). Recent studies indicate that a substantial proportion of characterized transcripts in humans and other species experience alternative splicing of a terminal exon or alternative polyadenlyation (polyA) site usage (10–12). For example, careful curation of mRNA sequence data shows that at least one-third of genes analyzed in human, mouse and *Arabidopsis*, and over 10% in *C. elegans*, express transcripts that share a terminal exon but use different polyA signal (PAS) sites, resulting in 3′UTRs of different lengths [(12); D. and J. Thierry-Mieg, personal communication]. Both 3′UTR isoforms and regulation can vary in a tissue-specific manner (13,14), and a significant fraction of predicted miRNA target sites in human genes are located in alternative UTR segments (15). These studies suggest that heterogeneity and combinatorial control of 3′UTR isoforms are likely to play a more significant role in regulation of gene activity than previously appreciated.

*To whom correspondence should be addressed. Tel: +1 212 998 8236; Fax: +1 212 995 4015; Email: kcg1@nyu.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Increased interest in 3′UTRs has spawned several new resources focused on 3′UTRs and their functional elements, such as UTRdb and UTRsite (16), PACdb (17), Poly_A db (18), PicTar (19), TargetScan (20) and miRanda (21), which use cross-species alignments and EST data to predict or highlight elements within UTRs that may have a functional role in RNA maturation or post-transcriptional gene regulation. However, only some of these contain data specific for *C. elegans* and none are dedicated as a comprehensive archive for all aspects of 3′UTR biology within a specific tractable model system. We have therefore developed a database focused on *C. elegans* 3′UTRs and their functional elements, UTRome.org, intended as a comprehensive resource for 3′UTR biology in *C. elegans*. The design and implementation we have established for UTRome.org could easily be adapted for the analysis of 3′UTRs in other species, including human.

OVERVIEW

The UTRome database provides up-to-date information on 3′UTR structures and functional elements for every *C. elegans* mRNA based on combined data from public repositories such as WormBase (8,9) and continuously updated results from an ongoing high-throughput pipeline we have developed to define 3′UTRs and their isoforms (Figure 1A). Information about functional elements within 3′UTRs currently includes computationally predicted miRNA-binding sites [derived from the PicTar (19,22) and MiRanda (21) algorithms], putative PAS sites [computed based on Ref. (23)], and predicted secondary structures [using the MFOLD algorithm (24)]. For each 3′UTR, users can view or download secondary structure prediction diagrams and browse graphical coordinate-based displays illustrating gene models, 3′UTR products from our cloning pipeline, previously annotated evidence for 3′UTRs from ESTs and mRNAs, putative PAS sites and predicted or validated miRNA-binding sites. We also provide a detailed description of data produced by our cloning pipeline, including status of cloning and annotation, ABI trace files, BLAT (25) and BLAST (26) alignments to the genome, and annotated agarose gel images of RT-PCR products used for cloning. As new data become available, UTRome.org will grow substantially over time to encompass individual isoforms for the majority of genes, improved predictions for miRNA-binding sites based on updated 3′UTR annotations and additional sequenced genomes, and results from *in vivo* analyses of 3′UTR structure and function, including experimental characterization of specific functional sequence elements.

DESIGN AND IMPLEMENTATION

UTRome.org uses an Apache web server and a collection of Perl CGI scripts coupled to a MySQL database to provide an intuitive user interface for 3′UTR data. The main UTRome database schema archives sequence and functional information on 3′UTRs and their corresponding genes, coding sequences (CDSs) and functional elements. It also serves as an electronic lab notebook to track all stages of our in-house 3′UTR cloning and annotation pipeline: from initial RT-PCR through generation of first-pass UTR sequence tags (USTs) based on automated BLAT and BLAST analysis, final sequence verification of 3′UTRs, and annotation of functional elements (a full description of this pipeline will be published elsewhere). A second light-weight GFF database (27) stores coordinate-based data for generating graphical displays of sequence-based annotations, which are generated dynamically using Bio::DB::GFF (part of BioPerl, http://www.bioperl.org) and the Generic Genome Browser (GBrowse) (27). An automated set of scripts generates first-pass annotations from our cloning pipeline from batches of raw sequence traces using BLAT and BLAST and deposits the raw sequence data, USTs, and validated 3′UTR sequences into the database on an ongoing basis. Data are extracted from external data sources using Perl scripts [e.g. from WormBase's AceDB engine (28,29)] and imported using Perl or MySQL scripts.

The UTRome database currently contains a comprehensive collection of all ~26 000 *C. elegans* transcripts from WormBase release WS180 and 3′UTR sequence annotations from our cloning pipeline. All coordinate-based data will be updated regularly and synchronized with each new WormBase freeze. The entire UTRome.org database and data processing framework could easily be adapted for any other organism by coupling the system to data import protocols compatible with different public repositories [e.g. FlyBase (30), etc.].

USING UTROME.ORG

Searching UTRome.org

The Welcome page contains a query box in the top right corner (mirrored in each page of the website), which lets the user search for a specific 3′UTR or for multiple 3′UTRs using wildcards. The accompanying pull-down menu allows users to search across the entire genome ('UTRome & Genome') or to limit queries to genes targeted by our cloning pipeline ('UTRome Only'). A productive search returns a comprehensive list of genes and 3′UTRs matching the query (Figure 1C). For each gene in the result list, we provide general information such as the Cosmid ID, Locus name, Chromosome and a brief description (accessible by mousing over any Gene or 3′UTR). The first column indicates whether the corresponding gene is targeted by our pipeline (blue if in the UTRome project, empty otherwise). If the 3′UTR has been annotated by WormBase or the annotation from UTRome has been finalized, we indicate its length in base pairs. For 3′UTRs in our cloning pipeline, we assign a color-coded flag (green, orange or red circles) as an indicator of confidence as to whether a given UST is a *bona fide* 3′UTR for the targeted gene. These preliminary annotations will be updated to final curation status on an ongoing basis as the project evolves. At the bottom of this and every page in on the website, we include a menu bar

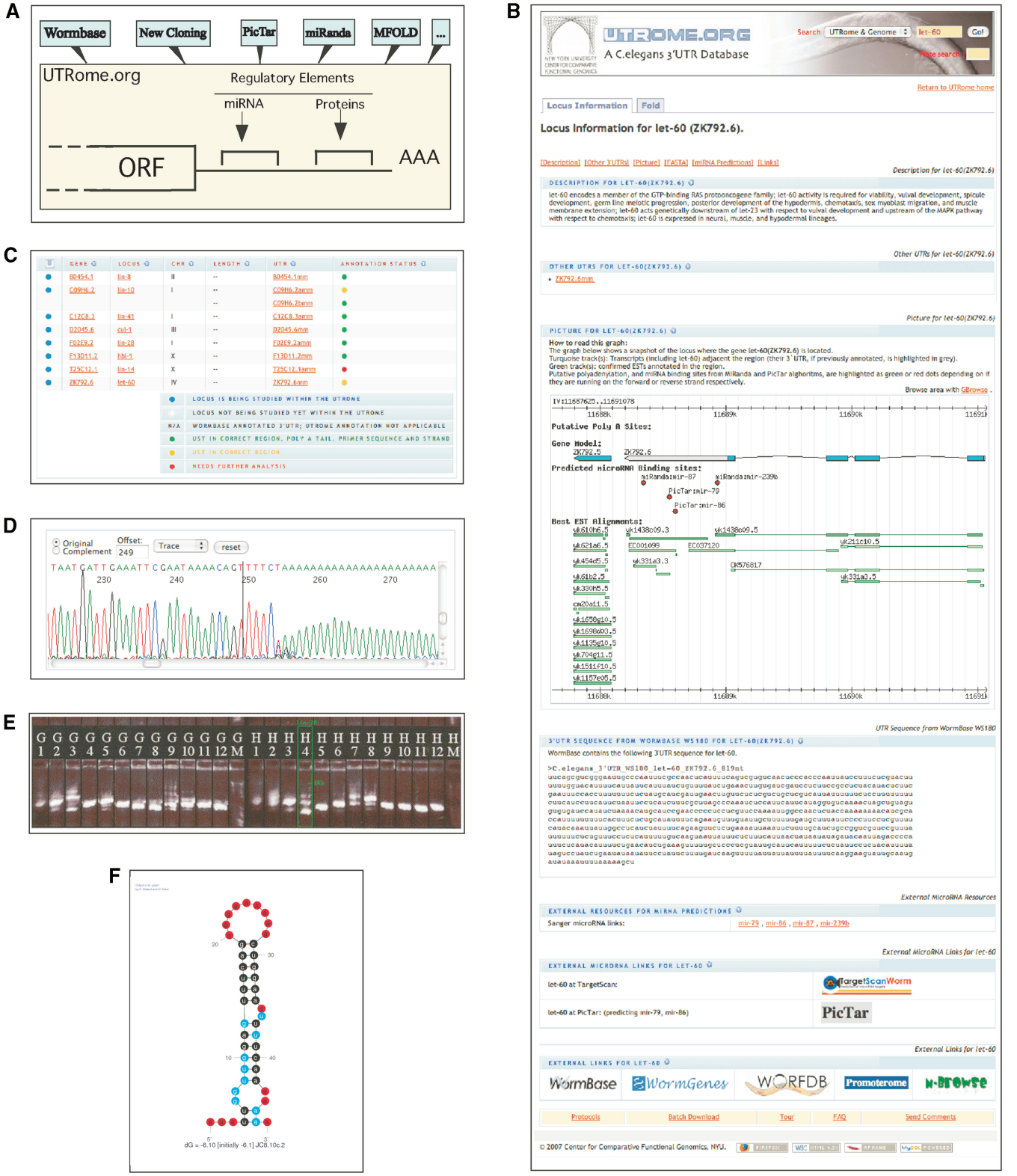

Figure 1. Overview of UTRome.org. The UTRome database integrates diverse information on *C. elegans* 3′UTRs. (**A**) Data on 3′UTR boundaries and predicted or experimentally validated functional elements, collected from multiple database sources or analyzed using various computational algorithms, are displayed in a series of user-friendly web pages. (**B**) 'Locus Information' page: a sample snapshot of aggregated data. (**C**) Results returned for the query 'lin' in a search limited to genes targeted by the UTRome project. (**D**) 'ABI trace files' page: a Java applet shows sequence traces for a UST including part of the polyA tail. (**E**) Excerpt from a 'Gel' page: PCR products from a 96-well cloning experiment indicate evidence for multiple 3′UTR isoforms in well H4 (automatically highlighted by a green box). (**F**) 'MFOLD' page: secondary structure prediction for a 3′UTR showing putative stem-loop structure.

containing links to protocols, batch downloads, a tour of the site, a FAQ page and email for feedback.

Browsing 3′UTR data

Each gene or 3′UTR present in the database can be browsed by clicking on its hyperlink in the Results list, which brings the user to a tabbed menu of data display options for the selected gene or 3′UTR. The set of tabs opens by default on a 'Locus Information' page providing general information for the given gene or 3′UTR (Figure 1B): a gene description, a list of alternate 3′UTR isoforms for this gene (if any), 3′UTR sequence in FASTA format (if annotated), a graphical display of the locus along with annotated functional elements, and separate tables listing the miRNAs predicted to target the gene [hyperlinked to their corresponding records at miRBase (31)], external miRNA–target prediction sites providing more detailed data and sequence alignments [PicTar (19), and TargetScan (20)], and links to other external database resources [WormBase (8,9), WormGenes (12), WorfDB (32), Promoterome (33) and N-Browse (19)]. Mousing over any of these links displays a brief description of the external resource. The graphical display shows the transcript model(s) for the given gene and, if available, previously mapped ESTs and mRNAs (from WormBase), predicted miRNA-binding sites (from both PicTar and miRanda), and sequence conservation with the *C. briggsae* genome. Additional conservation tracks will be included in future releases. A link to a local installation of GBrowse allows the user to study the region in more detail if desired, including zooming in to the nucleotide level. A web form near the bottom of the page allows users to submit (anonymously, if desired) comments, suggestions or requests (e.g. for inclusion of additional data) to the database administrator.

A second tab labeled 'Fold' links to a webpage displaying the predicted secondary structure for the 3′UTR region of the corresponding transcript (Figure 1F), calculated using the MFOLD algorithm (24). Secondary structures in RNA molecules may influence the accessibility of sequence-specific recognition motifs by factors such as miRNAs and can also serve as structural features recognized by some RNA-binding proteins (6). Although MFOLD predictions are not experimentally validated, they represent a valuable starting point to model the interaction of the given 3′UTR with RNA-binding factors. Taken together, these resources provide a powerful tool to study *C. elegans* 3′UTRs by synthesizing all the publicly available information for 3′UTRs genome-wide.

If the given 3′UTR has been cloned by our group, additional options will appear in the tabbed menu bar at the top of the page: 'UTR cloning', 'ABI trace file', 'Gel' and 'Plate'. The 'UTR cloning' page provides detailed cloning information and a graphical interpretation of new 3′UTR annotations produced by our pipeline (Figure 2 shows several examples). Here a brief description of the gene is followed by a 'Cloning status' table, which includes the sequence of the primer used for cloning, its melting temperature (T_m) and the contiguous length of the best BLAT alignment of the UST to the *C. elegans* genome for the 3′UTR clone of interest. The next panel, '3′UTR bioinformatic analysis', contains a computer-generated summary of the first-pass annotation from our pipeline, indicating cloning progress and UST quality (e.g. whether the sequence contains a poly-A tail, aligns at the expected locus, and contains portions of the primer used for RT-PCR). A human-curated summary is also included when further manual analysis has been performed. The third panel, 'Picture', provides a graphical depiction of the 3′UTR region of the transcript along the chromosome. Color-coded tracks show BLAT and WU-BLAST alignments of the UST to the genome in the vicinity of the given transcript: 'Green' glyphs represent USTs that passed our internal quality-control tests, 'Orange' glyphs indicate USTs that have been partially validated and 'Red' glyphs depict USTs that failed our validation tests and have been re-submitted to the cloning pipeline. Also displayed are PicTar and miRanda predictions for miRNA-binding sites, any putative PAS motifs, ESTs and mRNA evidence that support the current transcript models, and conservation with *C. briggsae*. Additional data on functional elements and sequence conservation will be incorporated as new data become available. This 'Picture' panel thus provides a comprehensive snapshot of the 3′UTR and any known or predicted functional elements within it.

The remaining three tabs document raw data for 3′UTRs in our cloning pipeline. First, the 'ABI trace file' page (Figure 1D) allows the user either to view the chromatogram produced by the ABI sequencer corresponding to the given UST, or to download it in SCF format. The chromatogram is rendered graphically using a Java applet, which enables the user to browse the entire sequence trace from 5′ to 3′, to extract the sequence in FASTA format, and view comments produced by the ABI sequencer. This page enables interactive access to the raw sequence data and its inspection at a great level of detail. Similarly, the 'Gel' page (Figure 1E) shows an agarose gel image containing the PCR bands for a set of 96 cloned USTs, with the UST of interest highlighted for easy reference. This raw data can provide information about 3′UTR heterogeneity since additional bands could indicate the presence of multiple, previously undocumented, isoforms in the original mini-pool. We are following up on all such cases to isolate individual alternative 3′UTR isoforms. Finally, the 'Plate' page, designed for internal use, features cloning information such as plate coordinates corresponding to the frozen stocks and barcode information for the various stages in the cloning pipeline.

FUTURE DIRECTIONS

One of the primary goals of the UTRome database is to provide continuous improvements to the comprehensive annotation of 3′UTRs and their functional elements in *C. elegans*. Part of this mission is to provide an interface for our cloning pipeline for curation and quality control, and ultimately to use our data to improve the 3′UTR annotations in genomic repositories like WormBase. As part of the modENCODE Consortium, an initiative

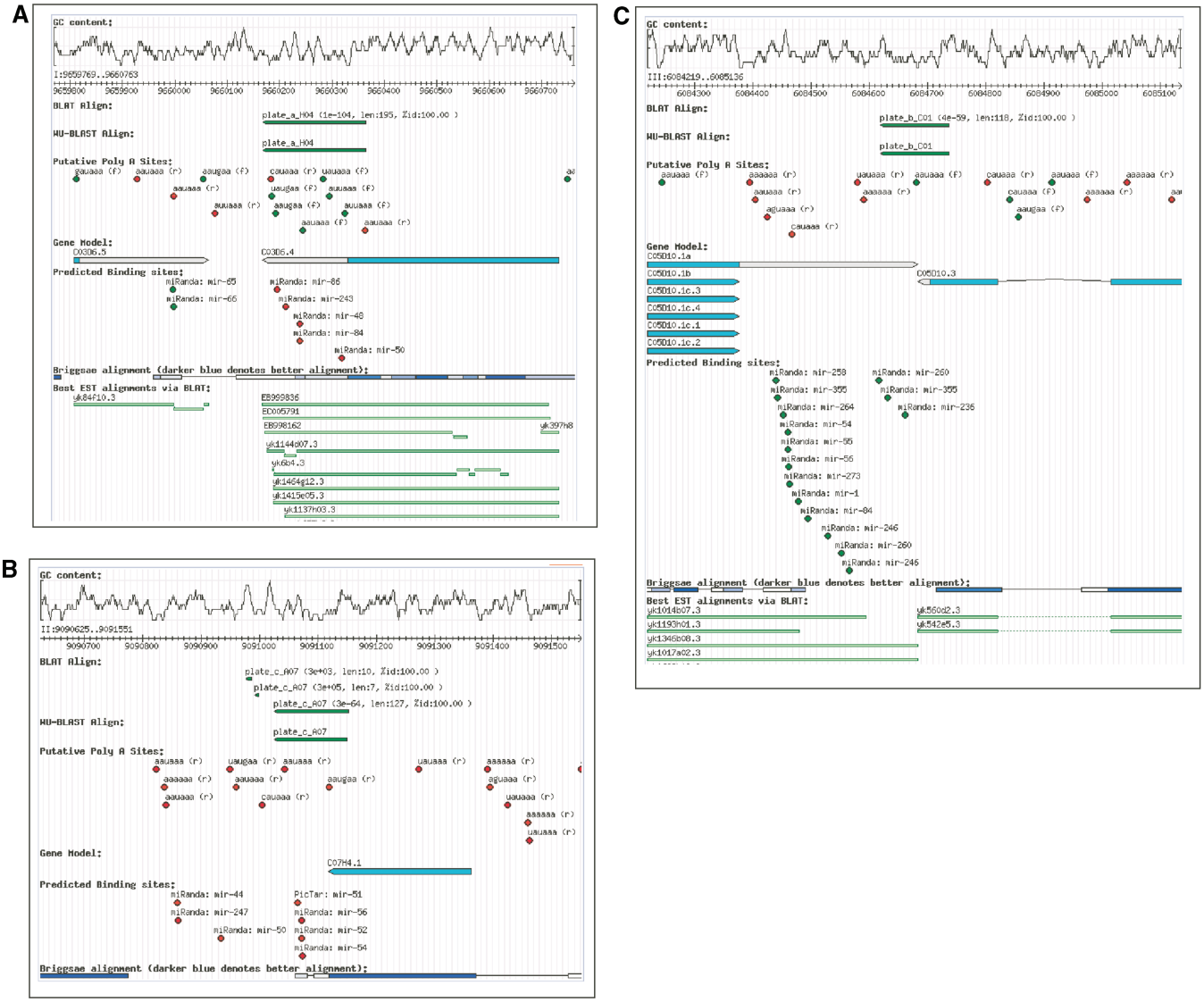

Figure 2. Experimental analysis of 3′UTRs. Three examples of USTs from our cloning pipeline aligned to the *C. elegans* genome using BLAT and WU-BLAST algorithms. (**A**) Validation of previously annotated 3′UTR: the length of the UST produced by our pipeline matches that of the annotated 3′UTR for transcript C03D6.4. (**B**) New experimental 3′UTR evidence for a transcript with no previous experimental support (C07H4.1). The UST contains a putative polyA addition site and a polyA tail, suggesting a true end for this 3′UTR, and overlaps several predicted miRNA-binding sites. (**C**) Improved 3′UTR annotation: evidence for a longer alternate 3′UTR isoform for a gene with a short previously annotated 3′UTR (C05D10.3). The UST sequence overlaps the 3′UTR of another transcript (C05D10.1a) in the opposite orientation. miRNA-binding sites have been predicted for C05D10.1 but not for C05D10.3. If both genes are transcribed simultaneously, these overlapping transcripts could potentially lead to the production of double-stranded RNA and endogenous small interfering RNAs (siRNAs). In all panels, putative polyA signal and miRNA-binding sites are indicated in green or red for transcripts oriented left-to-right or right-to-left, respectively.

from the National Human Genome Research Institute (NHGRI) to provide genome-wide characterization of sequence-based functional elements in the *C. elegans* and *Drosophila melanogaster* genomes (see http://www.modencode.org), we have been tasked to generate high-quality 3′UTR annotations for one-third of the *C. elegans* genome (~7000 genes). 3′UTR data (USTs and validated 3′UTRs) from this set will also flow into the modENCODE Data Coordination Center (DCC) database (to be hosted at http://www.modencode.org). We are continuously updating the UTRome database with new 3′UTR data from our cloning pipeline and plan to extend the project to the entire genome of *C. elegans*. We have also prototyped an *in vivo* pipeline, using fluorescent reporter constructs, to identify functional elements mediating post-transcriptional gene regulation within cloned 3′UTRs (19). We plan to extend and scale up this approach using the library of 3′UTR clones we are currently generating and to incorporate these data into the UTRome database. Over the next few years, we also anticipate a new influx of data for *C. elegans* on expression patterns of miRNAs, 3′UTR isoforms, and improved prediction of functional elements, which we envision incorporating into the UTRome along with additional analysis tools. Our vision for UTRome.org is to provide a comprehensive resource to access and analyze these data,

thus greatly enhancing our overall understanding of 3′UTR biology and helping the scientific community achieve a better understanding of the mechanisms used by cells to control post-transcriptional gene regulation in this and other organisms.

ACKNOWLEDGEMENTS

We thank Danielle and Jean Thierry-Mieg for sharing statistics on alternative transcript isoforms and insightful discussions on sequence curation, Ravi Sachidanandam for kindly providing us with the TraceView Java applet, Michael Zuker for suggestions on how to install and configure MFOLD, Victor Chistyakov for help with the AJAX auto-suggest feature, Nikolaus Rajewsky and his research group for fruitful collaborations on 3′UTR biology, Kevin Chen for helpful comments on the manuscript, and the modENCODE Consortium for propelling this project forward. This work was supported by grants from the National Human Genome Research Institute (R21HG003971 and 1U01HG004276). Funding to pay the Open Access publication charges for this article was provided by NHGRI award 1U01HG004276.

Conflict of interest statement. None declared.

REFERENCES

1. Wickens,M., Bernstein,D.S., Kimble,J. and Parker,R. (2002) A PUF family portrait: 3′UTR regulation as a way of life. *Trends Genet*., **18**, 150–157.
2. Chabanon,H., Mickleburgh,I. and Hesketh,J. (2004) Zipcodes and postage stamps: mRNA localisation signals and their trans-acting binding proteins. *Brief Funct. Genomic. Proteomic*., **3**, 240–256.
3. de Moor,C.H., Meijer,H. and Lissenden,S. (2005) Mechanisms of translational control by the 3′ UTR in development and differentiation. *Semin. Cell Dev. Biol*., **16**, 49–58.
4. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
5. Bushati,N. and Cohen,S.M. (2007) microRNA Functions. *Annu. Rev. Cell. Dev. Biol*., **23**, 175–205.
6. Keene,J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet*., **8**, 533–543.
7. Rajewsky,N. (2006) microRNA target predictions in animals. *Nat. Genet*., **38 Suppl**, S8–S13.
8. Stein,L., Mangone,M., Schwarz,E., Durbin,R., Thierry-Mieg,J., Spieth,J. and Sternberg,P. (2001) WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res*., **29**, 82–86.
9. Bieri,T., Blasiar,D., Ozersky,P., Antoshechkin,I., Bastiani,C., Canaran,P., Chan,J., Chen,N., Chen,W.J. *et al*. (2007) WormBase: new content and better access. *Nucleic Acids Res*., **35**, D506–D510.
10. Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Maeda,N., Oyama,R., Ravasi,T., Lenhard,B. *et al*. (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
11. Takeda,J., Suzuki,Y., Nakao,M., Barrero,R.A., Koyanagi,K.O., Jin,L., Motono,C., Hata,H., Isogai,T. *et al*. (2006) Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. *Nucleic Acids Res*., **34**, 3917–3928.
12. Thierry-Mieg,D. and Thierry-Mieg,J. (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol*., **7**(Suppl 1), S12–14.
13. Hughes,T.A. (2006) Regulation of gene expression by alternative untranslated regions. *Trends Genet*., **22**, 119–122.
14. Sood,P., Krek,A., Zavolan,M., Macino,G. and Rajewsky,N. (2006) Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl Acad. Sci. USA*, **103**, 2746–2751.
15. Majoros,W.H. and Ohler,U. (2007) Spatial preferences of microRNA targets in 3′ untranslated regions. *BMC Genomics*, **8**, 152.
16. Mignone,F., Grillo,G., Licciulli,F., Iacono,M., Liuni,S., Kersey,P.J., Duarte,J., Saccone,C. and Pesole,G. (2005) UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res*., **33**, D141–D146.
17. Brockman,J.M., Singh,P., Liu,D., Quinlan,S., Salisbury,J. and Graber,J.H. (2005) PACdb: PolyA Cleavage Site and 3′-UTR Database. *Bioinformatics*, **21**, 3691–3693.
18. Zhang,H., Hu,J., Recce,M. and Tian,B. (2005) PolyA_DB: a database for mammalian mRNA polyadenylation. *Nucleic Acids Res*., **33**, D116–D120.
19. Lall,S., Grun,D., Krek,A., Chen,K., Wang,Y.L., Dewey,C.N., Sood,P., Colombo,T., Bray,N. *et al*. (2006) A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol*., **16**, 460–471.
20. Lewis,B.P., Shih,I.H., Jones-Rhoades,M.W., Bartel,D.P. and Burge,C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
21. Enright,A.J., John,B., Gaul,U., Tuschl,T., Sander,C. and Marks,D.S. (2003) MicroRNA targets in *Drosophila*. *Genome Biol*., **5**, R1.
22. Krek,A., Grün,D., Poy,M.N., Wolf,R., Rosenberg,L., Epstein,E.J., MacMenamin,P., da Piedade,I., Gunsalus,K.C. *et al*. (2005) Combinatorial microRNA target predictions. *Nat. Genet*., **37**, 495–500.
23. Hajarnavis,A., Korf,I. and Durbin,R. (2004) A probabilistic model of 3′ end formation in *Caenorhabditis elegans*. *Nucleic Acids Res*., **32**, 3392–3399.
24. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*., **31**, 3406–3415.
25. Kent,W.J. (2002) BLAT – the BLAST-like alignment tool. *Genome Res*., **12**, 656–664.
26. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol*., **215**, 403–410.
27. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W. *et al*. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res*., **12**, 1599–1610.
28. Durbin,R. and Thierry-Mieg,J. (1994) The AceDB Genome Database. In Shuhai,S. (ed.), *Computational Methods in Genome Research*. Plenum Press: New York, pp. 45–55.
29. Stein,L.D. and Thierry-Mieg,J. (1998) Scriptable access to the Caenorhabditis elegans genome sequence and other ACEDB databases. *Genome Res*., **8**, 1308–1315.
30. Crosby,M.A., Goodman,J.L., Strelets,V.B., Zhang,P. and Gelbart,W.M. (2007) FlyBase: genomes by the dozen. *Nucleic Acids Res*., **35**, D486–D491.
31. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*., **34**, D140–D144.
32. Vaglio,P., Lamesch,P., Reboul,J., Rual,J.F., Martinez,M., Hill,D. and Vidal,M. (2003) WorfDB: the *Caenorhabditis elegans* ORFeome Database. *Nucleic Acids Res*., **31**, 237–240.
33. Dupuy,D., Li,Q.R., Deplancke,B., Boxem,M., Hao,T., Lamesch,P., Sequerra,R., Bosak,S., Doucette-Stamm,L. *et al*. (2004) A first version of the *Caenorhabditis elegans* Promoterome. *Genome Res*., **14**, 2169–2175.